



La langue en harmonie Influences de la formation musicale sur l'apprentissage précoce d'une langue étrangère

Christelle Dodane

► To cite this version:

Christelle Dodane. La langue en harmonie Influences de la formation musicale sur l'apprentissage précoce d'une langue étrangère. Linguistique. Université de Franche-Comté (Besançon), 2003. Français. <NNT : >. <tel-02573458>

HAL Id: tel-02573458

<https://hal.science/tel-02573458v1>

Submitted on 14 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ de FRANCHE-COMTÉ

Discipline : Phonétique

Présentée et soutenue publiquement

par

Christelle DODANE

Le 11 février 2003

La langue en harmonie

*Influences de la formation musicale sur l'apprentissage précoce
d'une langue étrangère*

Volume I

Directeur de thèse :

Professeur Gabrielle KONOPCZYNSKI

JURY

Mme Caroline DRAKE, Université René Descartes de Paris (Rapporteur)

Mme le Professeur Elisabeth GUIMBRETIERE, Université Paris VII (Membre du jury)

Mme le Professeur Gabrielle KONOPCZYNSKI, Université de Franche-Comté (Directeur de Thèse)

M. le Professeur Philippe MARTIN, Université Denis Diderot, Paris VII (Rapporteur)

M. Jacques MONTREDON, Maître de Conférence HDR, Université de Franche-Comté (Membre du jury)

Sommaire

Sommaire	1
Tables des signes phonétiques	3
Grille à l'usage des non musiciens	7
Tables des sigles et des abréviations	9
Introduction générale	13
Première partie : contexte théorique	
Chapitre I – Pourquoi rapprocher la langue de la musique ?	25
Chapitre II - Indifférenciation de la langue et de la musique au début de la vie	123
Chapitre III - Développement linguistique et musical de l'enfant : différenciation des deux systèmes	189
Chapitre IV - L'apprentissage précoce d'une langue étrangère	217
Deuxième partie : exploration empirique	
Chapitre V - Catalogue d'interférences français-anglais	247
Chapitre VI - Constitution des groupes	315
Chapitre VII - Effets de la formation musicale sur la perception de l'anglais	363
Chapitre VIII - Effets de la formation musicale sur la production de l'anglais	405
Conclusion	481
Table des matières	489
Bibliographie générale	499
Discographie	519
Partitions	520

Table des signes phonétiques

Voyelles orales

- /i/ de livre, pile, mille
- /e/ de épée, blé, thé
- /ɛ/ de mer, rêve, terre
- /a/ de plat, bal, tasse
- /u/ de poule, roux, mou
- /ɔ/ de port, bol, corps
- /œ/ de jeune,
- /ɑ/ de pâte, pâle
- /o/ de peau, seau
- /ø/ de peu, bleu
- /y/ de mule, rue, sur
- /ə/ de petit, ramener

Voyelles nasales

- /ɔ̃/ de pente
- /ɔ̃/ de pont
- /ɛ̃/ de pain
- /œ̃/ de brun

Table des signes phonétiques

Voyelles courtes

- /ɪ/ de **p**ink, **s**ix, **f**ish, **b**aby, **o**range
- /e/ de **r**ed, **s**even, **t**en, **t**welve
- /æ/ de **b**lack, **c**at, **p**arrot, **r**abbit
- /ʊ/ de **b**ook
- /ʌ/ de **o**ne, **d**uck, **u**mbrella, **b**utterfly
- /ɒ/ de **o**range, **d**og, **d**oll
- /ə/ de **b**utterfly, **s**even, **b**anana, **p**arrot, **u**mbrella.

Voyelles longues

- /i:/ de **g**reen, **s**heep, **t**hree, **t**hirteen.
- /ɑ:/ de **c**ar, **b**anana.
- /ɔ:/ de **h**orse, **f**our
- /u:/ de **b**lue, **t**wo
- /ɜ:/ de **b**ird, **p**urple, **t**hirteen, **g**irl

Diphtongues

- /eɪ/ de **e**ight, **b**aby, **t**eddy, **c**ake, **g**rappe
- /aɪ/ de **w**hite, **f**ive, **n**ine, **b**utterfly
- /ɔɪ/ de **b**oy
- /əʊ/ de **y**ellow
- /aʊ/ de **b**rown, **c**ow
- /ɪə/ de **f**ear
- /eə/ de **f**air
- /ʊə/ de **m**oor, **p**oor

Triphthongues

- /aɪə/ de **f**ire, **h**igher, **l**iable
- /eɪə/ de **p**layer, **c**onveyor, **p**raye
- /ɔɪə/ de **r**oyal, **e**mployer, **e**njoyable
- /aʊə/ de **t**ower, **c**oward, **o**ur, **l**ower
- /əʊə/ de **g**oer, **s**lower, **m**ower

Grille à l'usage des non-musiciens

Correspondance des intervalles musicaux en demi-tons

	Nom de l'intervalle	Abréviation	Ecart en demi-tons	Notation musicale	Rapport
DO-DO	Unisson		0		1:1
DO-DO #	Seconde mineure	2de m	1		15:16
DO-RE	Seconde majeure	2de M	2		8:9
DO-RE#	Tierce mineure	3ce m	3		5:6
DO-MI	Tierce majeure	3ce M	4		4:5
DO-FA	Quarte juste	4te J	5		3:4
DO-FA #	Quarte augmentée ou quinte diminuée	4te A	6		
DO-SOL	Quinte juste	5te J	7		2:3
DO-SOL#	Quinte augmentée ou sixte mineure	5te A ou 6te m	8		5:8
DO-LA	Sixte majeure	6te M	9		3:5
DO-LA#	Septième mineure	7 ^{ème} m	10		5:9
DO-SI	Septième majeure	7 ^{ème} M	11		8:15
DO-DO	Octave	8ve	12		1:2
DO-DO# à l'octave	Neuvième mineure	9 ^{ème} m	13		
DO-RE à l'octave	Neuvième majeure	9 ^{ème} M	14		

Table des sigles et des abréviations

ANA : Anglais Nord Américain.

BBC : British Broadcasting Corporation.

BO : Bonne Oreille (musicale).

CAE : Chant Adressé à l'Enfant.

CE : Cour Elémentaire (1 et 2).

CM : Cour Moyen (1 et 2).

CNDP : Centre National de Documentation Pédagogique.

CRDP : Centre Régional de Documentation Pédagogique.

CNR : Conservatoire National de Région.

CP : Cour Préparatoire.

CSP : Catégorie Socio-Professionnelle.

CSP1 : groupe rassemblant les enfants issus des catégories CSP 3 et 4 (selon le classement établi par l'INSEE).

CSP2 : groupe rassemblant les enfants issus des catégories CSP 2, 5, 6, 7 et 8 (selon le classement par l'INSEE).

CV : Consonne-Voyelle (syllabe).

CVC : Consonnes-Voyelle-Consonne (syllabe).

DA : Durée Absolue.

dB : décibel.

DH : Discrimination Harmonique (test musical).

DM : Discrimination Mélodique (test musical).

DMP : Dernier Mot Potentiel (travaux de Grosjean).

DR : Durée Relative.

DR : Discrimination Rythmique (test musical).

EILE : Enseignement d'Initiation aux Langues Vivantes.

ENM : Ecole Nationale de Musique.

EU : Encyclopedia Universalis.

Fo : Fréquence fondamentale.

G1 : Groupe 1 (dans la répartition de la population d'élèves).

G2 : Groupe 2 (dans la répartition de la population d'élèves).

GG : Category Goodness Difference.

HBH : Haut-Bas-Haut (contour mélodique de forme).

Hz : Hertz.

I : (groupe d'enfant qui pratiquent un) Instrument.

IH : Identification Harmonique (test musical).

ILV : Initiation aux Langues Vivantes.

IM : Identification Mélodique (test musical).

INSEE : Institut National de la Statistique et des Etudes Economiques.

IR : Identification Rythmique (test musical).

IRCAM : Institut de Recherche et de Coordination Acoustique/Musique (Centre Georges Pompidou, Paris).

IUFM : Institut Universitaire de Formation des Maîtres.

L2 : langue seconde.

LAE : Langage Adressé à l'Enfant.

LE : Langue Etrangère.

LM : Langue Maternelle.

LSF : Langue des Signes Française.

LV : Langue Vivante.

M : (classe) Musicale (ou enfants musiciens).

M : Majeur.

m : mineur.

MO : Mauvaise Oreille (musicale).

ms : millisecondes

N : Natif.

NA : Non Assimilable (cf. PAM, modèle développé par Best).

NI : (groupe d'enfant qui ne pratiquent pas d') Non Instrument.

NM : (classe) Non Musicale (ou enfants non musiciens).

NN : Non Natif.

OM : Oreille Moyenne (musicale).

OP : Oreille Passable (musicale).

PAM : Perceptual Assimilation Model (modèle de perception développé par Best).

PCS : (nomenclature des) Professions et Catégories Socio-Professionnelle (nouvelle classification INSEE de 1994).

PS : Partie Stable (d'une voyelle).

Q : Queue (d'une voyelle).

RD : Répétition Différée.

RI : Répétition Immédiate.

RP : « Received Pronunciation », appelé également « B.B.C. English », anglais britannique considéré comme standard.

SC : Single Category (cf. PAM, modèle développé par Best).

SMM : Système Musical Maternel.

T : Tête (d'une voyelle).

TBO : Très Bonne Oreille (musicale).

TC : Two Categories (cf. PAM, modèle développé par Best).

TEP : Tomographie à Emission de Positons (technique d'imagerie cérébrale).

TMO : Très Mauvaise Oreille (musicale).

UC : Uncategorized versus categorized (cf. PAM, modèle développé par Best).

UU : Both Uncategorizable (cf. PAM, modèle développé par Best).

Introduction

Lorsque les voix se mélangent au point de ne plus former qu'un tout vaste et indifférencié, nous essayons, en « tendant » l'oreille, de tirer une organisation de ce chaos sonore.

Et soudain, du bruit ambiant, émerge la musique d'une langue :

« (...) Dans toute la salle, pendant les intervalles des entr'actes, les conversations italiennes et la musique d'une langue étrangère parlée par des femmes ajoutaient encore à l'enchantement de cette soirée (...) ». Baudelaire, *les Paradis Artificiels, Mangeur d'opium*, III.

Si cette musique nous est familière, nous l'oublions bien vite en essayant de fixer notre attention sur le contenu pour saisir le sens des paroles qui sont échangées.

En revanche, si cette musique nous est inconnue, nous ne pouvons qu'en suivre le chant. Nos habitudes perceptives sont désorientées par cette nouvelle organisation musicale et nous échouons à repérer les mots, car nous ne savons pas déchiffrer cette organisation. A l'inverse, lorsqu'un locuteur étranger s'exprime dans notre langue maternelle, il transpose souvent la musique de sa propre langue sur son contenu segmental. L'harmonie maternelle se brise pour donner naissance à un chant différent, qui lui donne une couleur savoureuse, un exotisme fascinant. « *Toute les langues chantent, mais pas sur le même air* » et c'est tout le charme des accents étrangers que de nous le révéler. De même, on dit de certains accents régionaux qu'« ils chantent ». L'oreille, charmée, apprivoise cette nouvelle musique et quelquefois nous donne l'envie de chanter nous aussi : on « attrape » l'accent sans même s'en rendre compte.

A cause de l'importance qui est donnée à l'écriture dans notre culture, nous ne sommes plus très sensibles à ce chant, que nous oublions souvent au profit des mots. Pourtant les éléments prosodiques que sont le rythme et l'intonation constituent les fondations de la langue et ils entretiennent un rapport de proximité maximale avec le rythme et la mélodie de la musique.

Cette proximité de la langue et de la musique a largement été exploitée et ce, depuis très longtemps, par les philosophes, les linguistes, les phonéticiens, les compositeurs et les musiciens. Le rapprochement entre la langue et la musique ne constitue donc pas une approche nouvelle, ni originale, mais à notre niveau, il représente une démarche assez personnelle.

Si nous avons voulu, à notre tour, nous intéresser à la musique de la langue, c'est avant tout parce que nous sommes musicienne et ce, depuis l'enfance. La musique a toujours fait partie de notre vie et elle constitue pour nous une véritable passion. L'associer à la discipline qui nous tient le plus à cœur au sein d'un travail de recherche, représente le plus bel hommage que nous puissions lui consacrer, ainsi qu'à ceux qui lui donnent vie en la composant et en l'interprétant.

La musique et la phonétique

C'est au tout début de notre enseignement de phonétique que la nature essentiellement musicale de la langue nous est apparue comme une évidence, notamment, grâce aux travaux de Delattre et à la découverte de son système de transcription des intonèmes du français (cf. chapitre I). Pour représenter l'évolution de l'intonation au cours du temps, il utilise une portée musicale légèrement modifiée (cf. chapitre I). L'utilisation de la notation musicale implique donc une forte similitude entre la mélodie de la langue et celle de la musique.

L'hypothèse d'un transfert de compétence

A cause de ces similitudes, une oreille entraînée par la pratique musicale pourrait donc se montrer plus habile à traiter la musique de la langue. C'est après avoir établi ce constat que nous avons décidé de débiter une recherche comparée sur la langue et la musique. En outre, nous avons pu remarquer un transfert de compétence de la pratique musicale à la pratique de la phonétique. Lorsque nous avons commencé à travailler avec les représentations instrumentales utilisées en phonétique, c'est essentiellement notre oreille, entraînée par la pratique musicale, qui nous a guidée dans nos premières interprétations et nos premières analyses. En effet, grâce à la transcription musicale, il est très courant en musique et plus spécifiquement en solfège, de passer du phénomène sonore en tant qu'objet perçu à sa représentation visuelle et réciproquement. Cet exercice de va et vient est difficile et exige beaucoup de concentration. Ainsi, le simple fait de dessiner une courbe mélodique ne constitue en rien une démarche naturelle, mais est le fruit d'un héritage culturel (l'association du registre grave à la partie basse et du registre aigu à la partie haute d'une représentation visuelle résulte de l'histoire de la notation musicale, cf. chapitre I, 1.2.2.2.). Certaines personnes éprouvent d'ailleurs des difficultés à dessiner l'évolution du contour d'une courbe mélodique sur une feuille de papier ou dans l'espace, à l'aide d'un geste. Dans notre pratique de phonéticienne, nous avons donc constaté un transfert de compétence de la formation musicale à l'interprétation auditive des données phonétiques.

L'apprentissage précoce des langues

C'est sur cette hypothèse de transfert de compétence que s'articule notre travail. Mais, comment parvenir à mettre à jour un tel transfert de compétence dans une recherche de type expérimental ? Il est tout d'abord nécessaire de commencer par comparer les performances des musiciens à celles des non musiciens afin de pouvoir déterminer si une oreille entraînée permet de mieux percevoir et de mieux restituer les caractéristiques phonétiques d'une langue. Il est cependant difficile de travailler avec la langue maternelle car son acquisition se fait de manière très précoce et elle est déjà bien installée chez les enfants de 5 ans. Or, à cet âge, tous les enfants n'ont pas commencé une formation musicale et il serait très malaisé de rassembler un échantillon d'enfants musiciens. En revanche, il serait intéressant de travailler avec une langue étrangère, car comme nous l'avons souligné au début de ce texte, elle possède une organisation prosodique, et donc musicale, différente de notre langue maternelle. Or, au cours du développement de l'enfant, il existe une période pendant laquelle l'apprentissage d'une langue

étrangère est relativement proche de l'acquisition de la langue maternelle. Bien sûr, les processus mis en jeu sont différents mais avant l'âge de 8 ans, cet apprentissage se fait de manière « naturelle » avant de devenir complètement conscient et réfléchi. Lorsque l'apprentissage d'une langue étrangère intervient avant l'âge de 11 ans, âge qui correspond en France à l'entrée au collège, on parle d'apprentissage « précoce ». Nous avons donc choisi de travailler avec des enfants en apprentissage précoce d'une langue étrangère et de comparer leurs performances en fonction du fait qu'ils sont musiciens ou non musiciens.

Perception / production

Dans une formation musicale traditionnelle (en conservatoire par exemple), c'est l'oreille qui est éduquée en premier (solfège), les premières production musicales intervenant plus tardivement (chant et instrument). La priorité est donc donnée à la perception et la production n'est jamais traitée séparément de la perception (lorsque l'on joue d'un instrument, on contrôle sa production grâce à son oreille exactement comme lorsque l'on parle). Cependant, la phonétique reste une discipline le plus souvent focalisée sur la production, notamment à cause de la multiplication des représentations de type instrumental qui permettent de visualiser l'onde sonore. Il est vrai, et les Laboratoires Haskins l'ont montré dans les années 60, qu'il est difficile d'établir des relations biunivoques entre des indices acoustiques et la perception de ces indices.

De ce fait, perception et production sont souvent étudiées de manière isolée, les chercheurs étant formés exclusivement dans l'un ou l'autre domaine. Indirectement, le parallèle avec la musique en souffre car il repose sur les liens qu'entretiennent la perception et la production. En revanche, parmi les chercheurs qui travaillent sur la perception, certains abordent le parallèle entre langue et musique avec plus de profondeur et c'est finalement en psychologie cognitive que l'on trouve le plus de travaux à ce sujet (et notamment en France, Bigand, McAdams, Drake, Clarke... et à l'étranger Trehub, Jones et Yee, Peretz, Carlson, Krumhansl, Lerdahl, Jackendoff,...).

Nécessité et limites d'une inter / multidisciplinarité

Un travail qui s'intéresse au parallèle langue / musique se doit d'être avant tout inter et multidisciplinaire et comme le souligne le compositeur et musicologue Schaeffer (1966 : 31),

« une attitude plus ambitieuse consiste à proposer entre toutes, la musique comme une activité « globalisante », comme une interdiscipline proprement dite, une activité qui, recoupant de multiples disciplines spécifiques, vérifie par synthèse leurs apports partiels, tant sur le plan des faits que sur celui des idées, et se présente au même titre qu'elles comme une activité de découverte, qui vise autant, sinon plus, à fonder une connaissance qu'à créer des œuvres ».

Si une telle approche se révèle très fructueuse, elle est aussi dangereuse de par son ampleur. Il est en effet difficile de faire une synthèse de travaux relevant de plusieurs disciplines qui ne constituent pas nos disciplines de spécialité. Le risque de mauvaises interprétations est plus grand et le danger de ne pas approfondir suffisamment ou, au contraire, de se noyer dans les

détails, est constant.

Dans les disciplines pour lesquelles nous ne sommes pas spécialiste, nous avons donc essayé, dans la mesure du possible, d'éviter les erreurs d'interprétation en faisant appel à des spécialistes, notamment des musicologues pour le premier chapitre. Mais ces précautions n'évitent pas le fait que nous ayons avant tout une approche de phonéticienne doublée de celle d'une musicienne, ce qui limite forcément notre regard à cause de notre spécialisation. Nous avons également conscience que cette approche peut introduire, non seulement de mauvaises interprétations, mais également la méconnaissance d'autres travaux qui ont abordé cette problématique.

L'exercice est difficile, mais à l'inverse, le rapprochement entre la langue et la musique ne peut se faire que d'une manière multidisciplinaire. C'est le paradoxe d'une telle entreprise. En effet, comme le souligne Edgard Morin :

« A quoi serviraient tous les savoirs parcellaires sinon à être confrontés pour former une configuration répondant à nos attentes, à nos besoins et à nos interrogations cognitives ? »
(Edgard Morin, Transversales Sciences Culture, n°29, 8).

Contenu de la thèse

L'objectif principal de notre partie théorique sera donc de constituer une revue de littérature la plus complète possible (sans pour autant nous départir d'une certaine limitation) et de poser les jalons d'une réflexion la plus ouverte possible, en essayant d'identifier et d'éviter les préjugés qui pourraient nous conduire à poser des jugements simplistes sur les rapports qu'entretiennent la langue et la musique. La première partie se décompose en quatre chapitres :

Le premier chapitre propose une description des similitudes de formes qui existent entre langue et musique et un relevé des cas où cette similitude est maximale (comme dans les langues dites « musicales ») de façon à asseoir solidement le parallèle entre la langue et la musique. Une étude historique montre ensuite comment ces similitudes de formes ont été exploitées par ceux qui manipulent la substance sonore, d'un côté les phonéticiens, et de l'autre les compositeurs et les musiciens. Le chapitre s'intéresse enfin aux similitudes structurales qui ont été tirées de ces similitudes de formes par les linguistes, les philosophes et les musicologues.

Le second chapitre part du constat que, s'il existe des similitudes de forme et de structure entre la langue et la musique, il doit exister des similitudes de traitement. Pour le savoir une description du développement linguistique et musical de l'enfant au cours de la première année de sa vie s'impose. En effet, pendant cette période, la musique et la langue sont encore largement indifférenciées et la sensibilité de l'enfant est essentiellement musicale. Ce chapitre montre comment, avec le temps, une différenciation progressive entre la langue et la musique

se met en place pour permettre l'acquisition progressive de la langue maternelle et du système musical maternel.

Cette description repose essentiellement sur un bilan des travaux réalisés en psycholinguistique. Elle s'organise en fonction des trois principes suivants : au début de la vie, la perception précède la production ; la perception est globale avant d'être analytique et enfin, l'organisation prosodique sous-tend l'organisation segmentale.

Le troisième chapitre explore les deux voies nettement séparées qu'empruntent ensuite la langue et la musique au cours de l'enfance, en essayant toutefois de faire apparaître des similitudes dans l'acquisition de la langue maternelle et celle du système musical maternel. Cette description sera moins détaillée car la quantité de travaux consacrée au développement linguistique de l'enfant est inversement proportionnelle à son âge (sauf au niveau lexical et syntaxique). En fonction du bilan réalisé dans ce chapitre et dans le chapitre précédent, nous exposerons nos hypothèses de recherche sur le transfert de compétence de la musique à la langue, ainsi que le type de situation et de population qui serait le plus approprié pour une étude de cette nature.

Le quatrième chapitre interroge la littérature sur la période « idéale » pour l'apprentissage d'une langue étrangère, période pendant laquelle les capacités de l'enfant seraient maximales pour une bonne restitution de la prononciation. Ce chapitre confronte ensuite ces informations à la situation de l'apprentissage précoce d'un point de vue institutionnel. Il en réalise un bref historique pour ensuite dresser un bilan des réformes mises en place en France au cours des dix dernières années. Il pose enfin la question de savoir si l'état actuel de l'apprentissage précoce s'accorde avec ce que l'on sait sur le développement de l'enfant et l'idée d'une période optimale.

La deuxième partie expose l'exploration empirique qui tentera de mettre à jour un transfert de compétence de la formation de l'oreille musicienne sur la perception et la production d'une langue étrangère. Les cinquième et sixième chapitres exposent le travail préparatoire (réalisation d'outils d'investigation) et méthodologique nécessaire à la mise en place de l'expérimentation qui sera présentée dans les septième et huitième chapitres. La deuxième partie se décompose donc également en quatre chapitres.

Le cinquième chapitre apporte une justification au choix de l'anglais comme langue utilisée dans notre expérimentation, puis il propose une description comparée des systèmes prosodiques et vocaliques de l'anglais et du français, description qui nous permettra de prédire les interférences qui devraient amener un apprenant francophone à faire des erreurs en anglais (réalisation d'un catalogue d'interférences) et à partir de cette prédiction, de réaliser deux batteries de tests de perception qui seront utilisées dans les études n°1 et n°2 présentées dans le chapitre VII.

Le sixième chapitre expose la méthodologie que nous avons utilisée pour définir les deux groupes d'enfants qui constitueront notre échantillon, enfants musiciens et enfants non musiciens, mais également les obstacles rencontrés et les difficultés inhérentes à la nature de ce travail.

Le septième chapitre expose deux études réalisées pour déterminer les effets de la formation musicale sur la perception de l'anglais : étude n°1 « Perception de la prosodie » et étude n°2 sur la « Perception de contrastes vocaliques », fondées sur les résultats aux tests de perception (élaborés à partir de l'expertise réalisée dans le chapitre V) par deux groupes d'enfants scolarisés en CE1, un groupe de musiciens et un groupe de non musiciens tels qu'ils ont été définis dans le chapitre VI, tous en apprentissage précoce de l'anglais.

Le huitième chapitre expose trois études réalisées pour déterminer si le fait d'avoir une oreille entraînée par la pratique musicale améliore la production de l'anglais : étude n°3 « Production de la prosodie », étude n°4 « Productions des noyaux vocaliques » et étude n°5 « Stabilité de timbre ». La population étudiée est identique à celle des études n°1 et n°2 et leurs productions ont été recueillies au cours d'un entretien réalisé avec un locuteur anglophone natif.

Une discussion générale vient faire la synthèse de tout ce qui a été abordé au long de ces pages et tente de tirer les fruits d'un rapprochement langue / musique à un niveau pratique et pédagogique.

Remarques bibliographiques

Chaque chapitre comprend au début, une table des matières détaillée et à la fin, une bibliographie complète. Nous avons choisi de présenter une bibliographie par chapitre car certains chapitres possèdent une relative indépendance au niveau de leur contenu et des disciplines abordées. Pour cette raison, nous avons voulu différencier leurs références de façon à ne pas les noyer dans une bibliographie générale. Une bibliographie générale qui propose un classement par ordre alphabétique de la totalité des références, se trouve bien évidemment à la fin de ce volume.

Par ailleurs, les annexes ont été rassemblées dans un second volume de manière à en faciliter la lecture. Un CD d'annexes sonores accompagne le présent volume et les numéros des pistes sont indiqués en caractères gras dans le corps du texte pour permettre au lecteur d'écouter les extraits sonores simultanément avec sa lecture.

[†]Company's, citi par Konopczynski (1980: 327).

Première partie

Contexte théorique

Chapitre I

Pourquoi rapprocher langue et musique ?

« Toute oreille, même non musicienne, sait discerner le rythme de la phrase, les inflexions de son intonation, les accentuations et les fluctuations de son intensité, comme les formes et les variantes de son timbre puisque c'est leur assemblage qui fait sens. Autrement dit, dans la parole, tout est déjà composé [...] On le sait, un flux parlé génère une mélodie, celle de la langue à travers son locuteur. Aujourd'hui, les outils d'analyse du son montrent qu'elle contient aussi une harmonie, une harmonie naturelle en sorte », Claudy Malherbe (disponible sur le site : <http://sol.ircam.fr/instruments/musicweb.fr>).

Table du premier chapitre

1. Éléments pour une analogie entre langue et musique

- 1.1. Un parallèle implicite posé entre prosodie et musique
- 1.2. Similitudes et divergences de forme
 - 1.2.1. Rythme musical et rythme verbal
 - 1.2.1.1. Définition commune
 - 1.2.1.2. Musique et métricité
 - 1.2.1.3. Parole et périodicité sommaire
 - 1.2.2. Mélodie de la musique et mélodie de la parole
 - 1.2.2.1. Successivité et intervalles
 - 1.2.2.2. Le fait musical au travers du filtre de la transcription
 - 1.2.2.3. Des précautions à prendre
 - 1.2.2.4. Tension tonale dans la musique et dans la langue
 - 1.2.2.5. Différenciation mélodie langue / musique
 - 1.2.3. A mi chemin entre musique et langue : les clichés mélodiques
 - 1.2.4. Degré de mélodicité et de rythmicité, la clé de la musicalité ?
- 1.3. Cas de fusion avec la musique : les langues « musicales »
 - 1.3.1. Langues sifflées
 - 1.3.2. Langues tambourinées
 - 1.3.3. Langue artificielle musicale
- 1.4. L'origine de la langue dans la musique ?

2. Exploitation des similitudes de forme

- 2.1. Lorsque la musique devient langue
 - 2.1.1. Transcription musicale de la prosodie
 - 2.1.1.1. Potentialités de la transcription musicale
 - 2.1.1.2. Historique des applications de la transcription musicale en prosodie
 - 2.1.1.3. Transfert de la terminologie musicale vers la prosodie
 - 2.1.2. D'autres systèmes de transcription : leur validité par rapport à la transcription musicale
- 2.2. Lorsque la langue devient musique : Evolution du traitement de la similitude entre la langue et la musique au cours de l'histoire de la musique occidentale
 - 2.2.1. Rapports texte et mélodie : information vs musicalité
 - 2.2.1.1. Le chant grégorien : une musique du texte
 - 2.2.1.2. Le récitatif dans l'opéra :
 - 2.2.1.2.1. A la frontière de la parole et de la musique
 - 2.2.1.2.2. Trois récitatifs, trois langues, trois musiques
 - 2.2.1.2.1.1. Anglais : « *Didon et Énée* » de Purcell
 - 2.2.1.2.1.2. Allemand : Les « *Passions de St Matthieu* » de Bach
 - 2.2.1.2.1.3. Italien : « *Don Giovanni* » de Mozart

- 2.2.1.3. La musicalité des langues dans la mise en musique :
 - 2.2.1.3.1. Langues musicales versus langues non musicales ?
 - 2.2.1.3.2. Un obstacle à la traduction des œuvres musicales ?
- 2.2.2. La crise de la mélodie ou le retour à l'intonation
 - 2.2.2.1. Debussy et la musicalité de la langue française
 - 2.2.2.2. Le « Sprechgesang »
- 2.2.3. Musicalisation de la parole dans la création contemporaine
 - 2.2.3.1. La langue traitée comme matériau : la voix chez Boulez et Berio
 - 2.2.3.2. Le musicien « phonéticien »
 - 2.2.3.3. Typologie des effets « langue » dans la musique

3. Similitudes de formes étendues à des similitudes « structurales »

- 3.1. Universaux
- 3.2. Similitudes structurales
- 3.3. Peut-on parler de grammaire et de langage à propos de la musique ?
- 3.4. Capacité cognitive commune

Bibliographie du chapitre I

Si notre intuition nous suggère une situation de proximité maximale entre les éléments prosodiques et la musique, il convient de transformer cette intuition en une description précise des similitudes de forme qu'elles entretiennent, qui viendra justifier l'analogie. Mieux connaître ce qui rapproche le rythme verbal du rythme musical et la mélodie de la langue de celle de la musique, c'est également mieux identifier la frontière ténue qui les sépare. Cette frontière devient très floue dans les clichés mélodiques et elle s'abolit complètement dans les langues « musicales », qui n'utilisent plus que le niveau musical comme vecteur du niveau linguistique. Des linguistes et des philosophes ont d'ailleurs vu l'origine de la langue dans la musique, certains posant l'hypothèse d'une origine commune. Face au manque de système de notation pour les éléments musicaux de la langue, c'est donc tout naturellement que les phonéticiens ont fait appel à la transcription musicale et l'ont adaptée à leur objet. A l'inverse, les similitudes avec la langue ont toujours inspiré les compositeurs et les rapport entre le « texte » et la « musique » constituent même la trame de fond de l'histoire de la musique vocale occidentale. L'étude détaillée de ces rapports nous permettra de compléter et d'enrichir notre approche de phonéticienne grâce aux réflexions très subtiles menées depuis des siècles par les spécialistes de l'autre « versant ». Enfin, il n'est guère étonnant que ces similitudes de formes aient à leur tour attiré les théoriciens qui ont cherché à les étendre à des similitudes structurales.

1. Éléments pour une analogie entre langue et musique

1.1. Un parallèle implicite posé entre prosodie et musique

L'existence de liens étroits entre musique et prosodie a poussé certains chercheurs à parler d'éléments musicaux pour désigner le rythme et l'intonation, notamment Faure (1962 : 7) dans sa recherche sur le rôle des éléments musicaux dans la prononciation anglaise :

« Au cours de la présente étude, nous entendrons par éléments musicaux de la prononciation anglaise tout ce qui, dans cette langue, se combine à l'articulation des sons pour permettre au discours d'assurer la communication d'un certain contenu de pensée ou d'émotions. Ces éléments sont essentiellement constitués par les variations de hauteur qui affectent, de diverses façons, les syllabes successives d'un énoncé et qui, s'accompagnant presque toujours de variations concomitantes d'intensité et de durée, en déterminent à la fois le rythme et la mélodie ».

Or, parler d'éléments musicaux à propos de la langue, c'est poser implicitement un parallèle entre musique et langue. Et selon Menezes (1993 : 81) :

« du point de vue de la réception linguistique, c'est indubitablement à travers l'intonation prosodique que l'on s'aperçoit de façon immédiate de la contiguïté innée entre langage et musique ».

En se penchant sur l'étymologie de ces termes, il est d'ailleurs frappant de constater qu'ils ont une origine musicale. En effet, le terme « prosodie » provient de l'ancien grec « ôdê », qui veut dire chant ou plus précisément chant accompagné d'une musique instrumentale. Il a ensuite été utilisé pour désigner la « science de la versification » et la métrique, qui gouverne la voix humaine lorsqu'elle est en train de lire de la poésie. Quant au terme « intonation », il a lui aussi une origine musicale : il vient du latin « intonare », qui signifie tonner, faire retentir¹. Rossi (1999) explique que :

« par une fausse étymologie avec tonus (ton), intonation est à l'origine un terme musical utilisé pour désigner l'action d'entonner un air. Ce n'est guère avant le XIXe siècle qu'on l'utilise également pour désigner les tons de la parole. A la fin du XVIIIe, son usage linguistique n'est pas encore consacré, puisque Walker, en 1787, dans son ouvrage, The Melody of Speaking Delineated, consacré à l'intonation de l'anglais, ne l'utilise pas encore. Au XIXe, le terme intonation désigne, chez Chateaubriand, « les tons de la voix » ; il est donc synonyme de musicalité, mélodie ; déjà le glissement de sens est manifeste par rapport à la signification originelle».

Pour Scripture (1902 : 472), la description de la mélodie de la parole devrait prendre en compte le traitement de la mélodie dans le chant, et spécialement les chants primitifs et les chants spontanés des enfants².

¹ Intonation (1372) : « Golein, du latin intonare, tonner, faire retentir, rattaché par fausse étymologie à tonus, ton ». Dictionnaire Étymologique, Larousse, 1964.

² Scripture (1902 : 472) : « a discussion of the melody of speech should, perhaps, include a treatment of melody in song, particularly the primitive song of uncultured peoples and the spontaneous song of children ».

S'intéressant justement aux premières productions vocales de l'enfant, qu'elles soient langagières ou musicales, Papousek et Papousek (1981 : 166) soulignent cette interdépendance entre langage et musique :

« The evolvement and interdependence of language and music are phenomena common to all human cultures, and their universality need not to be documented ».

Et pour mieux asseoir ce parallèle, ils parlent également d'éléments musicaux pour désigner les éléments prosodiques (1981 : 164), exactement comme le faisait Faure :

« Musical elements – pitch, tone, interval, melody, loudness, accent, tempo and rhythm – or their acoustic correlates – frequency, harmonics, amplitude and temporal patterns – are matters of a primary interest to musicologists but have not attracted much attention of linguists or phoneticians studying language acquisition³ ».

L'utilisation de l'expression « éléments musicaux » permet de concilier l'approche linguistique et l'approche « musicologique », cette double approche étant nécessaire à la description linguistique, tout au moins au début de la vie :

« Dealing in details with the infant's earliest perception and production of vocal elements which represent musical element to a musicologist, suprasegmental features to a phonetician, and prosodic or paralinguistic elements to a linguist, we are going to speak of “musical” rather than “prosodic” elements. Thus we want to facilitate our discussion of both linguistic and musicological aspects. For the same purpose our procedures include musical transcriptions as well » (Papousek et Papousek, 1981 : 164).

1.2. Similitudes et divergences de forme

Si le parallèle posé entre la langue et la musique est né des similitudes de forme qu'elles entretiennent, encore faut-il décrire ces similitudes. Mais la plupart des auteurs ne se donnent pas la peine de le faire ou alors se contentent de quelques arguments donnés au compte-goutte. Il s'agit donc de justifier le fait qu'on puisse parler de musique à propos de la langue, et plus précisément, de déterminer ce qui est commun au rythme musical et au rythme verbal et ce qui est commun à la mélodie de la musique et à la mélodie de la parole.

³ Après 1980, les phonéticiens ont comblé leur retard et nous verrons dans le chapitre II que de nombreux travaux ont été menés sur la mise en place des éléments musicaux dans l'acquisition du langage.

1.2.1. Rythme musical et rythme verbal

1.2.1.1. Définition commune

Qu'est-ce que le rythme ? A la base de la sensation de rythme, il faut une séquence d'au moins deux éléments qui se distinguent sur le plan perceptif. Comment certains éléments se distinguent-ils des autres ? Parce qu'ils sont marqués par une proéminence, par un accent (l'accent est défini comme un « *fait local de proéminence* » par Di Cristo, 1999 : 185). Comment se manifeste cette proéminence ? Par un agencement spécifique des paramètres acoustiques que sont la fréquence, la durée et l'amplitude, qui va engendrer une mise en relief au niveau perceptif (hauteur, durée et intensité) et au niveau syntagmatique. C'est donc l'alternance des éléments saillants et des éléments non saillants qui va donner naissance au rythme. Selon la définition du Littré, le rythme dans la langue se définit par une **succession** de syllabes **accentuées** (sons forts) et de syllabes non **accentuées** revenant à certains intervalles. Dans ce sens, les syllabes constituent donc des unités rythmiques.

« Le rythme est fondé sur l'organisation temporelle des syllabes métriquement fortes et des syllabes métriquement faibles », (Di Cristo, 1999 : 186).

Appliqué à la musique, il désigne une **succession** de **temps forts** et de **temps faibles**. La définition est donc similaire pour la langue et pour la musique. Si l'on revient à l'étymologie du mot, *ruthmos* (du grec *rein*, qui signifie « couler » au sens d'un ruisseau qui coule) signifie chez les auteurs grecs :

« forme distinctive, figure proportionnée, disposition », ou encore « arrangement caractéristique des parties dans un tout », « manière particulière de fluer »,

et, chez Platon, pour la première fois,

« forme du mouvement que le corps humain accomplit dans la danse » (Benveniste, 1951).

Sauvanet (2002) pose donc la définition globale suivante :

*« tout phénomène, **perçu ou agi**, auquel on peut attribuer au moins deux des qualités suivantes: structure, périodicité, mouvement» et la définition restreinte : « tout phénomène, perçu ou agi, auquel on peut attribuer chacune de ces trois qualités ».*

Cette définition s'applique à la fois à la production et à la perception de la musique et de la langue. Un rythme fort combine les trois : la musique et la langue sont des **rythmes forts**. Elles combinent la structure (la forme, avec l'agencement du tout et des parties), la périodicité (le retour régulier de temps forts ou des syllabes accentuées) et le mouvement (exécution musicale et parole). Mais, dans certaines pièces de musique contemporaine, on trouve une structure, un mouvement, mais pas forcément de périodicité.

Il existe une tradition de description du rythme musical qui s'enracine dans ces similitudes entre rythme verbal et rythme musical. En effet, elle utilise un système basé sur les pieds de la poésie pour décrire le rythme musical. Cette tradition remonte au temps d'Aristoxène et

« il est possible d'en suivre la trace depuis les théoriciens de la Renaissance et du XVIIIe, jusqu'à Riemann, le théoricien de la musique du XIXe » (Clarke, 1989 : 26).

Plus récemment, elle fonde la description que font Cooper et Meyer (1960) de la structure rythmique de la musique : ils ont en effet constitué un répertoire de cinq types rythmiques inspirés de la prosodie grecque (le trochée, le dactyle, l'iambe, l'anapeste et l'amphibraque), définis selon la répartition des éléments accentués et non accentués (tableau n°1).

Trochée	Dactyle	Iambe	Anapeste	Amphibraque
				

Tableau n°1 : figures rythmiques correspondant aux cinq types rythmiques définis par Cooper et Meyer (1960).

1.2.1.2. Musique et métricité

Malgré ces similitudes, le cadre rythmique de la musique est pourtant différent de celui de la langue. Selon Jones et Yee (1994 : 108), la musique occidentale

« semble nous inviter à frapper le rythme avec le pied et à le suivre, ce qui suggère une apparente métricité ».

Sur la partition musicale, les structures sont effectivement organisées de façon mesurée : des points d'appui périodiques délimitent les mesures, la pulsation se définissant comme le retour régulier de ces points d'appui. Chacun d'entre eux est associé à un temps en musique.

Afin de rendre le rythme lisible, la notation musicale répartit ces temps au sein d'unités temporelles de taille égale, les mesures qui sont délimitées par des barres de mesure. On trouve ainsi des mesures comportant 2, 3, 4, 5 temps, etc... Selon Clarke (1989 : 27) :

« le rythme s'organise autour d'un cadre d'unités isochrones répétées (la pulsation), de telle sorte que de grandes périodes temporelles sont reliées l'une à l'autre par des rapports de nombres entiers. On obtient ainsi un cadre stable qui, entre autres choses, facilite la coordination des différents interprètes dans les ensembles ».

Lorsque ces unités sont isochrones et reliées entre elles par des rapports temporels simples, les auditeurs sont capables de se synchroniser avec le rythme et d'en extraire facilement la pulsation en la frappant (Michon, 1967 ; Wing et Kristofferson, 1973a, b), et ce, même après la fin de la séquence. Povel (1981) a montré en les chronométrant, que les battements coordonnés avec le plus de précision étaient ceux qui reproduisaient un rapport temporel simple et notamment, le plus simple d'entre eux, le rapport 2:1.

Les auditeurs sont donc capables d'engendrer et d'appliquer des relations métriques pour lier des groupes et des sons. Mais ils vont plus loin qu'extraire la seule pulsation. Nous avons vu que le rythme musical se définissait comme une succession de temps forts et de temps faibles. Quel que soit leur « poids » respectif, ces temps s'écoulaient régulièrement. Dans une mesure, le premier temps est toujours fort, le deuxième et le troisième sont faibles, sauf dans la mesure à quatre temps où le troisième est fort. Chaque temps se divise lui même en une partie initiale forte et une partie finale faible. Après avoir extrait la pulsation, les auditeurs, sensibles à cette hiérarchie rythmique, sont capables de repérer le ou les temps forts au sein de la mesure.

« Dans le cas d'un rythme de valse par exemple, l'auditeur repérera tout d'abord un premier niveau de battement régulier au niveau de la noire, puis il interprétera un de ces battements sur trois comme étant le plus important » (Bigand, 1994 : 273).

Pour un même morceau de musique, l'auditeur peut extraire une pulsation différente, dont les battements seront plus rapprochés (tempo rapide) ou au contraire plus éloignés (tempo lent). Ces différences de pulsation reflètent en réalité des niveaux hiérarchiques différents. Lerdahl et Jackendoff (1983) ont montré que la musique répondait à une organisation hiérarchique d'unités de durées variables : au plus bas niveau, les unités sont constituées d'un seul élément.

Ces unités se combinent ensuite pour créer d'autres unités plus longues, à des niveaux plus élevés. Plus l'auditeur sera expérimenté (acculturation avec le système musical classique et niveau de formation musicale), plus il aura accès à un grand nombre de niveaux. Il pourra ainsi organiser les événements sur de plus grands groupes temporels et « choisira » donc spontanément une pulsation plus lente, reflétant un niveau hiérarchique élevé (Drake et al., 2000a). La figure n°1 extraite de Drake et al. (2000b : 253) nous donne un exemple des différents niveaux de hiérarchie qui affectent un même extrait musical. Un musicien aura accès aux niveaux les plus élevés, c'est-à-dire qu'il sera capable de battre la pulsation tous les deux temps ($n+1$) ou tous les quatre temps ($n+2$).

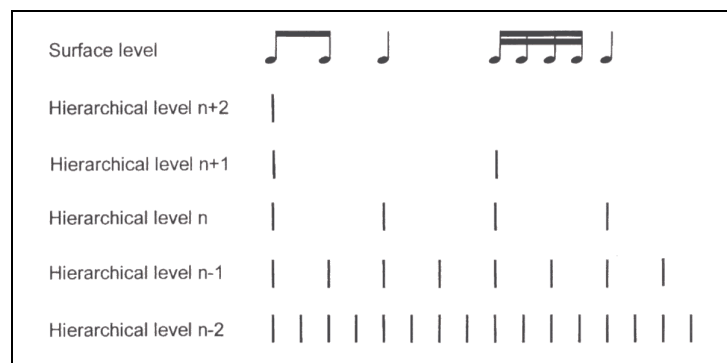


Figure n°1 (emprunté à Drake et al., 2000b : 253).

Il reste cependant un paradoxe au sein d'une organisation temporelle, qui paraît aussi réglée que la mécanique d'un horloge suisse. Alors que l'organisation mesurée de la musique et son organisation temporelle hiérarchique apparaissent sans ambiguïté sur la partition musicale, il n'en va pas de même au niveau de l'interprétation musicale. L'interprète ne peut jouer avec le paramètre de durée, sans risquer de violer le cadre rythmique. Cependant, il semble que l'auditeur soit sensible à des différences minimales de durée entre les notes. Dans une séquence composée de notes de durées semblables, des différences de durée aussi petites que 10 ms peuvent être détectées (Van Noorden, 1975). Les musiciens semblent profiter de cette forte sensibilité à la perturbation des durées chez l'auditeur pour communiquer l'accentuation. S'ils prennent la liberté de marquer l'accentuation par un ajout de durée, cela produira un ralentissement du tempo.

Le musicien virtuose joue avec le cadre rythmique qui lui est imposé, jusqu'aux limites de ce cadre, mais sans jamais le violer ; la musicalité est à ce prix et il semble bien que l'oreille n'apprécie pas les rythmes trop réguliers, qui lui paraissent non naturels :

« Les interprètes contrôlent intentionnellement leurs réactions pour créer des patterns sonores musicaux qui entravent systématiquement les attentes des auditeurs concernant les relations temporelles hiérarchiques ; ils donnent de ce fait une interprétation de la partition » (Jones et Yee, 1994 : 109).

La régularité qui apparaît sur la partition n'existe donc dans la réalité de l'interprétation. Pourtant, nous avons vu qu'elle constituait une réalité psychologique pour l'auditeur, qui se montre capable d'extraire une pulsation et ce, malgré une forte sensibilité aux petites différences de durée. Le terme de « métrique » désigne ce processus d'abstraction :

« La métrique est donc une abstraction, une idéalisation, qui implique l'existence de processus cognitifs complexes conduisant l'auditeur à interpréter comme régulières des successions d'événements sonores qui ne le sont objectivement pas (si ce n'est dans le cas des exécutions musicales par ordinateur) » (Bigand, 1994 : 273).

Ce processus joue un grand rôle, puisqu'il crée des attentes perceptives chez l'auditeur : celui-ci attend le retour des temps forts. Certaines figures rythmiques jouent avec ces attentes : c'est notamment le cas du **contretemps** (temps fort ou partie forte du temps occupé par un silence et partie faible, par une note ; ceci viole l'attente que la note tombe sur la partie forte) et de la **syncope** (note qui commence sur un temps faible et se poursuit grâce à une **liaison rythmique**, sur un temps fort ; il en résulte une tension rythmique qui ne peut se résoudre que sur le temps fort).

Comment l'auditeur réussit-il à restituer la pulsation d'un morceau de musique et à accéder à différents niveaux qui reflètent une organisation temporelle hiérarchique ? Une telle organisation implique en effet une organisation régulière. Jones et Yee (1994 : 109) tentent de résoudre cette énigme en pointant le rôle de l'attention dans la perception des patterns rythmiques. Les auditeurs expérimentés apprennent à extraire certaines propriétés rythmiques de la musique, qu'ils projettent ensuite sur les pièces qu'ils entendent. Une étude de Palmer et Krumhansl (1990) citée par Jones et Yee (1994 : 111) révèle ce pouvoir « organisateur » de l'attention.

Les auteurs font entendre à des auditeurs musiciens et non musiciens une séquence de 9 battements isochrones. Un battement sonde d'une hauteur différente a été intercalé entre les cinq derniers battements. La tâche des auditeurs consiste à évaluer le degré d'intégration de ce battement au sein de la séquence rythmique. Si les auditeurs sont informés qu'ils vont entendre un rythme binaire (2 ou 4 battements), ils jugent que le battement intervenant près du début des groupes métriques binaires est très intégré au rythme de la phrase. Mais, s'ils sont informés qu'ils vont entendre un rythme ternaire (3 ou 6 battements), ils jugent que c'est le battement qui intervient près du début des groupes rythmiques ternaires qui est le plus intégré. Il semble donc que l'attention soit fondamentale, puisqu'elle permet d'orienter la perception du rythme, de prédisposer à faire certains groupements, reposant sur des hiérarchies temporelles. Cet effet de « prédisposition attentionnelle » semble d'ailleurs plus prononcé chez les musiciens, ce qui soulignerait le rôle de l'expérience musicale dans la formation de l'attention. Par ailleurs, les musiciens se montrent capables de compléter des airs folkloriques fondés sur des rythmiques simples (Jones et Boltz, 1989) et cette tâche est beaucoup plus précise lorsque ces airs sont structurés hiérarchiquement.

1.2.1.3. Parole et périodicité sommaire

Même si l'interprétation musicale ne se montre pas parfaitement régulière, elle est donc perçue par l'auditeur comme métrique. Il n'en va pas de même pour l'organisation temporelle de la langue. Il n'est qu'à essayer de battre la pulsation en même temps qu'un auditeur parle pour se rendre compte que la tâche est beaucoup plus difficile que pour la musique (Drake, 2002). Le rythme de la langue paraît plus complexe, mais cette complexité dépend également du type de discours. Le rythme peut être en effet très scandé et régulier, comme dans le cas de la récitation (prières), des sermons, de la poésie, du théâtre, des discours politiques... Mais nous nous arrêterons au rythme de la parole dans la conversation courante. Ce rythme paraît non régulier et il semble fondé sur une sorte de périodicité sommaire :

« L'organisation temporelle de la langue parlée n'est toujours pas bien comprise, mais il semble que, bien qu'elle se fonde sur une sorte de périodicité sommaire (ou « quasi-périodicité »-near-miss-periodicity-, selon Schaffer, 1982) et impose de fortes contraintes à certains aspects de la structure temporelle, elle n'utilise pas le même cadre métrique que celui qu'on trouve en musique » (Clarke, 1989 : 27-28).

Selon Fraisse (1974), le retour périodique d'une différence acoustique tend à être interprété comme un temps fort de la structure métrique. Mais quel est cet élément qui revient « quasi » périodiquement ? Il semble qu'il ne soit pas le même selon les langues et il est possible de classer les langues en fonction du type d'élément accentué qui assure cette périodicité : ainsi, il existerait selon Pike (1945) des langues à rythmicité syllabique (« syllable timed ») et des langues à rythmicité accentuelle (« stress timed »). Cependant, on pense aujourd'hui que les différentes langues, au lieu d'appartenir à l'une de ces catégories de manière nettement délimitée, se répartissent le long d'un continuum allant de la rythmicité syllabique à la rythmicité accentuelle (Dauer, 1987). Ainsi, on peut dire que l'anglais a une tendance à la rythmicité accentuelle car cette langue est rythmée par le retour régulier de l'accent tonique (le fameux « *drum beat* », voir chapitre V, 2.1.1.3.). La durée des groupes temporels entre les accents toniques est donc « invariable ». En fait, il serait plus juste de dire qu'elle est perçue comme invariable par l'auditeur, car on note des différences de durée relativement importantes entre les groupes au niveau acoustique. Les auditeurs attendent donc une régularité rythmique et il semble que les locuteurs jouent de cette attente lorsqu'ils perturbent cette régularité, en insérant des pauses ou des accents là où on ne les attend pas par exemple.

Cette attitude se rapproche beaucoup de celle du musicien lorsque celui-ci accentue certains éléments en augmentant légèrement leur durée. Musique et langue se rejoignent finalement dans le fait qu'elles tirent toutes deux de l'expressivité lorsqu'elles jouent et altèrent la métricité. Cependant, la marge de manœuvre semble plus grande pour la parole, car le locuteur n'a pas besoin d'adhérer à un cadre temporel strict. Il peut donc jouer plus librement avec les différents paramètres pour la réalisation de l'accent, dans la mesure où il respecte certaines règles que lui impose sa langue maternelle (durée pour le français ; hauteur, intensité et durée pour l'anglais). Dans la musique classique, cette marge de liberté est beaucoup plus réduite que dans la langue car l'emploi de l'accent est régi par de fortes contraintes, notamment au niveau de la durée. Les chercheurs ne savent pas encore exactement comment le rythme de la parole est perçu, ni quelles unités marquent les niveaux inférieurs et supérieurs de la structure temporelle. Selon Jones et Yee (1994 : 118), les hiérarchies temporelles sont peut-être plus complexes pour la parole que celles de la musique, et donc plus difficiles à faire apparaître.

1.2.2. Mélodie de la musique et mélodie de la parole

1.2.2.1. Successivité et intervalles

Dans l'Encyclopedia Universalis, la mélodie est définie comme :

« une succession de sons ayant entre eux des rapports d'intervalles et de durée. Telle est la définition la plus large et la plus indéterminée qui puisse en être donnée » (Barraud, 2002).

Cette **succession de sons** correspond au niveau acoustique à l'évolution de la fréquence fondamentale au cours du temps et au niveau de la perception, à l'évolution de la hauteur au cours du temps.

« Comme la musique, le langage est sonore et se déroule dans le temps. Il est intéressant de comparer les emplois, structures et perceptions qui divergent à partir de cette base commune » (Schaeffer, 1966 : 284).

Selon Menezes (1993 : 82), le jeu des oppositions est syntagmatique à la fois pour les traits prosodiques et pour les « traits musicaux » et la séquence temporelle est la condition de leur existence :

« la séquence ne jouerait-elle pas à vrai dire le rôle le plus important aussi bien pour la musique que pour le langage ? ».

Langue et musique se caractérisent donc par leur **dimension temporelle** et leur **caractère séquentiel**. Cette similitude entre la structure des trames temporelles en musique et la structure prosodique du langage est également soulignée par Lerdahl et Jackendoff (1983). Dans la musique et dans la langue, les sons qui composent la ligne mélodique entretiennent entre eux des rapports d'**intervalles**, c'est-à-dire une distance entre leurs hauteurs respectives. Cependant, c'est dans la transition entre les sons que réside « la » différence toujours invoquée entre la langue et la musique : en musique, le passage d'une note à une autre se fait de manière discontinue, alors que dans la langue, il se fait de manière progressive et continue, de telle manière que les hauteurs ne peuvent être clairement délimitées et isolées par l'oreille.

D'après Scripture (1902 : 472), **Aristoxène**, le premier musicologue européen avait déjà signalé cette différence :

« It was pointed by Aristoxenus that the difference between the tones in song and those in speech lies in the fact that the voice in singing proceeds by jumps from one note to another, while in speech it continually slides up and down ».

Passy (1890 : 68) cite deux exemples de transcription musicale, le premier (à gauche de la figure n°2) d'un extrait musical et le second (à droite de la figure n°2), d'un extrait de parole. Ce faisant, il tente de démontrer que la transcription musicale se montre inapte à transcrire la mélodie de la parole, car elle ne permet pas de représenter les glissements qui la caractérisent.



Figure n°2 : transcription musicale de l'énoncé « Viens tu ? » dans une version musicale et dans la parole (emprunté à Léon, 1992 : 124).

1.2.2.2. Le fait musical au travers du filtre de la transcription

Cette différence pousse Crystal à affirmer que l'analogie entre langue et musique n'est pas vraiment fondée (1987 : 173). Mais, avant d'en arriver à une telle conclusion, il est nécessaire de prendre certaines précautions et en premier lieu, en ce qui concerne le système de notation musical. La plupart des auteurs oublient que ce système n'est pas la musique, mais bel et bien une **représentation** de la musique. C'est comme si on décrivait la langue parlée en se référant à l'écrit. Et comme toute représentation, elle obéit à un ensemble de critères qui ne sont pas objectifs. En effet, la notation musicale telle que nous la connaissons (voir figure n°17) est profondément ancrée dans une tradition musicale spécifique, la musique occidentale classique. Celle-ci a été fortement marquée par la pratique de l'écriture, à tel point que composition et écriture ont fini par se confondre : le compositeur dispose en effet d'un système d'écriture tellement puissant qu'il lui permet d'agencer des partitions d'une telle complexité qu'aucun musicien ne saurait les interpréter. L'étymologie du mot composition renvoie d'ailleurs à une idée d'arrangement graphique plutôt qu'à un arrangement des sons en fonction du temps. Il est donc dangereux de penser que le système de codification actuel, qui n'a guère évolué depuis le milieu du XVIIIe siècle et qui tend aujourd'hui à s'universaliser, soit le mieux approprié au langage musical, d'autant plus que ce système tend à privilégier la fixation de la hauteur et de la durée.

Au travers du filtre de la notation classique, on en arrive donc à la conclusion, trop simpliste à notre goût, que la musique bondit, alors que la langue, tel un skieur élégant, glisse d'une note à l'autre.

Et bien souvent, cette constatation suffit à enterrer définitivement toute discussion sur les similitudes de forme entre langue et musique. Les choses sont souvent bien moins simples qu'elles y paraissent au premier abord et il suffit pour s'en convaincre de nous projeter, pour un bref voyage, dans l'histoire de la notation musicale occidentale. Et de découvrir qu'à son origine était... la langue.

Ainsi, Massin et Massin (1983 : 126) nous apprennent qu'au II^e siècle de notre ère, Aristophane de Byzance avait mis au point un système de notation pour la langue grecque, où l'accent aigu représentait l'élévation de la voix, l'accent circonflexe, une faible élévation, le point, la chute puis l'arrêt de la voix à la fin d'une phrase. D'ailleurs, d'après son étymologie, le terme même d'accent renvoie à l'origine de la constitution mélodique⁴. Ainsi, les symboles graphiques que nous utilisons aujourd'hui pour transcrire « é, è et ê » étaient utilisés à l'origine pour représenter les modulations de la voix. Au VII^e siècle est apparue une notation utilisée pour transcrire la prosodie des textes de la Bible, et notamment celle du grec et du latin : la notation neumatique. Lors de la lecture des Écritures, la mélodie était plus proche de la déclamation que du chant et les neumes représentaient l'évolution de la ligne mélodique sous forme de « gestes » stylisés graphiquement et probablement inspirés des gestes du chef de chœur dans la cantillation byzantine.

« Cette mise en relief « spatialisée » des inflexions vocales dans des formes de notation neumatique est d'ailleurs présente dans d'autres traditions que la nôtre (cf. la psalmodie bouddhique, le nô japonais, etc.) » (Massin et Massin, 1983 : 126).

⁴ « Du latin *accentus*, élévation de la voix sur une syllabe, puis son d'un instrument » (Dictionnaire étymologique Larousse).

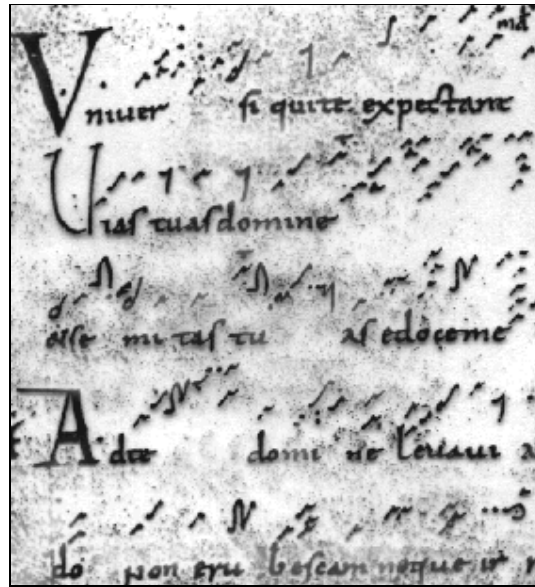


Figure n°3 : Fragment de neumes du manuscrit de Laon 239 - Metz vers 930 (emprunté à Schuyesmans, 2002⁵).

Les neumes sont inscrits au dessus du texte et chacun d'entre eux correspond à une syllabe et couvre une ou plusieurs notes (figure n°3). Par exemple, nous pouvons voir sur la figure n°4 que « torculus »⁶ représente un mouvement mélodique circonflexe bas-haut-médium.

SIGNES NEUMATIQUES à partir du VIIe	NOTATION CARRÉE à partir du XIIIe	NOTATION MODERNE à partir du XVe
virga		
punctum		
clivis		
podatus		
torculus		
porrectus		

Figure n°4: évolution des symboles utilisés pour la transcription musicale : signes neumatiques à partir du VIIe, notation carrée à partir du XIIIe et notation moderne à partir du XVe (emprunté à Massin et Massin, 1983 : 129).

⁵ Schuyesmans (2002), adresse de la page : http://www.schuyesmans.be/gregoriaans/FR/FRmu_05.htm

⁶ « Le **Torculus** se compose de trois notes, la médiane étant toujours la plus élevée. Les intervalles entre la première note et la plus haute et entre celle-ci et la suivante peuvent varier » (Schuyesmans, 2002).

Cette spatialisation des hauteurs permet de symboliser la hauteur relative des notes et donc, de transcrire les intervalles. Elle a notamment permis de visualiser ce qui relevait jusqu'alors du « perçu » et donc d'établir une correspondance entre aigu et haut, grave et bas. Mais dans ce genre de notation, les hauteurs des sons ne sont pas encore fixées, en raison de l'absence d'une « hauteur référence ». Ce point de référence apparaît avec la notation diastématique (de « *diastéma* », intervalle), dans laquelle une ligne horizontale indique la hauteur de la note fa. A partir de cette ligne qui est l'ancêtre de notre portée actuelle, il est donc possible de déterminer la hauteur des autres sons (figure n°5). Dans le système de notation mis au point par Guido d'Arezzo (995-1050) en 1025, le nombre de ces lignes « référence » va se multiplier facilitant ainsi la lecture ; ces lignes sont espacées par des tierces et les lignes au-dessus desquelles se trouvent un demi-ton sont colorées (do et fa).

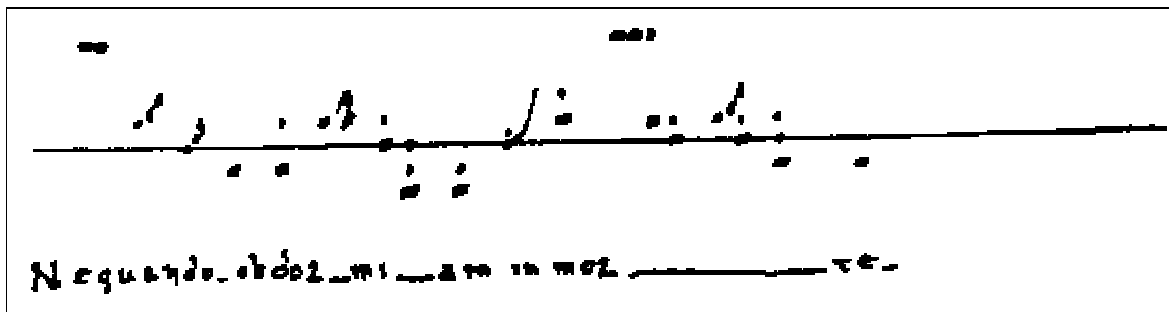


Figure n°5 : notation diastématique, manuscrit Paris B.N. 903 de Saint Yrieix (emprunté à Schuyesmans, 2002)⁵

A la fin du IXe siècle, la notation alphabétique s'associe aux neumes, de manière à fixer les hauteurs (figure n°6) encore plus solidement. La notation neumatique va ensuite peu à peu disparaître pour être remplacée par la notation carrée utilisée par les troubadours à partir du XIIe siècle (figure n°4). Celle-ci va à son tour céder progressivement la place à la notation mesurée au cours du XIIIe siècle, qui permet de fixer avec une grande précision les rapports de durée entre les notes, précision nécessaire à la coordination des différentes voix dans la polyphonie.

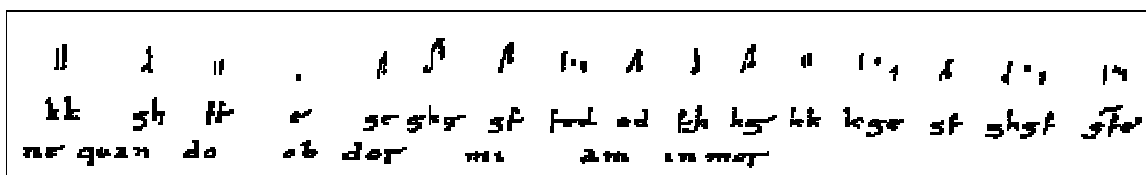


Figure n°6 : double notation musicale, alphabétique et neumatique - Manuscrit de Montpellier H 115 (emprunté à Schuyesmans, 2002)⁶

L'avènement de l'imprimerie au cours du XVe siècle va contribuer à la standardisation de la notation musicale (figure n°4) :

« les procédés de gravure influencent désormais la forme des figures de notes : de carrée ou en losange, elle devient arrondie » (Massin et Massin, 1983 : 133).

Elle se stabilise au cours des XVIe et XVIIe siècle, avec l'apparition de la barre de mesure. Pendant l'époque baroque, la notation garde encore beaucoup de flexibilité du fait de la place centrale accordée à l'interprète :

« jusqu'au XVIIIe siècle, la flexibilité de la notation et, consécutivement, du jeu musical, ne laisse aucun doute ; l'importance des ornements, d'un mode d'écriture comme la basse chiffrée, d'un mode d'expression vocale comme le récitatif nous le confirme incontestablement ; de même, cette surprenante déclaration de Couperin : « Nous écrivons une chose et en jouons une autre » » (Massin et Massin, 1983 : 136).

Cependant, peu à peu, c'est la fidélité au texte qui va s'imposer et la notation se caractérisera par une grande rigidité de la transcription de la hauteur et de la durée. C'est cette transcription musicale « classique » que nous utilisons encore aujourd'hui et qui est enseignée dans les écoles de musique et les conservatoires. Elle n'a plus guère évolué depuis le milieu du XVIIIe siècle. Selon Leipp (1971 : 82) :

« cette notation s'avère extrêmement efficace lorsqu'il s'agit d'instruments traditionnels : il est vraisemblable qu'on ne puisse plus guère l'améliorer. Elle contient la quantité d'information maximale que l'exécutant puisse intégrer pour réaliser en temps réel le programme qui lui est imparté. Elle possède le caractère de ce qui est classique : on ne peut rien en supprimer et l'adjonction d'autres signes la rend confuse, illisible, impraticable ».

La hauteur y est tellement fixée qu'un compositeur comme Bartók a dû introduire un nouveau symbole musical pour représenter les effets de glissement que l'on trouve dans les mélodies populaires (figure n°7) :

« Dans les mélodies populaires [...], il y a beaucoup de sons étrangers, certains glissements de voix, des sons dont la hauteur ne peut être exactement précisée » (cité par Massin et Massin, 1983 : 143).



Figure n°7 : notation du glissando en musique (emprunté à Massin et Massin, 1983 : 143).

1.2.2.3. Des précautions à prendre

Dès lors, nous voyons bien qu'il est difficile de tracer une frontière nette entre la langue et la musique et ce petit détour par l'histoire de la notation musicale montre qu'il faut se munir de beaucoup de précautions quand on essaie de les comparer. Comme le souligne Schaeffer (1966 : 288), il faut se méfier en premier lieu de l'écriture qu'on retrouve dans la note musicale « *en nous la faisant considérer comme un signe préexistant à sa réalisation* ». La transcription « classique » n'est pas la musique, nous l'avons déjà souligné, mais de plus, elle ne permet pas de transcrire toutes les musiques. N'oublions jamais qu'elle se réfère à une **tradition musicale spécifique**, fortement ancrée au niveau historique et géographique. Et il n'est pas envisageable d'aborder sérieusement les similitudes et les différences entre la langue et la musique, sans en tenir compte. Nous voyons peu à peu se dessiner le vide qui existe en la matière et une véritable comparaison entre langue et musique nécessiterait une étude musicologique et ethnomusicologique approfondie, couplée à une étude linguistique et ethnolinguistique comparative. Cette entreprise dépasse largement le cadre de ce travail et nous nous contenterons seulement de préciser les précautions à prendre. La première d'entre elles serait donc de toujours définir la tradition musicale dans laquelle on se place.

1.2.2.4. Tension tonale dans la musique et dans la langue

Selon Menezes (1993 : 84) :

« le fait d'intoner, c'est-à-dire d'être-dans-le-ton se présente, en tant que loi de l'intonation, comme expression de la pensée et de l'émotion aussi bien dans le langage verbal que dans le langage musical. La distinction entre les deux s'avère à ce moment-là impossible à être réalisée face à la « tension tonale » commune à ces deux sphères de la communication humaine ».

Cette tension tonale que nous retrouvons à la fois dans la langue et la musique donne vie à la mélodie. En effet, la mélodie n'est pas une simple succession de sons dans le temps, chacun ayant le même poids que le précédent et que le suivant et la notion de successivité n'implique pas la notion d'égalité : les sons ne sont pas égaux entre eux et certains ont plus d'importance que d'autres. Plus exactement, ils ont plus de poids au niveau perceptif.

Au niveau psychologique, nous sommes sensibles à cette hiérarchie et nous attendons les sons « importants » avec plus d’impatience : cette attente fait naître en nous une sorte de tension qui se résout après que ces sons se soient produits et lorsque l’attente est contrariée, il se crée un effet de surprise. Ces **schémas de tension et de détente** structurent notre perception de la mélodie et sont en grande partie responsables des émotions que nous ressentons à son écoute. Dans la musique tonale, cette hiérarchie porte un nom, la **hiérarchie intra-tonale** :

« à l’intérieur des tonalités, la théorie musicale définit des hiérarchies : la note qui donne son nom à la tonalité (la tonique) est hiérarchiquement la plus importante » (Bigand, 1994 : 254) ;

elle est considérée comme la note la plus stable et les mélodies finissent typiquement sur cette note.

« Puis vient la cinquième note de la gamme, la dominante qui, comme son nom l’indique, domine les autres » (Bigand, 1994 : 254).

De la mélodie émerge l’harmonie puisque un accord de dominante (V^e degré) crée des attentes de résolution sur l’accord de tonique (I^e degré). Du point de vue mélodique, les notes de l’accord de tonique, c’est-à-dire les première, troisième et cinquième notes (voir figure n°8) sont intériorisées mentalement comme les notes les plus stables (Dowling et Harwood, 1986).

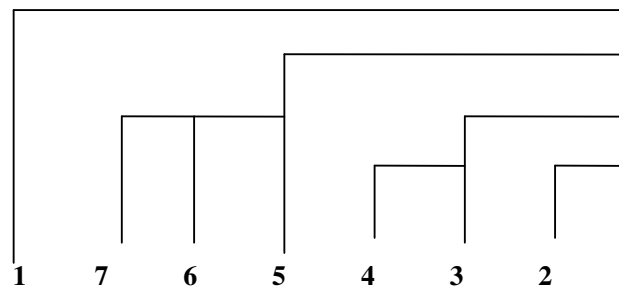


Figure n° 8 : le système de hiérarchie intra-tonale dans la mélodie
(d’après Bigand, 1994 : 254).

Par exemple, une mélodie qui aboutit sur la septième note tendra à se résoudre sur une note structurellement plus importante, c’est-à-dire sur la cinquième. Tant que cette résolution ne sera pas accomplie, nous aurons l’impression perceptive de non aboutissement. Arrivé sur la cinquième note, la tension sera toujours présente, mais moins intense. Elle s’apaisera totalement sur la première note, point d’aboutissement de la mélodie.

Les notes peu importantes constituent des points de passage sur lesquels la mélodie ne s'arrête pas. Pour cette raison, on parle de notes de passage. Elles sont notamment utilisées dans les ornements, tels que les appoggiatures, les mordants, les gruppetto et les trilles où la mélodie s'attarde avant de se fixer sur des notes structurellement plus importantes. Elles instaurent de fortes tensions qui seront résolues par l'arrivée de notes plus stables.

Même les adultes qui n'ont pas de formation musicale ont une connaissance tacite des différentes fonctions des notes de la gamme et cette connaissance implique qu'ils soient capables d'attribuer un poids spécifique à chacune des notes au sein d'une mélodie tonale. Ces rapports entre les notes sont intériorisés chez les auditeurs adultes : dans une expérience très connue, Krumhansl et al. (1979) présentent à des sujets musiciens et non musiciens les sept notes de la gamme de Do majeur, puis une des douze notes de la gamme chromatique. La tâche consiste à évaluer sur une échelle de sept points le degré de cohésion d'une note isolée avec la gamme. Pour tous les auditeurs, les notes jugées comme s'accordant le moins bien avec la séquence sont les notes extérieures à la tonalité de Do majeur. Par contre, la tonique do est jugée comme la note la plus intégrée à la séquence. Chez les sujets les plus musiciens, la dominante, puis la tierce sont les notes jugées les mieux intégrées après la tonique. Ces rapports de hiérarchie semblent intériorisés de manière très précoce : en effet, dès 4 mois et demi, les bébés préfèrent des extraits tirés des menuets de Mozart segmentés correctement (Juszyk, 1989), car ils écoutent plus longtemps les passages interrompus par des pauses situées aux limites des phrases plutôt qu'à l'intérieur des phrases. Une phrase non résolue du point de vue tonal ne peut pas être considérée comme achevée. Bien sûr, d'autres paramètres peuvent expliquer cette préférence pour les « bonnes limites », tels que la cohérence temporelle et rythmique de la phrase, mais du point de vue perceptif, la hiérarchie tonale semble également jouer un grand rôle. Outre le fait que cette hiérarchie entre les notes permette à l'auditeur de structurer sa perception de la mélodie et par conséquent, de l'harmonie, elle lui donne la faculté de prédire les notes qui vont se produire. Grâce à l'intériorisation de ces schémas, nous sommes en effet capables de compléter une ligne mélodique s'il manque une note. Nous sommes également capables de prédire la fin d'une phrase musicale et ce, même pour un air que l'on n'a jamais entendu auparavant. Selon Jackendoff (1991), l'auditeur anticipe ce qui est à venir (« *prospective hearing* ») en fonction de ce qu'il a déjà perçu (« *retrospective hearing* »).

La musique dodécaphonique a remis en cause ce type de hiérarchie tonale en proposant des gammes où chaque demi-ton possède un poids équivalent à celui des autres demi-tons. Mais on peut se demander si au niveau perceptif, l'auditeur de ce type de musique ne continue pas à analyser les sons de manière hiérarchique, en raison de son acculturation au système tonal et ce, depuis l'enfance (voir chapitre II, 3.1.3. et chapitre III, 2). Pour Lerdahl (1985), l'absence de hiérarchie dans les musiques contemporaines non tonales rend plus difficile la structuration perceptive et la représentation intériorisée des œuvres. Selon lui,

« les systèmes compositionnels actuels tendent à être inaccessibles, même à des auditeurs expérimentés » (1985 : 110).

Mais, cette inaccessibilité ne proviendrait-elle pas tout simplement du manque d'« acculturation » aux musiques non tonales ? Après tout, la musique tonale résulte d'une tradition musicale spécifique, la musique occidentale classique (et notamment, la musique tonale où la mélodie est soumise à l'harmonie, contrairement à la musique modale grégorienne). Cette tradition s'est construite au cours des siècles, jusqu'à atteindre son apogée pendant la première moitié du XIX^e siècle. Elle a ensuite évolué naturellement vers l'atonalité, à cause de la saturation du système tonal (voir 2.2.2.). Il est vrai qu'aujourd'hui, la musique diffusée est majoritairement tonale, notamment dans les musiques de variété et qu'il est donc impossible de se soustraire à ses influences. Elle est en quelque sorte notre « *langue maternelle musicale* » (Viret, 1987 : 7). Mais, cela ne veut pas pour autant dire que le système tonal soit plus naturel que les autres systèmes, comme certains auteurs l'écrivent. Nous en prendrons pour preuve l'acculturation avec des systèmes musicaux qui nous paraissent beaucoup plus complexes, tels que les gamelans balinaï (voir chapitre II, 3.1.3.3.).

Existe-t-il un système de hiérarchie similaire dans la langue ? Il existe dans la parole des points « cibles » hauts et des points cibles bas, étrangement fixes chez un même locuteur. Si le locuteur s'interrompt en plein milieu d'une phase, la courbe mélodique de l'énoncé se trouve comme suspendue et attend la résolution, que ce soit sur un point cible haut comme dans la question ou sur un point cible bas, comme dans l'affirmation. Cette attente crée un fort sentiment de tension chez l'auditeur. Par exemple, « Je vais au marché ↑ » [...] « demain ↓ ».

Cet effet de tension peut être utilisé en rhétorique, en particulier en insérant des pauses là où on ne les attend pas (cf. les discours du général De Gaulle) : le locuteur réussit à « suspendre » l'auditoire à ses lèvres et il est fort probable que la mélodie soit déterminante dans ce phénomène.

Ainsi, les facultés de prédiction qui se manifestent pour la musique se retrouvent également dans la langue : nous sommes tous capables de compléter des phrases qui ne sont pas terminées (nous finissons la phrase de notre interlocuteur ou nous parlons en même temps). Dans cette prédiction, nous nous basons sur des indices syntaxiques et sémantiques, mais avant tout mélodiques, comme le prouvent les travaux de Grosjean à ce sujet. Selon Grosjean (1983), la mélodie permet en effet de faire ressortir un certain point de l'énoncé appelé « dernier mot potentiel » (DMP), à partir duquel l'auditeur pourrait prédire la fin de la phrase. Les paramètres utilisés dans cette mise en relief, tels que la rupture de hauteur pour la Fo et l'allongement de la durée du DMP, joueraient le rôle de clés prosodiques. Pour savoir si l'information sur la longueur de la phrase en anglais est disponible uniquement sur le DMP, Grosjean et Hirt (1996) ont mené une série de quatre expérimentations. Bien que la syntaxe et la sémantique soient utilisées dans la première partie de la phrase, il arrive un moment où elles n'apportent plus d'information sur la continuation potentielle de la phrase. Cette information, seule la prosodie est en mesure de la donner. C'est ainsi que les sujets réussissent à prédire trois types de phrase lors de la seule écoute du DMP : ils peuvent affirmer que la phrase est terminée, qu'il reste encore trois mots ou six mots.

Que se passe-t-il avec le français, une langue dont l'organisation prosodique est très différente de celle de l'anglais ? Là où il existe trois choix possibles pour l'anglais (énoncé terminé ou trois mots ou six mots et plus), il n'en reste plus que deux pour le français (énoncé terminé ou énoncé qui se continue). Les anglais et les français réagissent de manière différente pendant les tâches de prédiction car l'information donnée dans leur langue respective est différente, les clés prosodiques ne les informant pas de la même manière (cf. chapitre V, partie 2.). Des sujets manifestant des lésions de l'hémisphère droit rencontrent de graves difficultés dans l'acte de prédiction, surestimant les phrases se terminant sur le DMP et sous-estimant celles qui ne se terminent pas sur le DMP (cf. chapitre VII, 3, discussion sur l'hémisphère droit, l'« hémisphère musical »).

La mélodie semble donc jouer un rôle déterminant dans le traitement des énoncés, puisqu'elle fournit des informations sur leur gabarit temporel. Il nous est arrivé, en travaillant sur un corpus de bébés de 9 à 48 mois (Dodane et al., 1999), de rencontrer des difficultés à délimiter les différents énoncés. Selon les critères adoptés dans cette étude, nous posons une limite lorsque la pause dépassait une certaine durée (400 ms). Mais, il arrivait de temps à autre que, malgré une pause de grande longueur, nous ayons l'« intuition » que l'énoncé n'était pas terminé. Et seule la mélodie pouvait nous fournir cette information, en assurant la cohésion d'un énoncé dont les parties étaient éloignées dans le temps. En outre, elle contient des indices sur les frontières linguistiques importantes, qui permettent de segmenter les énoncés. Ainsi, de manière frappante, les bébés font preuve des mêmes capacités de segmentation que pour la musique. Dans une étude que nous avons déjà citée, Jusczyk (1989) a montré que des bébés anglophones de 4 mois ½ préféraient des passages interrompus aux limites des phrases plutôt qu'à l'intérieur. Ces résultats ont été confirmés chez des bébés de 6 mois par Hirsh-Pasek et al. (1987) et plus récemment par Nazzi et al. (2000).

1.2.2.5. Différenciation mélodie langue / musique

Comment mettre à jour les caractéristiques spécifiques de la mélodie de la langue par rapport à celle de la musique ? Il paraît difficile de le faire sans effectuer une comparaison acoustique et auditive des deux types de mélodie. Or, il existe un cas de figure qui pourrait nous faciliter la tâche : certains musiciens ont essayé d'imiter l'intonation des langues naturelles avec leur instrument (voir 2.2.3.4.). Jimi Hendrix est peut être le plus célèbre d'entre eux : dans « *Still Raining, Still Dreaming* » (1968), il débute son solo de guitare par une imitation de l'intonation humaine qui dure quelques secondes (2021 ms), puis à la phrase suivante, la mélodie « redevient » musique (**extrait musical n°1**). Puisqu'une guitare est éloignée de la voix humaine, le musicien est obligé de « forcer » le trait, tout comme le ferait un imitateur lorsqu'il imite la voix d'un homme politique ou un dessinateur lorsqu'il réalise une caricature. Pour que son imitation soit réussie, il lui faut exagérer les traits caractéristiques du modèle. C'est à cette condition que l'auditeur pourra reconnaître l'extrait en tant qu'« imitation » de l'intonation. Un tel extrait se montre donc précieux pour nous aider à révéler les différences entre l'intonation et la mélodie, car ces différences y seront forcément exagérées.

Les deux phrases ont d'abord été transcrites sur portée musicale, de manière à visualiser leur rythme aussi bien que leurs notes (figures n°9 et n°11). Leur ligne mélodique a été resynthétisée grâce au logiciel Praat afin d'isoler la guitare des instruments qui l'accompagnent (basse et batterie), ce qui pourrait perturber la détection de la Fo. Une première analyse de leur courbe mélodique nous révèle que la phrase « musicale » (figure n°10 ; **extrait musical n°2**) contient plus de glissements que la phrase « parlée » (figure n°12 ; **extrait musical n°3**) et que les hauteurs semblent moins fixées. Les instruments à corde et notamment la guitare électrique (effets obtenus notamment avec la pédale wah wah dans cet extrait) permettent en effet des effets de glissando, qui sont beaucoup plus difficiles à obtenir avec des instruments « à clés » dont les hauteurs sont fixées (flûtes traversières, saxophones, clarinettes...). Hendrix joue de ces effets de manière beaucoup plus marquée pour la musique que pour l'intonation : il commence sa phrase musicale par un glissando montant de 6 demi-tons (216-336 Hz) qui aboutit au fa³, suivi d'un vibrato centré autour du la³, très visible sur la représentation mélodique (indiqué par une flèche sur la figure n°10). On observe également des glissements pour la phrase parlée (voir figure n°12), notamment une chute vertigineuse de 641 Hz à 97 Hz (soit 33 demi-tons !) au début. Mais, les hauteurs suivantes sont plus fixes, surtout à la fin de la phrase. On voit donc bien que ces glissements ne sont pas l'apanage de la parole et qu'ils sont aussi difficiles à transcrire pour la musique que pour la langue, comme nous l'avons d'ailleurs expérimenté ici (figure n°9 et n°11). Passy reprochait à la notation musicale de se montrer inapte à transcrire la mélodie de la parole, mais il aurait dû préciser qu'il parlait de la notation musicale classique. Et c'est justement le filtre de cette notation qui lui fit supposer que la caractéristique essentielle de la langue par rapport à la musique était de glisser d'une hauteur à une autre. Cependant, si la musique contient plus de glissements dans cet extrait, d'où nous vient alors ce sentiment de « fixité » des hauteurs à son écoute ?

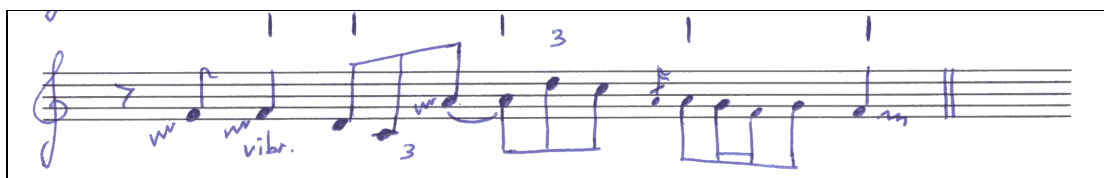


Figure n°9 : transcription musicale de l'extrait de « Still Raining, Still Dreaming » où Hendrix joue « normalement ».

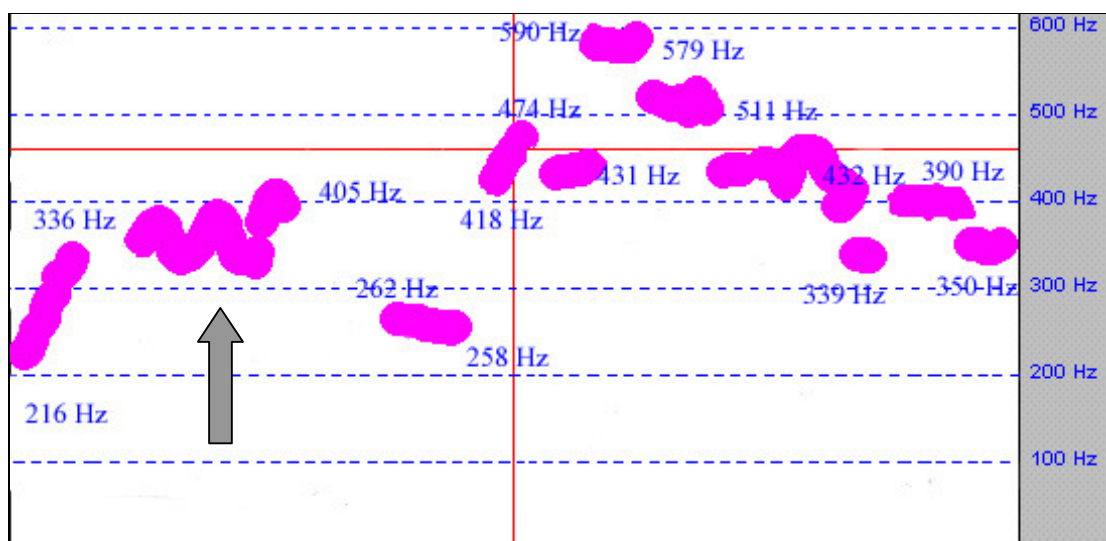


Figure n°10 : représentation de l'évolution de la Fo obtenue après resynthèse avec le logiciel Praat.

La mélodie de la phrase musicale est structurée autour de l'accord parfait Fa M, composé des trois notes structurellement les plus importantes de cet accord : fa (Tonique, I), do (Dominante, V) et la (Médiane, III). Après un glissando qui aboutit sur le fa³, la mélodie reste fixée sur cette note, qui possède le poids le plus important dans la tonalité (tonique). Cette note est allongée (croche liée à une noire, soit 1 temps 1/2) et placée sur un temps fort, qui renforce son poids. Elle est suivie d'un triolet dont la dernière note est le la³, c'est-à-dire la médiane. Grâce à la liaison rythmique, cette note est allongée et positionnée sur la partie forte du temps suivant. La mélodie passe ensuite par le ré (VII), puis par do (V), revient sur le la (III), notes d'ornement, puis fa (I). Elle pourrait s'arrêter sur cette note, mais retarde la résolution en incluant une nouvelle note de passage. La mélodie aboutit sur le fa³, placé sur un temps fort et allongé (noire). Dans cet extrait, les notes tonales importantes occupent donc une position forte (partie forte du temps ou temps fort dans la mesure). Cet « ancrage tonal » est renforcé par un allongement important. Ainsi, la tonique fa (I) et la médiane la (III) sont au minimum deux fois plus longues que les autres notes (voir tableau n°2). Les plus longues valeurs correspondent donc aux « bonnes » notes, c'est-à-dire aux notes tonales importantes. **La mélodie est animée, pour ne pas dire « rythmée » par l'attente et la venue des trois personnages principaux que sont dans un ordre hiérarchiquement décroissant la tonique, la dominante et la médiane.** La structuration rythmique, en réservant aux notes tonales les positions fortes, assure une forte cohésion avec la mélodie. Les « bons intervalles » sont ainsi fixés aussi bien au niveau tonal que rythmique. Par ailleurs, on observe une continuité de la mélodie dans le sens où le plus grand intervalle entre deux notes consécutives est une 6te M (do³-la³).

Note	Fa	Fa	Ré	Do	La	Ré	Do	La	La	Sol	Fa	Sol	Fa
Durée en ms	330	252	172	199	335	160	172	67	301	58	80	243	437

Tableau n°2 : durée des différentes notes composant la mélodie de la phrase « musicale » (en ms).

La mélodie de la phrase « parlée » commence par un mi^4 pour chuter très rapidement (au cours d'un triolet) jusqu'à un autre mi , le mi^2 , deux octaves plus bas. Cette dernière note est allongée (noire) et suivie d'une pause (demi-soupir), qui délimite une première unité. La mélodie, qui reprend au si^3 , effectue ensuite une série de sauts d'octave très rapides (au cours d'un triolet) : elle chute d'une octave (pour les correspondances des intervalles musicaux en demi-tons, voir p : X) sur un autre si (si^2), puis remonte d'une octave pour atteindre un do (do^4), puis un $\text{do}\sharp$ ($\text{do}\sharp^4$) et chute à nouveau de plus d'une octave jusqu'au sol (sol^2), qui est allongé et suivi d'une pause. Cette dernière note marque la fin d'une seconde unité et la fin de la phrase « parlée ». Les deux pauses servent donc à délimiter deux groupes au sein de la phrase.

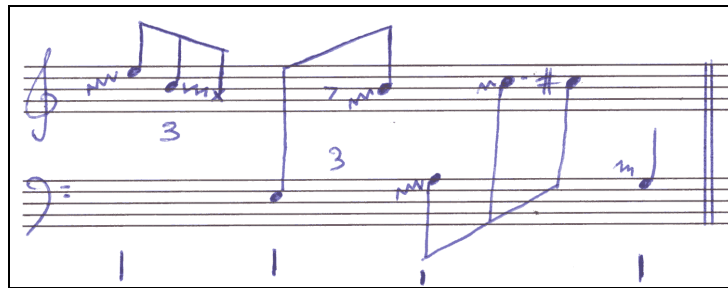


Figure n°11 : transcription musicale de l'extrait de « Still Raining, Still Dreaming » où Jimi Hendrix fait « parler » sa guitare.

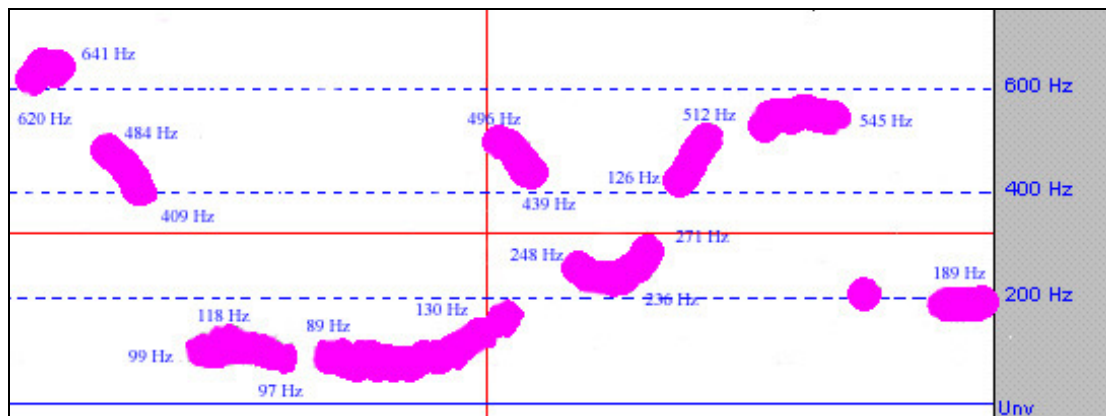


Figure n°12 : représentation de l'évolution de la Fo obtenue après resynthèse avec le logiciel Praat.

A l'intérieur de chaque groupe, la durée des différentes notes est sensiblement égale, ce qui donne à l'auditeur l'impression qu'elles s'écoulent régulièrement (tableau n°3). Il n'y a que la dernière note du deuxième groupe qui soit plus allongée (277 ms), mais le rapport avec les autres notes n'atteint jamais le double.

Note	Mi	Si	La	Mi	/	Si	Si	Do	Do#	Mi
Durée en ms	213	162	175	172	255	172	232	191	178	277

Tableau n°3 : durée des différentes notes composant la mélodie de la phrase « parlée » (en ms).

Aucune note à l'intérieur des groupes n'est mise en relief par un allongement de la durée, comme les notes tonales importantes pour la musique. La mélodie du premier groupe commence et se termine par un mi ; celle du second groupe commence par un si, qui est répété et aboutit sur un sol. Autrement dit, les notes importantes sont mi-sol-si, soit un accord parfait mineur. Nous n'irons pas plus loin, mais ces notes ne sont pas mises en relief aussi fortement que ne l'étaient les notes de l'accord parfait pour la musique.

Par ailleurs, on peut observer une grande discontinuité dans la ligne mélodique ; la mélodie franchit souvent l'intervalle d'octave entre notes successives (plus grand intervalle : 8ve + 5te A, soit sol2-do#4), alors que pour la phrase musicale les mouvements étaient conjoints. Selon Vos et Troost (1989), la plupart des types de mélodie utilisent typiquement des changements de hauteur de l'ordre de 1/12 à 1/6 d'octave et la subdivision de l'octave à travers les cultures tend à se placer dans cette étendue. Dans la parole, les variations de hauteur tendent à dépasser la moitié de l'octave et les relations entre les intervalles ne sont pas critiques (Patel et al., 1998). Outre leur grande étendue, ces changements de hauteur se montrent continuels à l'intérieur des groupes : il n'y a jamais d'arrêt sur une note particulière qui serait allongée, et de plus, on observe une succession rapide de montées et de descentes.

Ces caractéristiques semblent être les plus importantes pour différencier la mélodie de la langue et celle de la musique dans cet extrait : les écarts entre les notes sont trop grands et les notes importantes du point de vue tonal ne sont pas marquées au niveau du rythme. Plus la mélodie est rangée et plus les « bonnes » notes sont mises en relief, plus la mélodie est perçue comme musicale, plus les intervalles sont fixés (car on laisse à l'oreille le temps de les fixer ; avec la parole, la succession de notes est trop rapide). La cohésion mélodique et rythmique de la phrase musicale assure la « fixation » des intervalles et donc de la tonalité. C'est donc cette propriété qui différencie l'intonation de la musique, et non pas les glissements d'une note à l'autre.

1.2.3. A mi chemin entre musique et langue : les clichés mélodiques

Les clichés mélodiques (on parle également d'« intonations clichés » ou, en anglais, de « stylised tunes ») sont constitués par une séquence de mots dont la courbe mélodique est fortement contrainte, voire figée.

Ces clichés sont communs aux membres d'une même communauté et utilisés dans des situations bien spécifiques (Di Cristo, 1998 : 215), telles que les salutations par exemple (« *Coucou* » et « *Bonjour* » ; « *hello* » ou « *bye bye* ») ou le fait de répondre au téléphone (« *allô* » ; « *moshi moshi* » en japonais). On les retrouve en abondance dans le discours des commissaires-priseurs, des vendeurs et notamment sur les marchés, ... et ils sont très courants dans la parole de l'enfant, surtout dans la raillerie (« *Na-na-na-na-nère* »). Ils trahissent la langue maternelle des locuteurs qui s'expriment dans une langue étrangère et contribuent ainsi fortement à l'accent d'intonation (Fonagy et al., 1976).

Les clichés mélodiques sont très intéressants car ils révèlent la proximité de la langue et de la musique : ils se situent dans une position intermédiaire entre la parole et le chant (Hirst, 1998 : 71). Il est donc particulièrement pertinent de mettre à jour ce qui contribue à faire d'un cliché un motif musical, afin de pouvoir commencer à esquisser une frontière entre la langue et la musique.

Les clichés mélodiques se caractérisent par une stabilité de hauteur sur chaque syllabe (régularité micro-prosodique) :

« instead of consisting of a continuous sequence of movements from one target point to the next as in normal speech, the contour is produced as a sequence of static level tones » (Hirst, 1998 : 71).

Cette régularité contribue à donner au cliché son aspect chantonné (Fonagy et al., 1983 : 153). Elle permet de fixer les intervalles intersyllabiques et assure ainsi une régularité au niveau macro-prosodique. La transcription musicale est particulièrement apte à schématiser les clichés mélodiques (Léon, 1971 : 124), car elle fait apparaître cette régularité et cette stabilité au niveau de la distribution des hauteurs. Selon Fonagy :

« la mélodicité des énonciations constitue, à côté de la hauteur et du temps, une troisième dimension de l'espace imaginaire où évolue la mélodie » (Fonagy, 1983 : 155).

Dans l'exemple du cliché « *na-na-na-na-nère* » (voir transcription musicale, figure n°13), la ligne mélodique est tellement figée qu'elle peut être représentée par une série de notes musicales : sol-sol-mi-la-sol-mi, qui correspond à la succession d'intervalles : tierce mineure (sol-mi), quarte juste (mi-la), seconde majeure (la-sol) et enfin tierce mineure (sol-mi).

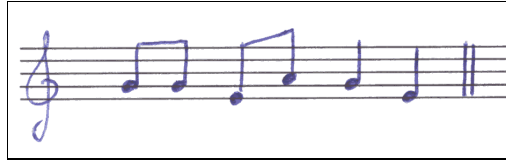


Figure n°13 : transcription musicale du cliché mélodique « na-na-na-na-nère »

C'est cette succession d'intervalles qui permet à l'auditeur de mémoriser et de reconnaître l'énoncé « *na-na-na-na-nè-re* », plus que le contenu segmental qui n'a d'ailleurs aucun sens. En anglais, on retrouve le même cliché avec le même contour mélodique, mais des syllabes légèrement différentes, « *Na-na-na-na-na-euh* ». Le pouvoir évocateur des clichés est souvent utilisé par les musiciens, notamment dans les musiques improvisées, car il suffit de reproduire le contour du cliché pour que les auditeurs le reconnaissent immédiatement. Dans « *The Jump* », le chanteur Bobby McFerrin (1984) incorpore le cliché musical « *Na-na-na-na-na-euh* » à une phrase musicale (**extrait musical n°4**). Cette citation est très intéressante, car bien que le texte soit différent (il utilise les syllabes « *le-le-le-le-le-le* »), nous identifions tout de suite le cliché mélodique correspondant. En revanche, dans la deuxième phrase musicale (voir transcription musicale, figure n°14), il modifie le contour en produisant un fa bécarré au lieu d'un fa dièse (soit un demi-ton plus bas), c'est-à-dire un intervalle de tierce majeure au lieu d'une tierce mineure, puis une quarte augmentée au lieu d'une quarte juste (pour les non musiciens, se reporter au tableau des correspondances p : 7). Bien qu'il contribue à l'expressivité de la phrase musicale, ce « petit » changement suffit à briser le cliché. Les intervalles ne souffrent donc aucun changement et la forme est complètement figée. Par ailleurs, la rythmique est régulière et facilement restituable grâce à la notation musicale :

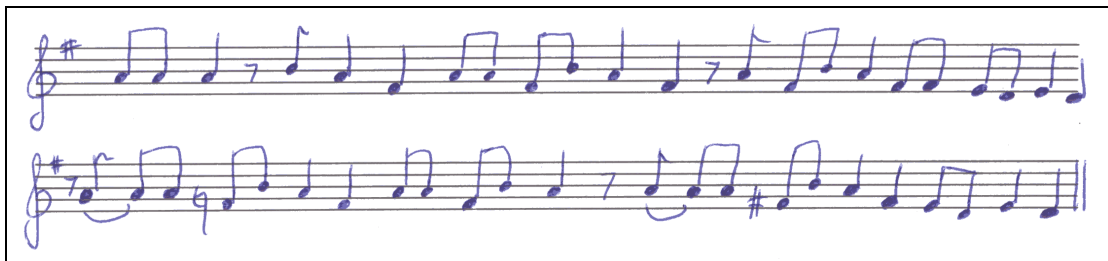


Figure n°14 : extrait de « The Jump » de Bobby McFerrin (1984).

1.2.4. Degré de mélodicité et de rythmicité, la clé de la musicalité ?

D'après Fonagy (1983 : 155) :

« c'est le degré de mélodicité plus ou moins élevé qui distingue les différents genres du discours : chant, récitatif, sermon, déclamation, discours politique, exposé, conversation ».

La mélodicité semble donc bien être un critère essentiel pour caractériser une parole musicale et donc pour la distinguer d'un discours « non musical ». D'après nos observations, le degré de mélodicité pourrait se définir par le degré de fixation des notes tonales importantes, par l'intermédiaire de leur allongement et/ou de leur emplacement au sein de la cellule rythmique. Le degré de rythmicité participe également à la musicalité d'un énoncé. Nous le définirons comme le degré de stabilité de la rythmique. Si ce degré est fort, la pulsation sera facilement reproductible par l'auditeur. Nous pourrions formaliser ces relations de la manière suivante : un axe vertical borné en haut par l'« intonativité » et en bas par la « musicalité ». Plus le degré de mélodicité et plus le degré de rythmicité seront importants, plus l'énoncé considéré sera musical. On peut donc classer les différents types de discours langagiers et musicaux sur ce même axe. C'est ce que fait implicitement Van Waesberghe (1957 : 374, cité par Léon et Martin, 1969 : 66-67) lorsqu'il classe les différents types de communication, de la parole normale à la musique. Sur la figure n°15, nous avons fait correspondre l'axe de musicalité à cette typologie : ainsi, la parole normale possède la plus forte « intonativité » alors que le chant avec rythme imposé possède la plus grande « musicalité ».

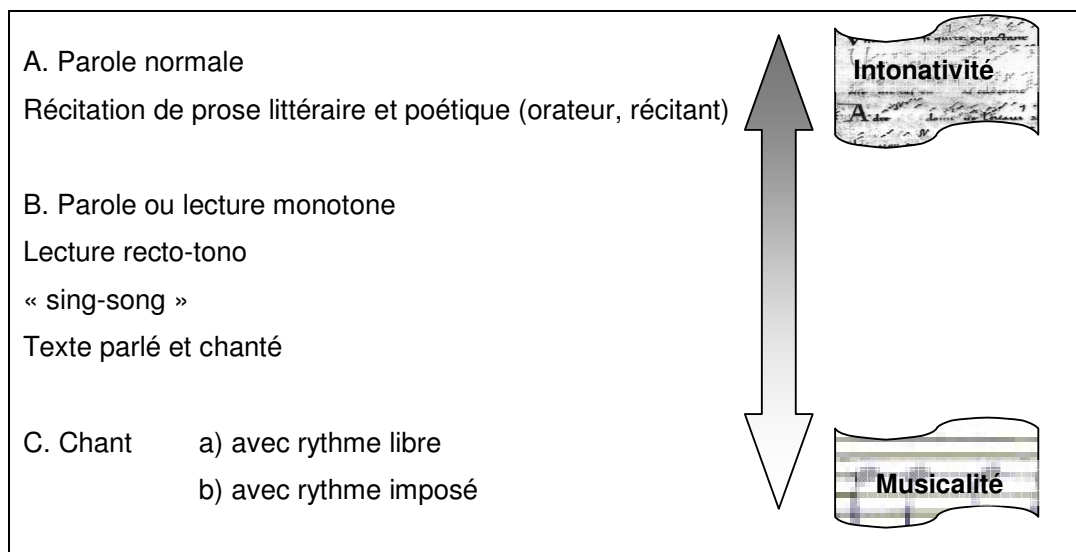


Figure n°15 : typologie des différents types de communication par Van Waesberghe (1957 : 374, citée par Léon et Martin, 1969 : 66-67).

Il nous semble donc beaucoup plus pertinent de classer les différents types d'énoncés en fonction de leur musicalité, c'est-à-dire de leur mélodicité et de leur rythmicité, plutôt qu'en fonction de leur appartenance à la musique ou à la langue.

1.3. Cas de fusion avec la musique : les langues « musicales »

Dans certaines langues, le message ne passe plus que par les éléments musicaux, comme c'est le cas dans les langues sifflées et les langues tambourinées. Ces langues « musicales » imitent le rythme, l'intonation et les tons de la langue parlée correspondante, par l'intermédiaire du sifflement ou par l'intermédiaire d'un instrument de musique. Ce procédé d'imitation est très semblable à l'effet de « guitare parlante » utilisé par Hendrix (cf. 1.2.2.5. et **extrait musical n°3**). C'est la musique de ces langues, c'est-à-dire sa prosodie qui est utilisée pour faire passer des messages normalement véhiculés par le contenu segmental. En l'absence de mots, l'ambiguïté est forcément plus grande, mais reste toutefois limitée en raison de la restriction des situations de communication pendant lesquelles les messages sont échangés. Dans ces langues, on assiste à une véritable fusion de la prosodie et de la musique.

1.3.1. Langues sifflées

Les langues sifflées sont utilisées pour communiquer sur de grandes distances. Grâce à une bande de fréquence étroite (2 KHz), la portée du signal permet en effet de couvrir aisément plusieurs kilomètres (jusqu'à 12 à 14 km pour le Silbo). On a recensé jusqu'à présent quatre langues sifflées. Elles sont la plupart du temps utilisées par des montagnards, et spécialement les bergers qui peuvent échanger des messages d'une vallée à une autre, en dépit du relief accidenté. Elles sont également utilisées dans la conversation courante. La radiocinématographie montre que, pendant la production d'un énoncé sifflé, les organes articulatoires bougent de la même façon que pour la phrase parlée équivalente. Selon les langues, le sifflement peut suivre le pattern du deuxième formant du langage ordinaire du locuteur (Shadle, 1997 : 53). Celui-ci est quasiment réduit à une fréquence pure qui imite les inflexions de l'intonation et les caractéristiques de la langue parlée correspondante. La langue sifflée des Indiens d'Oaxaca au Mexique est fondée sur le mazateco, qui est une langue tonale (Cowan, 1948, cité par Crystal, 1987 : 400). Chaque syllabe correspond à une syllabe de la parole et les 4 tons de la langue parlée sont utilisés. Ces tons sont classés de 1 (haut) à 4 (bas). Cowan transcrit la phrase équivalente en français à « Qu'est-ce que tu m'apportes là ? » de la façon suivante : 1,1,3,3,2,4 hme¹ č²a¹ ši³ ki³ –č²ai²-ve⁴.

Dans les îles Canaries, on trouve une autre langue sifflée, le Silbo de la Gomera qui est fondée sur un dialecte de l'espagnol (Busnel et Classe, 1976). Sur les bords de la Mer Noire, une autre langue sifflée est pratiquée dans le village de Kusköy en Turquie. Cette langue s'appuie sur le turc. Enfin, la dernière langue sifflée a été recensée en France, dans le village d'Aze situé dans les Pyrénées. Elle est fondée sur un sous-dialecte du béarnais, dont la matrice est espagnole.

1.3.2. Langues tambourinées

Les phénomènes rythmiques peuvent également porter le message linguistique : c'est le cas des fameux « tambours parlants » utilisés dans certains pays d'Afrique. Tout comme le sifflement, l'utilisation des percussions permet de délivrer des messages relativement complexes et sur de longues distances. Les langues tambourinées s'appuient sur l'imitation du rythme de la langue parlée correspondante, de manière plus ou moins codifiée. Mais elles peuvent également imiter des phénomènes mélodiques, c'est-à-dire l'inflexion de la voix et les tons. Par exemple, les membres de la tribu Jabo, à l'Est du Libéria utilisent des troncs de 2 mètres de long, dont la fente longitudinale permet de produire différents tons (Herzog, 1945). Ces instruments permettent non seulement de reproduire la rythmique du Jabo, mais également d'en reproduire les 4 tons, et ce qui lie ces tons entre eux, les glissandos. Sur la transcription réalisée par Herzog (figure n°16), la rythmique est indiquée par la valeur des notes (croches, double-croches) et des silences (demi-soupir), c'est-à-dire leur durée. Les tons sont indiqués par les chiffres affectant les différentes syllabes (de 1 à 4), comme c'était le cas pour le mazateco (cf. 1.3.1.).

Figure n°16 : transcription du Jabo (emprunté à Crystal, 1987 : 400).

Dans sa thèse, Essien (2000) décrit le fonctionnement des tambours parlants yoruba utilisés au Nigéria. Des tendeurs relient les deux membranes situées à chaque extrémité de l'instrument. Il est possible de faire varier la tension de ces tendeurs en exerçant une pression variable à l'aide du bras. Cette variation continue de la tension sur la membrane va engendrer une variation continue de la hauteur, lorsqu'on martèle rapidement la peau avec une baguette. Grâce à ce système, le tambour est donc « capable » de reproduire les glissements de l'intonation de la langue naturelle.

Essien explique que :

« des personnes dans l'audience, qui connaissent la mélodie et la chanson, peuvent substituer les sons de la parole aux hauteurs et entendre la « parole » dans la mélodie jouée » (Essien, 2000 : 63).

Ils ont dû au préalable passer par une phase d'apprentissage qui leur a permis d'intérioriser des phrases de la langue naturelle et leurs correspondances tambourinées. La langue yoruba se prête bien à cette « musicalisation », car selon Beier (1976, cité par Essien, 2000 : 63), les tons (trois tons, bas, moyen et haut) sont plus significatifs que les voyelles et les consonnes. Ainsi, beaucoup de mots se distinguent uniquement par leurs tons et non par leur contenu segmental. Pour cet auteur, la capacité du tambour à produire ces tons, équivaut donc à la capacité à parler. Et cette capacité a toujours fasciné les voyageurs, comme le souligne Essien. Il cite un extrait d'*Awake* (1997) qui raconte la transmission d'un message :

« In May 1976, the remarkable ability of drummers to communicate using drums was demonstrated by court musicians of a Yoruba chief. Volunteers from the audience whispered a series of instructions to the master drummer who, in turn, drummed the instruction to another musician located far from the courtyard. Responding to the drum instructions, the musician moved from one location to another and carried out whatever action he was requested to perform » (cité par Essien, 2000 : 62).

1.3.3. Une langue artificielle musicale

S'il existe des langues qui utilisent uniquement les éléments musicaux comme vecteurs linguistiques, en se fondant principalement sur les patterns rythmiques et intonatifs de la langue correspondante, alors, pourquoi ne pas imaginer une langue artificielle qui serait fondée uniquement sur le langage musical ?

Une telle langue pourrait avoir des prétentions universelles, puisque le langage musical « classique » et la notation qui lui est associée sont connus par un grand nombre de personnes à travers le monde. C'est ainsi qu'au début du XIX^e siècle, le compositeur français Jean-François Sudre (1787-1862) créa une « langue musicale universelle », une sorte d'espéranto musical baptisé « solrésil ». Ce projet fut soutenu entre autres par Victor Hugo, Alexandre Humboldt et Alphonse de Lamartine et fut très populaire au XIX^e siècle, notamment à la cour de Napoléon III. Il constitue l'une des plus vieilles tentatives connues de création d'un langage artificiel. En 1902, Boreslas Gajewski en rédigea la grammaire. Le solrésil est fondé sur les 7 notes de la gamme diatonique : **do, ré, mi, fa, sol, la** et **si** (transcrites **d, r, m, f, so, l, s**). Ces notes sont utilisées comme des syllabes et elles peuvent se combiner entre elles pour créer des mots. Produites de manière isolée, la note **si** signifie « oui » et la note **do**, « non ». La combinaison de deux notes engendre des mots grammaticaux, tels que « je » (**dr**) et « tu » (**dm**) ou d'autres mots tels que « bonsoir ; bonne nuit » (**mso**). La combinaison de trois notes est utilisée pour les mots courants, tels que « boire » (**rsol**), « jour » (**drm**), « du pain » (**dsom**), « mai » (**dll**), « soleil » (**soff**) et « siècle » (**drs**). Les mots de quatre notes sont rangés par classes, identifiées par la note initiale. Ainsi,

« les mots commençant par SOL ont rapport au théâtre, à la littérature, aux beaux-arts et aux sciences » (Gajewski, 1902),

somsor pour « musique », par exemple. Il y a en outre plus de 9000 combinaisons à 5 notes pour les noms d'animaux, de végétaux et de minéraux. Le solrésil permet d'obtenir des oppositions sémantiques simplement en inversant les syllabes : par exemple **sosf** pour « rire » et **fsso** pour « pleurer ». On peut indiquer le genre (indiqué par une barre horizontale : « femme », **dmfr** ⁻) et le nombre (indiqué par un accent aigu : « mesdames », ss'o^ˆ). Les verbes sont au participe présent, le temps verbal étant indiqué par une lettre (**D** pour l'imparfait de l'indicatif par exemple : « je commençais », **dr D sdf**). Chaque mot est suivi d'un bref silence ce qui permet à l'auditeur de bien les délimiter.

Le solrésil a été conçu pour pouvoir couvrir un éventail très large de situations de communication. Ainsi, il peut se transcrire avec plusieurs systèmes différents : en toutes lettres, avec des notes sur une portée musicale (la portée utilisée comporte seulement trois lignes, car la gamme utilisée s'arrête au do⁴), avec les 7 premiers chiffres ou des couleurs spécifiques ou des frappes rythmiques associés aux 7 notes. Il peut même être sténographié. Par ailleurs, il peut être indifféremment sifflé, chanté, parlé ou joué à l'aide d'un instrument.

Comme pour les langues sifflées et les langues frappées, il peut donc se transmettre sur de longues distances, de jour comme de nuit. En dépit de ses qualités, le solrésil n'a pourtant jamais atteint l'universalité à laquelle le prédestinait son concepteur. Aujourd'hui, très peu de personnes communiquent avec ce langage. Il nécessite en effet un long apprentissage, car les combinaisons sont difficiles à mémoriser (en raison du nombre limité d'unités). Dans sa version, chantée, sifflée ou jouée, il nécessite de la part du locuteur et de l'auditeur une très bonne oreille musicale, voire l'oreille absolue pour repérer la note initiale des mots de 4 notes. Par ailleurs, si l'auditeur est habitué à attribuer un sens bien précis à des combinaisons de notes en solrésil, il ne pourra plus s'abstenir de le faire lorsqu'il écoutera de la musique. Il est probable qu'il perde le plaisir de la musique, la musique étant pansémique. En plus d'une bonne oreille, il faut chanter juste en solrésil pour ne pas bruite le message. En réalité, le solrésil n'a de musical que le nom, car il utilise des principes très différents de la musique. Alors que :

« d'une manière tout à fait générale, on pourrait dire que, parmi toutes les manières possibles de structurer, à des fins quelconques, le continuum sonore, celles qu'utilisent la musique et le langage se situent à des niveaux différents et peuvent être réalisées simultanément sans qu'il se crée d'interférences » (Ruwet, 1972 : 50).

Pourvu que la fonction distinctive soit assurée dans le langage, il existe une grande liberté dans l'utilisation matériel sonore à des fins musicales. Mais, le problème du solrésil, c'est que le niveau du langage est justement attribué à la musique : les combinaisons de notes jouent le rôle des mots dans les langues naturelles, exactement comme si on utilisait la prosodie à la place des mots. La fonction du contour mélodique n'est pas de porter un sens aussi précis que ne le font les mots. Par ailleurs, dans sa version écrite, le solrésil n'est ni plus, ni moins qu'un langage de signaux semblable au morse. Ce n'est donc pas les similitudes entre langue et musique qui sont exploitées ici.

1.4. L'origine de la langue dans la musique ?

Les similitudes très fortes entre langue et musique et notamment entre la langue parlée et la langue chantée ont poussé certains linguistes, compositeurs et philosophes à poser l'hypothèse d'une origine commune du langage et de la musique. Cette idée se retrouve très fréquemment au cours du XVIII^e siècle. Ainsi, selon Rameau :

« on peut dire que la musique, simplement considérée dans les différentes inflexions de la voix, laissant le geste à part, a dû être notre premier langage » (cité par Didier, 1982 : 400).

Mais selon Kintzler (citée par Didier, 1982 : 400), chez Rameau :

« langue et musique n'ont pas la même identité, elles sont simplement en conformité acoustique, en tant qu'émettrices de vibrations sonores ».

Alors qu'il fonde cette origine uniquement sur les similitudes acoustiques entre la langue et la musique, la question des origines se place sur un terrain essentiellement philosophique pour Condillac et Rousseau. Tandis que Condillac, dans son *Essai sur l'Origine des Connaissances Humaines* (1746) évoque une époque où langage, musique et geste n'auraient fait qu'un, Rousseau, dans son *Essai sur l'Origine des Langues, où il est Parlé de la Mélodie et de l'Imitation Musicale* (1763) se plaît à imaginer la nature des premiers discours des êtres humains, où musique et langue n'étaient pas différenciées selon lui :

« Autour des fontaines dont j'ai parlé les premiers discours furent les premières chansons : les retours périodiques et mesurés du rythme, les inflexions mélodieuses des accens firent naître la poésie et la musique avec la langue, ou plutôt tout cela n'étoit que la langue [...] Les premières histoires, les premières harangues, les premières loix furent en vers ; la poésie fut trouvée avant la prose ; cela devoit être, puisque les passions parlèrent avant la raison. Il en fut de même de la musique ; il n'y eut point d'abord d'autre musique que la mélodie, ni d'autres mélodie que le son varié de la parole, les accens formoient le chant, les quantités formoient la mesure, et l'on parloit autant par les sons et par le rythme que par les articulations de la voix. Dire et chanter étoit autrefois la même chose dit Strabon » (1990 : 114).

Pour Condillac, la musique et la langue se sont ensuite progressivement séparées, selon un processus d'une extrême lenteur :

« la prosodie s'éloigna [...] du chant peu à peu, et à mesure que les raisons qui l'en avoient fait approcher davantage cessèrent d'avoir lieu. Mais elle fut longtemps avant de devenir aussi simple qu'elle l'est aujourd'hui » (1746 : 119, chapitre II, première section, seconde partie « De la Prosodie des Premières Langues »).

Avant cette séparation, la parole donnait sens à la musique, mais la variation musicale pouvait également changer le sens d'un mot. On relève d'ailleurs ce phénomène dans les langues tonales où le changement de ton est distinctif. Ainsi, Condillac apporte pour preuve de l'origine commune de la langue et de la musique l'existence d'une langue comme le chinois, car dans cette langue, la variable qui fait sens peut être d'ordre musical (le ton) ou d'ordre linguistique (le phonème) :

« dans l'origine des langues, la manière de prononcer admettoit donc des inflexions de voix si distinctes qu'un musicien eût pu la noter, en ne faisant que de légers changemens ; ainsi je dirai qu'elle participoit du chant. Cette prosodie a été si naturelle aux premiers hommes qu'il y en a eu à qui il a paru plus facile d'exprimer différentes idées avec le même mot, prononcé sur différents tons, que de multiplier le nombre des mots à proportion de celui des idées. Ce langage se conserve encore chez les Chinois. Ils n'ont que 328 monosyllabes qu'ils varient sur cinq tons, ce qui équivaut à 1640 signes » (Condillac, 1746 : 119, seconde partie, section première, chapitre II « De la Prosodie des Premières Langues »).

Pour Rousseau et Condillac, c'est au moment de la naissance de l'écriture, que la séparation entre la langue et la musique s'est amorcée. Près de deux siècles plus tard, le musicologue soviétique Assafjew (1884-1949) explique ce processus de différenciation de la manière suivante (cité par Menezes, 1993 : 87-88) : certains patterns mélodiques se sont peu à peu fixés et leur reconnaissance grâce à la mémoire auditive a permis la constitution d'une conscience musicale collective. C'est alors que la scission avec le langage s'est réalisée. La musique s'est notamment spécialisée en développant une véritable science des intervalles :

« La précision de plus en plus incisive de l'intervalle en tant qu'indicateur de la tension propre à l'intonation et l'évolution assez frappante des structures musicales dans l'histoire envisageant surtout la composition des intervalles constituent sinon la raison d'une séparation absolue, au moins une distinction assez nette entre musique et langage verbal (ou poésie). [...] L'harmonie, science des proportions d'intervalles, émerge, nous le constatons, en tant qu'élément le plus éminent du discours musical – l'intonation mélodique en constituant sa manifestation la plus élémentaire –, et l'intervalle sera vu comme une forme primordiale de la musique en tant qu'expression autonome, bien qu'il ait irréfutablement, et justement pour cette raison, son origine au sein même du Verbe » (Menezes, 1993 : 87).

Selon Menezes (1993 : 113), cette spécialisation de la langue et de la musique se serait manifestée par une différenciation des écritures verbales et musicales. L'écriture alphabétique a omis, avec l'établissement de l'alphabet, la transcription de l'intonation et des relations de durée des phonèmes dans leurs réalisations, en excluant la prosodie de son contexte. L'avènement plus tardif de l'écriture musicale est dû aux relations d'intervalles et de durées, développées de façon toujours plus systématique et exhaustive que dans le langage parlé. L'écriture musicale est donc née en raison de la spécialisation du code musical vis-à-vis des propriétés prosodiques du langage et a contribué à la démarcation historique des limites entre codes musical et linguistique.

Enfin, parmi d'autres théories, Jespersen (1922 : 401-402) cite celle de Noiré qui situe l'origine du langage dans le rythme : lors de tâches effectuées en commun, les hommes produisaient des grognements rythmés pour accompagner l'effort physique qu'ils fournissaient. Ces sons finissaient par s'associer à l'acte lui-même et servaient à le désigner :

« en conséquence, les premiers mots signifieraient quelque chose comme « hisser » ou « haler » » (Jespersen, 1922 : 402).

Cette théorie est surnommée la « yo-he-ho » theory ou si nous la traduisons en français, la théorie « eh-oh-eh-oh ». Jespersen refuse de spéculer sur l'origine du langage et utilise un raisonnement inductif fondé sur l'observation du langage enfantin, du langage des peuplades primitives et de l'histoire du langage. Comme Rousseau et Condillac, il en tire la conclusion que le premier langage se rapprochait plus du chant, que de la parole telle que nous la connaissons aujourd'hui :

« et il nous faut donc conclure que le langage des hommes non civilisés et primitifs contenait bien plus d'agitation passionnée que le nôtre et ressemblait plus à un chant, une musique » (Jespersen, 1922 : 406).

Selon Jespersen, l'origine du langage doit se rechercher en effet dans les aspects poétiques et romantiques de l'existence et notamment, dans l'expression des émotions :

« Le langage naquit à l'époque des amours de l'humanité ; et je veux croire que les premiers énoncés produits tenaient à la fois de la plainte nocturne d'une chatte amoureuse sur un toit et des mélodieux chants d'amour du rossignol » (Jespersen, 1922 : 419-420).

Dans ce langage « primitif » fondé exclusivement sur l'expression des émotions, la langue et la musique ne se différenciaient pas, elle formaient un tout. Ainsi, pour Jespersen, *« il a existé une époque où tout langage était chant, ou plutôt durant laquelle ces deux choses n'étaient pas encore distinctes »*. Il ajoute :

« L'homme a chanté ses sentiments bien avant d'être capable d'exposer ses pensées. Mais il ne faut pas, bien entendu, nous imaginer que « chanter » a dans ce contexte la même signification que dans une salle de concert moderne. Lorsque nous disons que la parole est née du chant, ce que nous voulons dire, c'est essentiellement que notre langage relativement monotone et que notre musique vocale très élaborée ne sont que les résultats de la différenciation d'énoncés primitifs, qui ressemblaient moins au premier qu'à la seconde » (1922 : 422).

2. Exploitation de ces similitudes de forme

Tout au long de ces lignes, nous avons vu que les rapports étroits entre la langue et la musique puisaient leurs origines dans les similitudes très fortes qui existaient entre la musique et la prosodie. Mais, d'après Menezes (1993 : 113), il existe un véritable paradoxe dans le fait que le lien entre la musique et la langue ne puisse se faire que par l'intermédiaire de la parole. En effet, depuis l'avènement de l'alphabet, la prosodie a été complètement exclue de la transcription orthographique ; celle-ci ne permet absolument pas de rendre compte de l'intonation et des rapports de durée entre les phonèmes. La prosodie appartient donc uniquement à la sphère de l'oral, c'est-à-dire à la parole. C'est pour cette raison que les linguistes et en particulier les phonéticiens ont dû créer des systèmes de notation de la prosodie pour combler le vide laissé par l'orthographe (cf. partie 2.1.1.2. et 2.1.2.). Au contraire, l'écriture musicale s'est spécialisée sur la transcription de la prosodie, notamment sur les relations entre les hauteurs et entre les durées des sons. Elle n'a pas cessé de s'affiner, jusqu'à atteindre une très grande précision dans la notation des intervalles, dans la fixation des hauteurs et des durées (voir historique de la notation musicale, partie 1.2.2.2. et 2.1.1.1.). C'est grâce à son écriture que la musique a pu différencier les niveaux acoustique, instrumental et stylistique, et par conséquent, qu'elle a pu exercer un contrôle absolu de chacun de ces niveaux. Les propriétés acoustiques ont ainsi pu se libérer des conventions auxquelles elles étaient soumises. L'écriture musicale a permis également d'approfondir les rapports entre les propriétés acoustiques et leurs propriétés auditives. Elle a donc permis l'évolution des méthodes de composition et l'évolution du style musical. C'est ce dynamisme qui distingue la notation musicale de la notation orthographique, cette dernière possédant un caractère éminemment immuable. Pour Menezes (1993 : 119) :

« L'écriture musicale est née en raison de la spécialisation du code musical vis-à-vis des propriétés prosodiques du langage et a corroboré de façon décisive la démarcation historique des limites entre les codes musical et linguistique ».

2.1. Lorsque la musique devient langue

2.1.1. Transcription musicale de la prosodie

2.1.1.1. Potentialités de la transcription musicale

Nous avons vu dans le paragraphe 1.2.2. que la mélodie de la musique et celle de la parole ont beaucoup de similitudes, notamment dans le fait qu'elles se caractérisent par une succession de sons au cours du temps (dimension temporelle) et que ces sons entretiennent entre eux des rapports de hauteur (intervalles), des rapports de durée, mais aussi des rapports d'intensité (ils ne sont pas tous prononcés avec la même force articulatoire). On relève également des similitudes entre le rythme de la musique et celui de la langue, dans le fait qu'ils se définissent tous deux par une succession de constituants forts et de constituants faibles. Ces constituants forts sont mis en relief par une combinaison particulière des trois paramètres acoustiques que sont la durée, la Fo et l'amplitude. Cette mise en relief correspond au phénomène d'accentuation, présent à la fois dans la langue et dans la musique. A la question de savoir comment représenter la mélodie et le rythme d'un énoncé, les phonéticiens se sont donc tout naturellement tournés du côté de la transcription musicale classique, qui permet justement de représenter la durée, la hauteur et l'intensité (voir figure n°17). Selon Leipp (1971 : 81),

« cette « partition » si simple et, point important, lisible de façon quasi instantanée, comporte, en fait, un nombre extraordinaire de renseignements sur les variables physiques et psycho-physiologiques en présence ».

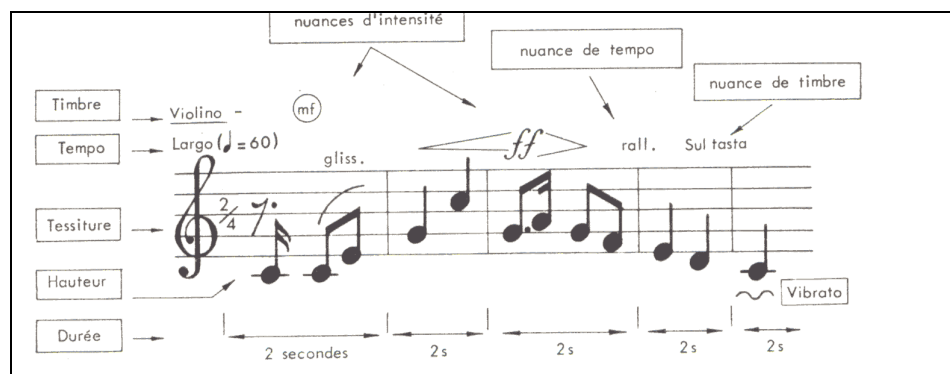


Figure n°17 : représentation musicale classique (emprunté à Leipp, 1971 : 81).

A la lumière du parallèle langue / musique, les indications qui s'avèrent particulièrement intéressantes pour la transcription la prosodie portent sur⁷ :

- **les indications de durée** : l'**axe horizontal** de la portée qui représente la dimension temporelle de la langue et de la musique ; les **valeurs des notes** (la noire dure un temps par exemple) **et des pauses** (la demi-pause dure deux temps par exemple) pour la durée des syllabes et des pauses dans la parole ; l'indication de vitesse d'exécution, c'est-à-dire le **tempo**, pour le débit (noire=60, 60 noires par minute ; on trouve d'autres indications de tempo qui permettent d'accélérer ou de ralentir au cours de l'exécution, comme « *accelerando* » ou « *rallentando* ») ; les indications qui concernent la mesure (le chiffrage 2/4 indique qu'il y a deux noires par mesure) et les barres de mesure (barres qui permettent de délimiter les mesures) semblent inappropriées, car elles impliquent une trop grande métricité, absente dans la langue. Il paraît donc difficile de faire tenir le rythme de la langue entre deux barres de mesure.
- **les indications de hauteur** : l'**axe vertical** de la portée qui représente l'étagement des hauteurs du plus grave au plus aiguë (de bas en haut), pour indiquer les rapports de hauteur entre les sons (notes et syllabes) ; la **clef** (ici la clef de sol & permet de fixer une référence en associant une hauteur absolue, la note sol³, 392 Hz, à la deuxième ligne de la portée) et les **cinq lignes de la portée** permettent de fixer la **hauteur** (une fois la référence fixée grâce à la clé, les lignes de la portée permettent de représenter les hauteurs en fonction de ce point) ;
- **les indications d'intensité** : symboles ajoutés en dessous de la portée (***pp*** pour « *pianissimo* », très doux ou ***ff***, pour « *fortissimo* », très fort par exemple) ; certains d'entre eux indiquent une augmentation (« *crescendo* ») ou une diminution progressive de l'intensité (« *decrescendo* »), voir figure n°18.

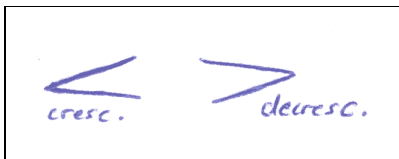


Figure n°18 : symboles utilisés dans la notation musicale pour indiquer une augmentation progressive de l'intensité (*crescendo*) ou une diminution progressive de l'intensité (*decrescendo*).

⁷ Pour plus de détails sur la notation musicale elle-même, se reporter à Leipp (1971 : 81-82).

2.1.1.2. Historique des applications de la transcription musicale en prosodie

Les potentialités de l'écriture musicale ont séduit nombre de linguistes et de phonéticiens et ce, depuis de nombreux siècles. Il y a eu beaucoup de tentatives d'adaptation de cette notation aux particularités de la prosodie des langues naturelles et il est possible d'en retracer rapidement l'historique en s'arrêtant aux noms les plus connus. Ces tentatives se heurtent aux obstacles des différences entre langue et musique ou réussissent au contraire à les dépasser. Elles permettent de compléter les différences et les similitudes mises à jour dans la partie 1.2.

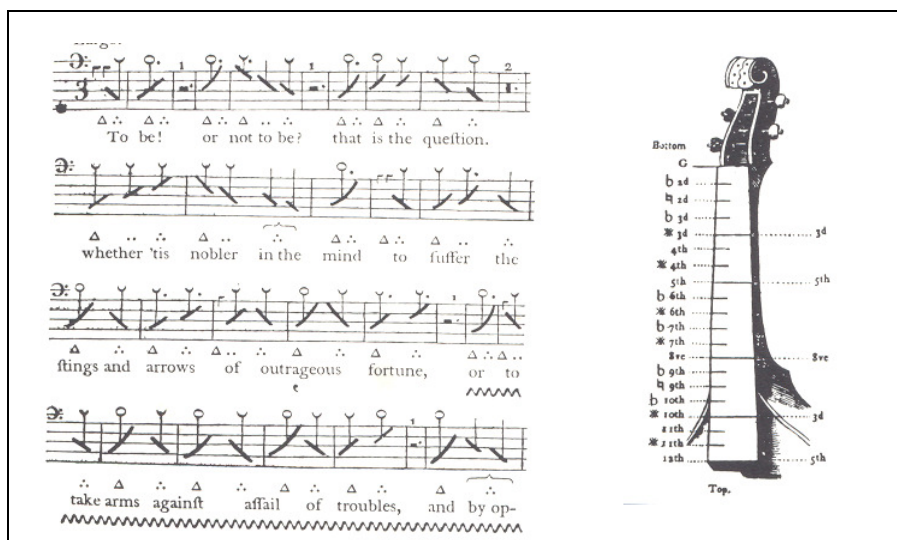
Dès le XVI^e siècle, le grammairien Louis Meigret, cité par Konopczynski (1978 : 82) tente de représenter l'évolution de l'intonation et de l'accentuation sur une portée musicale dans son *Treuvé de la Gramèze françoise* (1550). Son système n'avait recours qu'à une seule valeur de note, la blanche et se limitait à deux hauteurs différentes. Une transcription aussi primitive n'exploitait que très partiellement les possibilités de la notation musicale et ne pouvait bien sûr pas rendre compte de la richesse musicale d'une langue naturelle, tant du point de vue de ses variations de hauteur, que de ses durées. Près de deux siècles après Meigret, Joshua Steele fit une exploitation beaucoup plus approfondie de la notation musicale dans une étude portant sur l'intonation anglaise, « *Prosodia Rationalis or An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols* » (1775)⁸. A l'opposé de celle de Meigret, la transcription de Steele est extrêmement précise dans la notation des durées et de la hauteur. Faure reproche d'ailleurs à Steele ce degré de précision :

« l'auteur s'efforce d'autre part, ce qui est fort aléatoire, [...] de noter des différences de hauteur de l'ordre du quart de ton, du demi-ton et de trois quarts de ton entre les syllabes successives d'un énoncé et des repos allant jusqu'au huitième de pause » (1962 : 118).

En réalité, cette précision s'explique par le procédé adopté par Steele pour transcrire la mélodie (Crystal, 1987 : 170) : il a indiqué sur le manche d'une viole de gambe (basse de viole, même tessiture que le violoncelle) les hauteurs correspondant aux **frets** (à droite de la figure n°19). A la manière d'Hendrix, il a ensuite reproduit les variations de hauteur de la voix sur le manche de la viole et les a transcrites sur une portée musicale (à gauche de la figure n°19). La transcription présentée est celle d'un extrait de « *Hamlet* » de Shakespeare, le célèbre « *To be or not to be...* ». La durée de chacun des glissement est symbolisée par des signes verticaux, indiquant une durée relative (du plus court au plus long).

⁸ Cité par Faure (1962 : 18), Crystal (1969 : 23 ; 1987 : 170) et Konopczynski (1978 : 82).

Le degré d'emphasis figure en dessous de la portée ; les symboles sont également relatifs. Les pauses sont signalées par des symboles empruntés à la notation musicale : les silences.



Steele se rendit compte que les points de départ et d'aboutissement des glissandi se trouvaient très souvent entre deux frets et ne correspondaient donc pas aux notes de la gamme diatonique. Pour cette raison, ces points sont la plupart du temps localisés entre deux lignes de portées sur sa transcription. Un siècle plus tard, le français Léonce Roudet (1899) préconisa une adaptation de la notation musicale, car il avait remarqué lui aussi que les hauteurs ne correspondaient pas aux hauteurs des notes de la gamme. Pour cela, il utilisa une échelle logarithmique en Hz, sur laquelle étaient indiquées les correspondances avec les notes de la gamme. Ces notes étaient prolongées par des lignes horizontales qui remplaçaient la portée traditionnelle (10 lignes, étagées de mi^2 -165 Hz à sol^3 -396 Hz, soit une tessiture d'une dixième mineure). En outre, pour Roudet, la notation musicale s'avérait inapte à représenter les variations continues de hauteur. Pour palier ce problème, il les représenta donc par une ligne continue évoluant sur sa portée à 10 lignes. Les pauses étaient indiquées par une rupture de cette ligne mélodique (figure n°20).

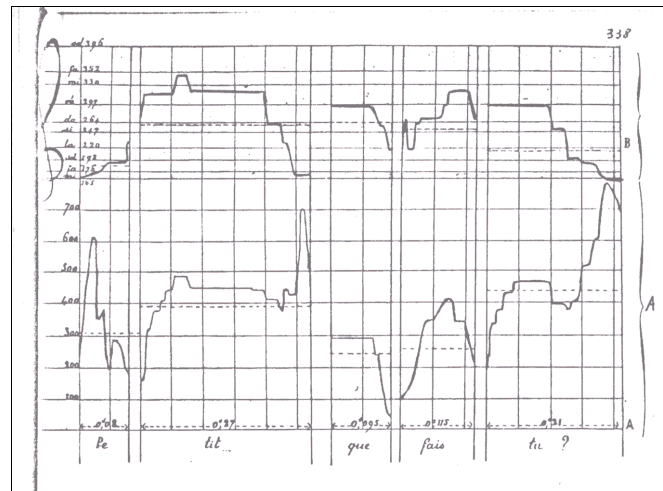


Figure n°20 : transcription utilisée par Roudet (1899),
emprunté à Konopczynski, 1978 : 93

L'axe horizontal représentait l'axe temporel comme dans la notation musicale. Mais la durée ne pouvait plus être indiquée par les valeurs des notes, ni par les barres de mesures, car elles avaient été supprimées par Roudet. L'axe temporel était donc gradué selon l'échelle suivante : 1 cs = 1,6 mm (Konopczynski, 1978 : 84). Quant à l'intensité, elle était représentée par une autre courbe, placée juste en dessous de la courbe mélodique. Le système de notation de Roudet ressemble beaucoup aux représentations des courbes mélodiques et des courbes d'amplitude que l'on obtient avec les logiciels d'analyse phonétique modernes. L'abbé Rousselot (1901-1908) utilisera, quelques années plus tard, le même type de transcription à quelques différences près. L'échelle utilisée pour les hauteurs n'est plus logarithmique, mais musicale et graduée par demi-tons. Une courbe d'amplitude accompagne la courbe mélodique, mais elle figure au-dessus d'elle (figure n°21). Quant à la durée, l'unité de mesure n'est pas précisée car Rousselot reproduit les enregistrements.

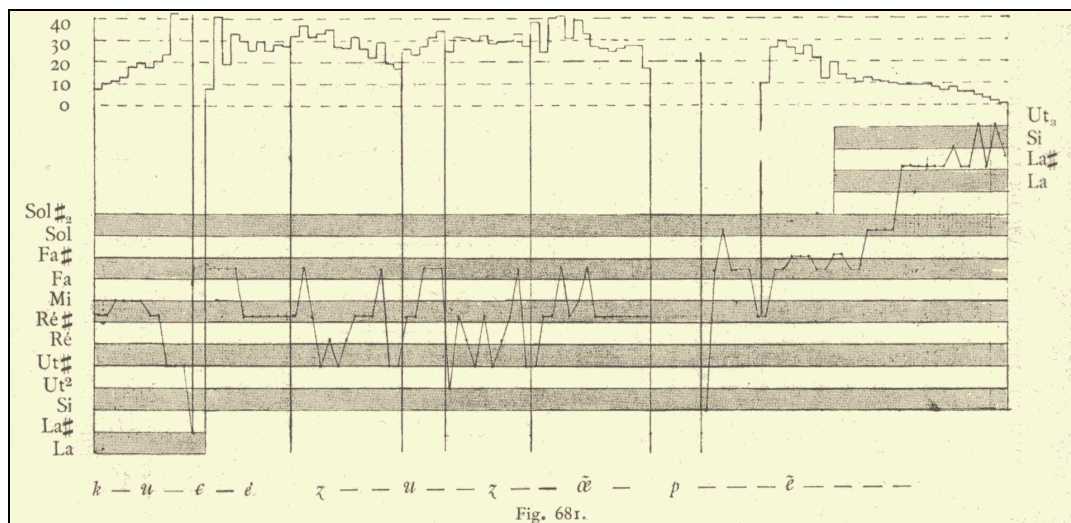


Figure n°21 : transcription utilisée par Rousselot (1901-1908 : 1005).

Ce système de transcription est donc nettement moins précis que ne l'était celui de Roudet. En 1909, paraissent les célèbres « *Intonation Curves* » de Daniel Jones. Dans cet ouvrage, Jones utilise le système de notation musicale pour transcrire la prosodie de textes dramatiques dans deux langues différentes : l'anglais et le français. Contrairement à Roudet et Rousselot, il ne modifie pas la portée musicale, mais comme eux, il représente l'évolution de la hauteur dans la parole par une courbe continue (continuité de la ligne mélodique anglaise). Cette courbe est fondée sur des bases auditives. Il est à souligner que le contour de la mélodie est interrompu à l'emplacement des sourdes (sons non voisés, pas de hauteur). La transcription phonétique (sons et accents) du texte figure en dessous de la portée et les différentes syllabes sont délimitées par des barres de mesures (voir figure n°22, empruntée à Konopczynski, 1978 : 95). La durée de chaque syllabe est représentée par la longueur relative de la mesure. En dehors de ces barres de mesure, il n'y a pas d'indications de durée. On ne trouve pas non plus d'indications d'intensité. Cette transcription est donc beaucoup plus simple que les trois précédentes, mais surtout beaucoup plus lisible. C'était le but recherché par Jones, dont l'approche était avant tout celle d'un didacticien. Les phonéticiens Roudet et Rousselot visaient au contraire à transcrire avec le plus de détails possibles l'évolution des paramètres acoustiques et ce, dans un but d'analyse. Et il est normal d'adapter le système de transcription à ses objectifs. De manière générale, tous ces auteurs ont essayé d'adapter le système de notation musicale, afin qu'il reflète les spécificités de la musique des langues naturelles.

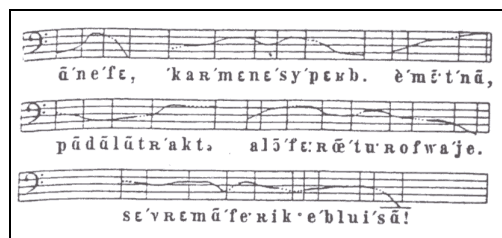


Figure n°22 : transcription utilisée par Jones dans « *Intonation Curves* » (1909)
(emprunté à Konopczynski, 1978 : 95).

Le système de transcription adopté par le psychologue allemand Wilhelm Wundt est un exemple intéressant, car dans un premier temps, cet auteur a transposé la notation musicale dans le domaine de la langue sans la modifier. Selon Konopczynski (1978 : 88),

« Wundt est si influencé par la musique qu'il en arrive à assimiler la phrase à un air dont il commente les caractéristiques ».

Il utilise même une armure, ce qui implique une tonalité. Steele et Roudet avaient justement abandonné les notes musicales, car elles symbolisaient des hauteurs fixes, correspondant à la gamme diatonique. Avec ce système le plus petit intervalle possible entre deux notes est le demi-ton (gamme tempérée) et les intervalles sont toujours décomposables en un nombre de demi-tons entier (multiples). On rencontre ce type d'intervalles dans la langue (voir tableau d'intervalles), mais également des intervalles avec des rapports plus « complexes ». Un des principaux danger avec la notation musicale, c'est d'interpréter ces intervalles complexes en fonction de la structure des « bons » intervalles du système tonal, que nous avons intériorisés depuis l'enfance. Cela nous rappelle étrangement le phénomène de « crible » phonologique, qui nous pousse à interpréter les phonèmes d'une langue étrangère en fonction de la structure des phonèmes de notre langue maternelle (cf. chapitre V, partie 1.2). Un musicien, lorsqu'il transcrit la mélodie de la langue, aura tendance à « normaliser » les intervalles et sa transcription sera faussée. On pourrait parler de crible « tonal » et il est fort probable qu'il s'applique à la perception des musiques non tonales (cf. chapitre II, 3.1.3.3.). Wundt s'est vite rendu compte qu'il fallait abandonner la rigidité des notes musicales en raison de cette

« différence essentielle entre parole et musique, cette dernière contenant toujours des intervalles harmoniques ; d'où la difficulté d'appréciation des intervalles de hauteur, généralement rapprochés des intervalles musicaux connus, et donc minimisés » (Konopczynski, 1978 : 89)

et que pour palier ce phénomène de crible, il était nécessaire de coupler une analyse instrumentale à l'analyse auditive.

En 1900, il adopte un nouveau système de transcription fondé sur une analyse instrumentale et dans laquelle les variations mélodiques sont représentées par une ligne continue, qui évolue sur une portée musicale à 5 lignes (figure n°23). Des notes sont ensuite affectées aux points d'inflexion de la mélodie. Les accents sont indiqués sur la transcription orthographique qui figure au-dessus de la portée (accent aigu et accent grave). Quant à la durée, elle est représentée par l'espacement relatif entre les différentes notes. Signalons la grande ressemblance avec le système de Jones.



Figure n°23 : transcription utilisée par Wundt (1900)
emprunté à Konopczynski, 1978 : 97.

Mis à part Rousselot et Roudet, tous les auteurs évoqués jusqu'à présent avaient conservé la portée musicale à cinq lignes. Pike (1945) sera le premier à bouleverser ce système en introduisant la notion de « niveaux » intonatifs (on parle également de registre). La tessiture qui correspond à une portée musicale (de re^3 à sol^4 , soit un intervalle de 17 demi-tons, 293,67 Hz-783,99 Hz) ne correspond pas à la tessiture couverte par la voix dans la parole. Pour complexifier encore les choses, cette tessiture n'est pas la même selon les langues. Elle est par exemple beaucoup plus grande en anglais (plus de deux octaves, voir chapitre V, partie 1.1) qu'en français. Il faut donc adapter la notation musicale à la tessiture de la langue étudiée. Pour répondre à ce problème de tessiture, il est possible de diviser l'espace fréquentiel couvert par la voix en un certain nombre de zones, appelées registres ou niveaux, relatifs bien sûr au registre de l'individu. Pike (1945) fut le premier à employer les niveaux intonatifs, dans la description de la prosodie de l'anglais américain. De 5 lignes, la portée passe à 4 lignes, chacune représentant un niveau distinct ; ces niveaux sont espacés par un même intervalle de hauteur. La portée est divisée horizontalement en intervalles de temps égaux ; pauses indiquées par la rupture de la ligne mélodique. Pour l'intensité, on trouve l'indication de la place et de la force de l'accent. Ce système de niveaux a été complété par Faure (1962) pour la description de l'anglais britannique. Cette langue se caractérise en effet par de fortes variations mélodiques qui lui donnent des propriétés très musicales (voir chapitre V). Faure est passé à un système à 5 niveaux, ce qui redonne à la portée le même « visage » que la notation musicale.

Ce système à 5 niveaux a été repris par Zwanenburg (1965) et Delattre (1969) pour le français (voir figure n°24). La notation de Delattre est essentiellement mélodique, elle ne comporte aucune indication de durée et d'intensité.

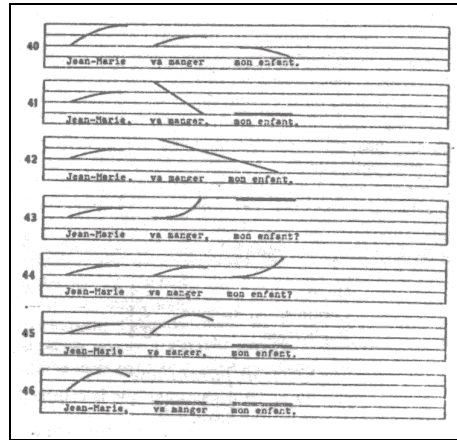


Figure n°24 : transcription de la phrase « Jean-Marie va manger mon enfant » par Delattre (1969 : 12).

Léon et Léon (1964) ont également utilisé une représentation par niveaux, mais au lieu de représenter la courbe mélodique, ils ont inséré une transcription orthographique de l'énoncé dont les différentes syllabes se répartissent sur quatre niveaux différents. Ce procédé est similaire à celui de Bolinger (voir plus loin, figure n°31), qui fait « chanter » les lettres en leur faisant suivre l'évolution de la mélodie, à la différence près que celui-ci ne se réfère pas à la notion de niveau et qu'il n'a pas recours à une « portée ».

Nous l'avons vu, toute transcription répond à des objectifs bien précis. Et lorsque le but de la recherche est justement de comparer, voire de rapprocher la langue de la musique, une transcription musicale authentique paraît la mieux appropriée. Fonagy et Magdics (1963) n'ont pas eu recours aux niveaux, car leur objet d'étude porte sur la comparaison des émotions dans la parole et dans la musique vocale européenne. Bien au contraire, l'utilisation de la notation musicale leur permet de comparer la ligne mélodique de la langue à celle de la musique. Ils sont donc restés très fidèles à la transcription musicale, comme on peut le voir sur la figure n°25.

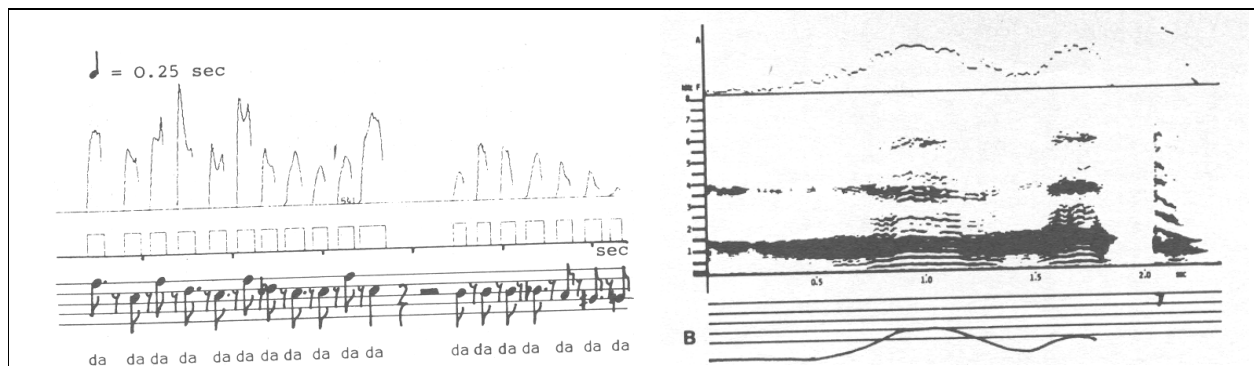


Figure n°25 : extrait de la transcription de Fonagy et Magdics (1963).

Papousek et Papousek (1981) ont réalisé une étude sur les productions vocales et musicales de leur enfant, Tanya. Pendant la première année de la vie, les éléments musicaux jouent un rôle de premier plan dans le langage émergent (cf. travaux de Konopczynski et chapitre II). La notation musicale s'avère donc très appropriée pour transcrire les premiers énoncés produits par l'enfant. Par ailleurs, comme c'était le cas pour Fonagy et Magdics, elle facilite la comparaison entre productions vocales et productions musicales. Elle ne peut donc pas être remplacée par une analyse instrumentale. Comme Papousek et Papousek (1981 : 217) le soulignent eux-mêmes,

« our present experience suggests that the importance of modern audiovisual records of behavior, spectrographic, and partially computerized digital analyses of vocal sounds do not yet outdate traditional musicological methods including subjective evaluation and musical transcription ».

Ils ont donc couplé une transcription musicale à une ou plusieurs représentations instrumentales. La figure n°26 nous en montre deux exemples : celle de gauche combine une transcription musicale en bas et une courbe d'amplitude au-dessus. La ligne intercalée entre les deux représentations symbolise l'évolution temporelle de l'énoncé et donne la référence rythmique (croches), qui permet d'interpréter correctement la rythmique de la transcription musicale. A droite, on peut voir une courbe de mélodie et une représentation spectrographique couplées avec une représentation semi-musicale (ligne continue qui évolue sur une portée musicale).



Enfin, à titre anecdotique, quittons la sphère des phonéticiens et des linguistes, pour citer le système de transcription adopté dans la méthode « *L'Anglais par le Rap* » (1992), citée par Ribière-Raverlat (1997 : 185). Cette méthode d'anglais s'adresse à des collégiens. Pour leur

permettre de visualiser facilement l'intonation en même temps que le rythme, un système de transcription a été mis au point (figure n°27) : l'énoncé, transcrit orthographiquement figure dans une large flèche dont les courbes suivent l'évolution de la ligne mélodique. Cette flèche vient se superposer à une portée musicale. Les syllabes dont les caractères sont indiqués en gras doivent être accentués.

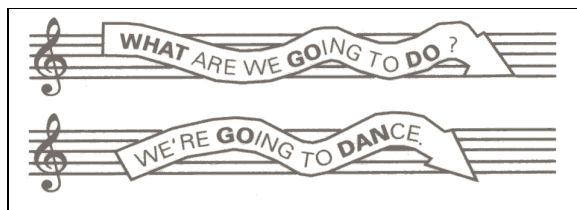


Figure n°27 : transcription musicale utilisés dans la méthode « L'Anglais par le Rap » (emprunté à Ribière-Raverlat, 1997 : 185).

Depuis la blanche de Meigret jusqu'aux niveaux de Delattre ou Faure, il a fallu toute une évolution historique pour que la notation musicale s'adapte à la « musique » de la langue. Toutes ces tentatives ont permis d'approfondir progressivement les connaissances sur les différences et les similitudes entre la langue et la musique. Par ailleurs, l'apparition des analyseurs de mélodie a permis de résoudre les problèmes liés à l'analyse auditive pure, qui concernaient principalement les erreurs d'appréciation des intervalles par l'oreille à cause du crible tonal. Mais à l'inverse, l'analyse instrumentale ne doit pas faire perdre l'usage de la notation musicale, car celle-ci obéit justement à une analyse auditive qui manque évidemment à l'analyse instrumentale. Le danger est donc de perdre l'ancrage auditif de la mélodie et donc son ancrage musical. Le couplage des deux types d'analyses paraît donc être un très bon compromis, comme Wundt l'avait déjà compris il y a plus d'un siècle. Les nouveaux systèmes de notation de la prosodie (cf. partie 2.1.2.) ne doivent pas sonner le glas de la transcription musicale. Chaque système répond en effet aux objectifs d'analyse bien spécifiques de son utilisateur. Des chercheurs comme Fonagy et Magdics, Papousek et Papousek ont utilisé la notation musicale car celle-ci constituait le système le mieux adapté à leurs objectifs de recherche. Massin et Massin (1983 : 124) faisaient remarquer à propos des différents types de transcription utilisés pour la musique que

« chercher un système de notation généralisable à toutes les musiques correspondrait à la volonté d'imposer un unique système de pensée et d'analyse ».

Cette remarque vaut tout autant pour les systèmes de transcription de la prosodie et elle a d'autant plus de résonance que la tendance actuelle en matière de transcription va justement vers une uniformisation des pratiques.

2.1.1.3. Transfert de la terminologie musicale vers la prosodie

Reflétant l'analogie entre langue et musique, de nombreux termes musicaux ont été transférés du domaine de l'analyse musicale à celui de la description prosodique. Il n'est qu'à se référer aux glossaires de deux livres parus récemment et concernant la prosodie du français : « *La Prosodie du Français* » de Lacheret-Dujour et Beaugendre (1999 : 271-288) et « *L'Intonation, le Système du Français : Description et Modélisation* » de Rossi (1999 : 203-212). Dans le premier ouvrage, on trouve 16 mots provenant du domaine musical : accent, ambitus, anacrouse, attaque, cadence, coda, gamme tonale, glissando, intervalle mélodique, octave, pauses, registre, tempo, tessiture et unité tonale. Dans le second ouvrage, on trouve 8 mots : accent, glissando, mélodie ou tunes, phonème tonal, registre, tessiture et ton. Par ailleurs, en A.P.I., il existe une extension des symboles et des signes diacritiques pour permettre la transcription étroite des énoncés. Parmi ces symboles, on retrouve certaines notations utilisées pour la transcription musicale et notamment, la transcription du tempo et des nuances. Ainsi, les variations de débit de parole sont indiquées par des indications de tempo comme **lento** (débit lent), **allegro** (débit rapide), **accelerando** (accéléré) et **rallentando** (ralenti) :

[XXX{ <i>allegro</i> XXX <i>allegro</i> }XXX]	débit rapide
[XXX{ <i>lento</i> XXX <i>lento</i> }XXX]	débit lent

Figure n°28 : indications de tempo utilisées pour la transcription du débit de parole en API.

Les variations d'intensité sont indiquées par des symboles de nuances comme ***f*** (**forte** : parole forte), ***ff*** (**fortissimo** : parole plus forte) et ***p*** (**piano** : parole douce) :

[XXX{ <i>f</i> XXX <i>f</i> }XXX]	parole forte
[XXX{ <i>ff</i> XXX <i>ff</i> }XXX]	parole plus forte
[XXX{ <i>p</i> XXX <i>p</i> }XXX]	parole douce

Figure n°29 : indications de nuances utilisées pour la transcription des variations d'intensité dans la parole en API.

2.1.2. D'autres systèmes de transcription : leur validité par rapport à la transcription musicale

D'autres systèmes que la notation musicale ont été mis au point pour transcrire la prosodie. Nous n'en ferons pas le détail ici, mais on peut toutefois dégager trois grands types de transcription prosodique :

- On peut symboliser les mouvements mélodiques par des flèches, telles que ↑ pour indiquer une montée et ↓ pour indiquer une descente. Le système INTSINT⁹ (Hirst et Di Cristo, 1998) permet par exemple de transcrire l'évolution de la Fo sur une ligne séparée, sous la transcription orthographique ou phonétique (de manière à être indépendante de la division en groupes accentuels). Les flèches indiquent les points cibles, c'est-à-dire les pics ou les vallées de la courbe, qui sont considérés comme des tons (jugés comme les moments les plus informatifs du point de vue de la perception et de la production). Les symboles < (« Upstep ») et > (« Downstep ») indiquent une descente ou une montée par paliers, par terrasses (figure n°30).

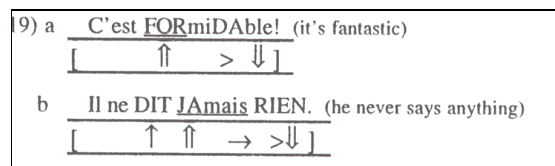


Figure n°30: transcription d'un énoncé à l'aide d'INTSINT (Di Cristo, 1998 : 209).

- A la place des flèches, on peut utiliser des lettres, comme **B** pour bas et **H** pour haut. C'est ce que fait par exemple le système ToBI¹⁰ (Silverman et al., 1992) pour l'anglais américain. Les points cibles mélodiques sont repérés par les lettres **H** (« High ») pour les tons hauts et **L** (« Low ») pour les tons bas. Les frontières sont indiquées par le symbole %.
- On peut tordre les lettres de manière à ce qu'elles représentent l'évolution de la courbe de Fo. C'est ce qu'a fait Bolinger (1998) pour décrire l'intonation de l'anglais américain (figure n°31). Ce faisant, il a réalisé un codage iconique de la Fo (Guaïtella, 2000 : 174), très facilement déchiffrable (cf. sur les portées primitives au IXe siècle, il n'y avait pas de neumes ou de notes, mais uniquement les syllabes du texte chanté, qui montaient ou qui descendaient en fonction de la progression de la notation musicale).

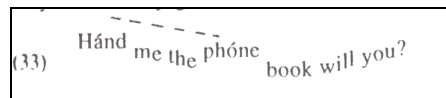


Figure n°31 : transcription de la mélodie par Bolinger (1998 : 51).

Tous ces systèmes de transcription ont été élaborés pour servir des visées théoriques bien précises. Mais est-ce que ces représentations se montrent aussi efficaces que la transcription musicale pour la lecture de la prosodie ?

⁹ « International Transcription System for Intonation ».

Pour le savoir, Fonagy et Bérard (1976) ont réalisé une étude pour comparer trois systèmes de transcription : la méthode graphique à cinq niveaux, les trois lettres H, M et B et la transcription musicale.

« Imaginons que les trois systèmes de transcription constituent trois canaux. Lequel des trois assure une meilleure transmission ? » (Guimbretière, 2000 : 157).

Pour le savoir, ils ont fait transcrire par des experts, dans chacun des trois systèmes, des énoncés assertifs, interrogatifs et répétitifs, des clichés mélodiques et des énoncés exprimant des attitudes spécifiques. Ils ont ensuite demandé à d'autres phonéticiens de reproduire la mélodie à partir de ces transcriptions. Les clichés mélodiques et les énoncés exprimant des attitudes spécifiques n'ont été correctement reproduits qu'à partir de la transcription musicale. Quant aux autres énoncés, la mélodie était aussi bien reproduite avec les trois systèmes. La transcription musicale s'avère donc le système le plus performant pour permettre l'oralisation du contour mélodique. Quant à la transcription par niveaux, on peut la considérer comme une variante de la transcription musicale, par rapport à des systèmes comme INTSINT et ToBI. La transcription musicale permet en outre une transcription « relativement » objective du contour mélodique. En effet, elle ne répond pas à une visée théorique spécifique de la prosodie comme c'est le cas pour INTSINT et ToBI. Avec ces systèmes de transcription, l'analyse est appliquée au moment même de la transcription. D'après Guaïtella (2000 : 173),

« L'intérêt d'une notation est de fournir le terrain le plus objectif possible pour permettre ensuite une interprétation plus libre. Si le système de codage se fonde sur des interprétations théoriques, alors l'analyse qui sera faite sur la base de ce codage présente une circularité dépendante du codage lui-même ».

¹⁰ « Tones and Break Indices ».

2.2. Lorsque la langue devient musique : Evolution du traitement de la similitude entre la langue et la musique au cours de l'histoire de la musique occidentale

« On pourrait se demander si la musique vocale ne représente pas un cas privilégié à l'intérieur de cette catégorie de systèmes, dans la mesure où elle unifie, en une temporalité unique, deux systèmes très différents. « Le langage est le plus vrai », disait Hegel, mais le langage isole, déplace, et à la limite je veux toujours autre chose que ce que je dis. « La musique exprime la pure vie intérieure », dit Boris de Schlæzer, mais elle est impuissante à nommer. On conçoit donc ce qu'il y a de séduisant dans une entreprise qui, en les combinant dans une fusion intime, grâce au truchement de leur organe commun, la voix, vise à donner l'illusion que, la béance qui est au cœur de l'un, l'autre viendra la combler, et réciproquement » (Ruwet, 1972 : 68).

Un compositeur pourra choisir de contraindre la ligne mélodique d'un texte parlé en accentuant la musicalité (fixation des intervalles et métricité du rythme) ou au contraire, la laisser évoluer plus « librement », comme elle le fait dans la parole naturelle. Cette dialectique constitue la trame de fond de l'histoire de la musique vocale occidentale, qui a toujours oscillé entre la priorité accordée à la mélodie (poème symphonique où le texte est absorbé par le niveau musical) et la priorité accordée à la prosodie (grégorien, récitatif), c'est-à-dire au texte. Pour Antonio Salieri, la musique doit primer sur la parole : « *primo la musica e poi le parole* » (1786). De même pour Mozart qui écrit que « *dans un opéra, il faut absolument que la poésie soit fille obéissante de la musique* » (lettre du 13 octobre 1781). Cependant, même chez Mozart, la musique doit savoir s'effacer au profit du texte dans le récitatif, de manière à ce que le spectateur puisse comprendre le déroulement de l'action. Chez un compositeur plus contemporain comme Debussy, l'écriture musicale est gouvernée par la musicalité de la langue. Il retrouve ainsi d'une certaine façon, près d'un millénaire après le chant grégorien, une musique du texte. Est-ce que la dialectique musique / texte tendrait à se résoudre dans la musique contemporaine ? C'est la question à laquelle nous essaierons de répondre dans les lignes qui vont suivre, mais quoiqu'il en soit, il est certain que son étude apporte un éclairage très subtil sur les rapports qu'entretiennent la langue et la musique.

2.2.1. Rapports texte et mélodie : information vs musicalité

2.2.1.1. Le chant grégorien, une musique du texte

Dans le chant grégorien, le texte revêt une importance fondamentale. Il s'agit en effet de transmettre les textes de la Bible aux fidèles par la lecture à haute voix et pour cela, l'intelligibilité des paroles doit être respectée. Cette exigence d'intelligibilité donne à la langue une place centrale. Le lecteur doit en connaître parfaitement les caractéristiques prosodiques s'il veut être compris par son public ; il doit prendre en compte la musique de la langue,

« ce chant latent (cantus obscurior) dont parle Cicéron, ou ce germe de musique (seminarium musices) selon l'expression de Martianus Capella » (Viret, 1987 : 38).

Bien qu'il subsiste quelques textes en grec, comme le « Kyrie Eleison », le latin d'église est la langue véhiculaire de cette liturgie. En conséquence, la musique appliquée à ces textes correspond à la musique de la langue latine, et notamment à son rythme :

« si ce texte influe sur la composition grégorienne au point de constituer son support essentiel, c'est qu'il ne se présente pas comme un simple assemblage de phonèmes que le compositeur traiterait à sa guise, mais bien comme un donné déjà investi par lui-même de caractères musicaux, principalement sous l'angle de la rythmique » (Viret, 1987 : 37).

Ce rythme d'essence linguistique évolue avec une certaine liberté et ne rentre pas dans un cadre métrique (cf. parole et périodicité sommaire, partie 1.2.1.3.). Ce cadre métrique est d'autant moins nécessaire que le chant grégorien est monodique, c'est-à-dire qu'il est chanté à cappella et sans accompagnement ; c'est un **plain-chant**. Il n'a donc pas besoin d'un rythme stable pour permettre de coordonner plusieurs voix entre elles, comme l'exige la polyphonie. Dans la langue, le rythme naît de la succession de syllabes accentuées et non accentuées (cf. définition du rythme, partie 1.2.1.). La rythmique du latin est organisée selon le retour des accents toniques qui affectent les mots. Pour cette raison, l'accent tonique est toujours indiqué dans les livres de chant.

Comment ces accents sont-ils traités dans le chant grégorien ? Viret (1987 : 39-40) nous donne un exemple du traitement des mots de 5 ou 6 syllabes. Ces mots peuvent être accentués soit sur l'antépénultième (mots proparoxytons), soit sur la pénultième (mots paroxytons). Les deux ou trois syllabes initiales préparent l'arrivée de l'accent, point culminant du mot.

Ce temps d'attente créé chez l'auditeur un sentiment de tension, qui ne peut s'apaiser qu'au moment où l'accent survient. Les syllabes atones qui suivent viennent conclure le mouvement et concourent ainsi à la détente survenue après l'arrivée de l'accent. L'accent constitue donc le « pôle rythmique » du mot et le schéma mélodique lui est subordonné : l'intonation monte avant l'accent, atteint son pic sur l'accent, puis redescend sur les syllabes atones terminales (figure n°32).

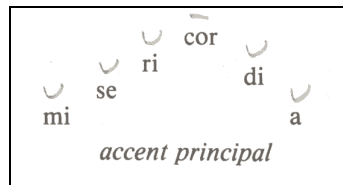


Figure n°32: schéma accentuel du mot « misericordia » (emprunté à Viret, 1987 : 40).

L'adjonction d'un accent secondaire sur la syllabe initiale peut venir complexifier la simplicité de ce schéma, par un dédoublement de la courbe (figure n°33), mais le pattern mélodique garde cependant sa configuration « en cloche ». Ces schémas de montée et de descente reposent sur les mêmes principes psychologiques que les schémas de tension et de détente dans la perception de la mélodie (cf. partie n°1.2.2).

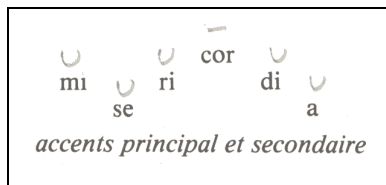


Figure n°33 : schéma accentuel du mot « misericordia » (tiré de Viret, 1987 : 40).

Selon Fraisse (1974 : 134),

« nul isochronisme n'apparaît. L'important reste un mouvement intensif et mélodique. Dom de Malherbe évoque, pour décrire ce rythme libre, les mouvements de la nature. « Il faut chanter comme une vague bondit ou s'étale, comme un champ de blé ondule, comme un arbre s'agit sous l'effort du vent, comme un fleuve roule ses eaux » (1946 : 24). Mais tous les restaurateurs du chant grégorien s'accordent pour invoquer le vol d'un oiseau comme la meilleure métaphore : après les syllabes préliminaires, la voix monte vers la syllabe accentuée, comme dans un élan, puis redescend vers le repos où se produit une nouvelle inspiration ».

La notation neumatique utilisée pour la transcription du chant grégorien ne comporte aucune indication de rythme, tout simplement parce que ce rythme est celui de la langue. Il suffit donc au chanteur de bien connaître la rythmique du latin.

Si le rythme du chant grégorien reste avant tout celui du mot latin, c'est sa mélodie qui va le faire évoluer vers une « musicalité » grandissante. Plusieurs facteurs concourent à cette évolution. Le premier d'entre eux concerne l'évolution du style du chant. Au départ, les chantres récitaient le texte sur une seule note, ce qui devait être relativement monotone pour les auditeurs. Mais, peu à peu, ils vont augmenter le ton aux endroits où la mélodie monte naturellement et le baisser, là où elle descend naturellement. Ces montées et ces descentes correspondent à des frontières linguistiques importantes. Comme pour le rythme, la ligne mélodique est donc calquée sur l'intonation de la langue latine. Le chant est syllabique (**extrait musical n°5** et partition correspondante, figure n°34), dans la mesure où on chante une note par syllabe, comme c'est le cas dans les antiennes (le chant qui sert d'introduction et de conclusion à la psalmodie, chant des psaumes).

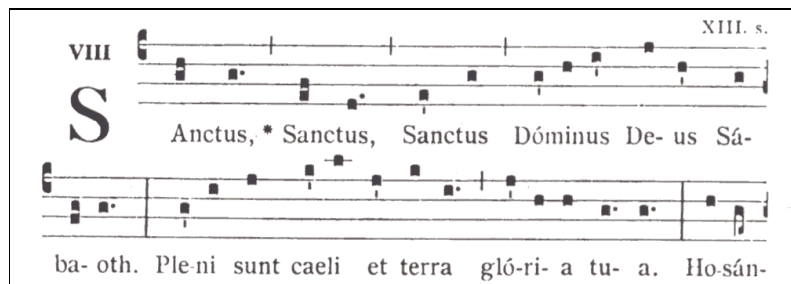


Figure n°34 : chant syllabique - Sanctus XIII, VIIIe Mode (tiré du livret d'accompagnement du CD « Chant Grégorien, Chant de Toujours », p : 15).

Mais peu à peu, la mélodie va se complexifier. Le nombre de notes par syllabes va augmenter pour donner naissance aux ornements, les **mélismes**. La figure n°35 nous en donne un exemple : sur la voyelle « e » de « Kyrie », la ligne mélodique se prolonge sur une douzaine de notes (**extrait musical n°6** et partition correspondante, figure n°35).



Figure n°35 : chant mélismatique - Kyrie XIII, IVe Mode (tiré du livret d'accompagnement du CD « Chant Grégorien, Chant de Toujours », p : 13).

Ce chant mélismatique nécessite une technique vocale beaucoup plus poussée que dans le chant syllabique, ce qui va transformer les moines lecteurs en véritables moines chanteurs. Les mélismes préfigurent déjà les **vocalises** de l'opéra italien. On les retrouve dans l'alléluia (notamment dans le « jubilus »), qui est un chant d'acclamation de louanges.

Mais, cependant, il y a une limite à cette ornementation : le chant grégorien doit resté compréhensible, car il est avant toute chose, une musique du texte. La compréhensibilité constitue donc un frein naturel à toute évolution plus radicale du chant grégorien.

« Rien de tout cela ne saurait éclipser le sens du texte liturgique revêtu d'une mélodie destinée à le commenter, l'éclairer, le mettre en valeur. C'est donc lui qu'il faut interroger d'abord... On ne saurait être trop catégorique sur ce point » (Dom Gajard, 1951 : 39¹¹).

Un autre facteur contribue à faire évoluer la mélodie et il est d'une importance capitale dans l'histoire de la musique occidentale : au début du chant grégorien, les hauteurs ne sont pas encore fixées par la notation musicale (cf. partie 1.2.2.2.). L'apparition d'une hauteur référence va permettre cette fixation, et par conséquence, la fixation des intervalles. Cette évolution de l'écriture pousse inéluctablement la mélodie vers une plus grande « musicalité » (cf. 1.2.4.) et concourt à creuser le fossé entre la langue et la musique.

2.2.1.2. Le récitatif dans l'opéra

2.2.1.2.1. A la frontière de la parole et de la musique

Progressivement, la musique vocale évolue vers la **polyphonie** et cette multiplication des voix va venir brouiller l'intelligibilité du texte. La polyphonie nécessite une rythmique plus métrique nécessaire pour que les différentes voix s'accordent entre elles, mais très éloignée de la liberté rythmique de la parole. Ces bouleversements éloignent de plus en plus la musique du texte, ce qu'illustrent les oeuvres de l'Ecole Franco-Flamande et notamment, celles de Roland de Lassus (1531-1594) et de Giovanni Palestrina (1525-1594). Chez ces compositeurs, l'équilibre de l'architecture polyphonique prime sur la compréhension des paroles et l'interprète perd son individualité. En fait, il n'a d'existence qu'au sein du groupe vocal. C'est en réaction contre cette complexification de l'écriture vocale qu'apparaît le **récitatif**.

A Florence, un cercle d'humanistes rassemblés autour du Comte Giovanni Bardi (Camerata Bardi) entreprend de rénover la musique vocale selon le modèle de la tragédie grecque classique. Selon eux, il faut retrouver l'intelligibilité du texte en revenant à la monodie :

¹¹ Le chant grégorien a été rénové au cours du XXe siècle principalement par les moines de l'Abbaye de Solesmes. La théorie de Dom Mocquereau sur le rythme libre musical s'illustre particulièrement et a été sa mise en musique par le maître de chœur de Solesmes, Dom Gajard dans les années 60.

« *il faut renoncer au contrepoint et revenir à la simplicité du mot* » (Vincenzo Galilei, Camerata Fiorentina).

Autrement dit, la musique doit se plier au texte, elle doit servir la poésie.

Jacopo Peri qui fréquente la Camerata Bardi sera le premier compositeur à appliquer ces préceptes dans les deux premiers opéras de l'histoire de la musique, « *Dafne* » (1597) et « *Euridice* » (1600). « *Euridice* » se caractérise par l'utilisation d'un nouveau style, le *recitar cantando* qu'on pourrait traduire par « *parler en chantant* ». Ce nom illustre parfaitement l'ambiguïté d'un style qui neutralise les oppositions habituelles entre la lecture et le chant. Non seulement, la musique se lie au verbe, mais elle devient celle du verbe : la ligne mélodique et l'écriture rythmique suivent fidèlement l'intonation et le rythme de la langue utilisée dans le texte. Comme dans le chant grégorien, l'interprète doit connaître les caractéristiques prosodiques de la langue qu'il emploie. En se transformant en orateur-chanteur, il recouvre la liberté qui est propre à la parole : c'est lui qui imprime le rythme, qui « conduit » l'accompagnement. Il n'y a plus besoin du chef d'orchestre car les instruments accompagnateurs s'alignent sur le chanteur. Le texte redevient central et avec lui, l'information qu'il véhicule. Un des problèmes majeurs dans « *Euridice* » vient justement du fait qu'il ait été presque entièrement écrit en *stile recitativo*. En effet, il souffre d'un manque de musicalité et d'expressivité.

Ce déséquilibre sera rectifié dans l'« *Orfeo* » (1607), le premier chef d'œuvre de l'opéra. Son auteur, Claudio Monteverdi n'utilise plus systématiquement le *recitar cantando*, mais il a également recours au contrepoint pour les parties symphoniques et au **madrigal** pour les parties chorales. L'**aria** n'apparaît pas encore, mais peu à peu, l'expression des sentiments va lui être confiée. La musicalité y reprend le dessus souvent aux dépens de la compréhension du texte. Ce tiraillement entre l'exigence d'intelligibilité nécessaire à la compréhension du drame et la musicalité structure toute l'évolution de l'opéra italien au cours du XVII^e siècle. Cette lutte ne prendra fin que lorsque la séparation de l'air et du récitatif sera consommée, c'est-à-dire à partir des années 1650-1660, et notamment dans l'œuvre d'Antonio Cesti (1623-1669). A partir de cette époque, l'aria et le récitatif ont des fonctions nettement délimitées dans l'opéra.

« *Dans l'opéra italien, la fonction du récitatif est de relier les airs en exposant les péripéties de l'action sous forme de monologue ou de dialogue* » (Massin et Massin, 1983 : 98).

Il doit donc être le plus intelligible possible, afin que les spectateurs puissent suivre le cours du récit. L'aria quant à lui permet au chanteur de donner libre cours à sa virtuosité et de montrer sa maîtrise de la technique vocale. Il constitue selon Massin et Massin (1983 : 97) « *la friandise des gourmets de la musique lyrique* ».

En France, c'est Jean-Baptiste Lully qui fait apparaître le récitatif dans la tragédie lyrique. Ce récitatif qu'on appelle « français » suit également le texte, mais se montre plus mélodique que son homologue italien. Au XVIII^e siècle, on trouve deux types différents de récitatifs : le **recitativo secco** qui est un récitatif simple, où la voix est accompagnée d'un clavecin ou d'une basse continue et le **recitativo accompagnato**. Ce dernier type de récitatif est accompagné par tout l'orchestre. Son rythme est moins libre que le recitativo secco, mais il se montre plus expressif et plus musical. Il se situe à l'intermédiaire du recitativo secco et de l'aria. Dans l'opéra bouffe dont le « *Don Giovanni* » de Mozart marque le sommet, le récitatif adopte la forme « *parlando* ». Son écriture est tellement souple et l'accompagnement tellement réduit (le plus souvent limité au clavecin) qu'il permet de restituer la rapidité du dialogue naturel, surtout dans les dialogues entre les personnages (**extrait musical n°11**, dialogue de Leporello et Don Giovanni, acte II, scène XI).

2.2.1.2.2. Trois récitatifs, trois langues, trois musiques

Comme le chant grégorien, le récitatif est une musique du texte qui montre une proximité maximale avec la parole. Certains types de récitatifs sont plus musicaux que d'autres, comme le recitativo accompagnato, mais ils se situent tous dans une zone frontière où l'on ne sait plus très bien distinguer la langue de la musique. Dans ce contexte, le problème de la musicalité intrinsèque des langues se pose avec une grande force. Il est évident que l'allemand ne chante pas comme l'anglais, qui ne chante pas comme l'italien. Les règles d'accentuation et les contraintes rythmiques sont différentes et chacune de ces trois langues possède sa prosodie spécifique. Comment vont-elles être « mises en musique » dans le récitatif ? Est-ce que leur musique intrinsèque va être respectée par le compositeur ? Pour le savoir, nous avons choisi trois récitatifs écrits dans trois langues différentes : ils sont tirés des œuvres « *Didon et Enée* » de Henry Purcell en anglais (**extrait sonore n°7**), les « *Passions de St Matthieu* » de Jean-Sébastien Bach en allemand (**extrait sonore n°9**) et « *Don Giovanni* » de Wolfgang Amadeus Mozart en italien (**extrait sonore n°11**). Le texte de chacun de ces récitatifs a ensuite été lu par trois locuteurs natifs anglais (**extrait sonore n°8**), allemand (**extrait sonore n°10**) et italien (**extrait sonore n°12**). Ces enregistrements nous permettront de comparer l'accentuation et l'intonation adoptées par le chanteur avec celles adoptées par le lecteur.

2.2.1.2.2.1. Anglais : « *Didon et Énée* » de Purcell

Didon et Enée est un opéra en trois actes créé par le compositeur anglais Henry Purcell à Chelsea, en 1682. Dans cette œuvre, Purcell utilise une forme de récitatif dont la ligne mélodique évolue plus souplesment que dans le *recitativo secco*. L'extrait intitulé « *Ruin'd ere the set of Sun ?* » est tiré de la scène de la grotte, qui prend place dans le deuxième acte de l'opéra et met en scène des sorcières et des enchanteurs. Le passage choisi concerne une des répliques des sorcières (**extrait sonore n°7** et figure n°36). La ligne mélodique est chantée, mais elle suit très fidèlement l'intonation de la langue (pour une description en détail du système prosodique de l'anglais, voir chapitre V). Ainsi, dans « *Trustly elf* », la ligne mélodique s'infléchit sur la syllabe atone « *ly* », comme elle le fait dans la langue (**extrait sonore n°8**). De même, la tendance de l'anglais à accentuer la syllabe initiale dans les mots lexicaux plurisyllabiques est respectée (« *MERcury* », « *TRUSTly* »). Quant à sa rythmicité accentuelle, elle apparaît moins au niveau de la partition, qu'au niveau de l'interprétation (figure n°37). Il est vrai que certaines syllabes accentuées sont allongées, comme « *elf* », « *self* », « *chide* » ou « *night* ». La valeur des notes correspondantes équivaut à une noire (un temps) au minimum, alors que pour les autres syllabes, les valeurs sont plus courtes, c'est-à-dire des croches (un demi-temps) ou des doubles-croches (un quart de temps). Cependant, les autres syllabes accentuées ne se démarquent pas sur la partition. C'est le chanteur qui les met en relief en les produisant avec une plus grande force articulatoire, comme dans « *As SENT from JOVE* ». Tous les mots lexicaux sont accentués et les syllabes atones sont réduites comme la syllabe « *ly* » de « *trustly* », ce qui n'apparaît pas non plus sur la partition. L'accompagnement instrumental suit l'accentuation linguistique et le rythme imprimé par le chanteur : les envolées du clavecin commencent toujours sur une syllabe accentuée ; quant à la ligne mélodique de la basse et des cordes, elle ne change que sur des syllabes accentuées. On observe un changement de rythme dans l'accompagnement à partir de « *sail tonight* » avec l'introduction d'un accent d'insistance sur « *all* » exprimé musicalement par un pic mélodique (montée d'une quarte pour aboutir au ré3 correspondant à la syllabe « *all* », puis descente d'une quinte, pour aboutir au sol3) et un allongement (croche pointée), suivie d'une imitation musicale des rires des sorcières. Dans tout l'extrait, il y a très peu d'ornementation, sauf sur les deux syllabes « *chide* » (appoggiature et trille) et sur « *charge* » (notes de passage). Ces deux syllabes étant « naturellement » allongées parce qu'elles portent l'accent, elles laissent le temps au chanteur de charger un peu la ligne mélodique. Chacune des autres syllabes de l'extrait correspond à une seule note, ce qui rend le texte très intelligible.

Si l'on compare cet extrait au texte lu par le locuteur anglophone (**extrait sonore n°8**), on relève une correspondance quasi parfaite dans la localisation des accents (voir tableau n°31). Tous les pics mélodiques correspondant aux syllabes accentuées dans le texte parlé, n'apparaissent pas dans le texte chanté. Cependant, le chanteur réalise les accents de mot en allongeant les syllabes.

Figure n°36 : partition de l'extrait, « Ruin'd ere the set of Sun ? », tiré de Didon et Aénée de Purcell (1682 : 32-33).

MUSIQUE (extrait sonore n°7)	PAROLE (extrait sonore n°8)
Sorceress¹² (extrait n°7) : But when they've done ↑, my trustly elf, In form of Mercury himself ↑, As sent from Jove , shall chide ↑↑ his stay , And charge him sail tonight ↑ With all ↑ this fleet away.	Sorceress (extrait n°8) : But when ↑ they've done ↑, my trustly elf, ↑ In form ↑ of Mer ↑cury himself ↑, As sent ↑ from Jove ↑, shall chide ↑ his stay ↑, And charge ↑ him sail ↑tonight ↑ With all ↑ this fleet away.

Figure n°37 : accentuation comparée des extraits chantés et parlés de « Ruin'd ere the set of Sun ? », les syllabes accentuées étant indiquées en caractères gras et les pics mélodiques par une flèche montante.

¹² **Traduction** : Mais, dès qu'ils en auront fini, mon fidèle lutin – Sous la forme de Mercure lui-même – Envoyé par Jupiter, lui reprochera de rester – Et lui ordonnera de faire voile ce soir avec toute sa flotte.

2.2.1.2.2 Allemand : « *Les Passions de St Matthieu* » de Bach

Le récitatif suivant est intitulé « *Da ging him der Zwölfen einer* » (**extrait sonore n°9**). Il est tiré des « *Passions de Saint Matthieu* » (appelée également Grande Passion), composées en 1727 par Jean-Sébastien Bach. L'extrait met en scène l'évangéliste et l'un des disciples de Jésus, Judas. L'évangéliste raconte que Judas est venu trouver les prêtres pour leur proposer de leur livrer Jésus, en échange de 30 pièces d'argent. La narration est interrompue par la question de Judas aux prêtres en style direct : « *Was wollt ihr mir geben ? Ich will ihn euch verraten* ». La ligne mélodique de l'Evangéliste est « chantée », mais l'intervention de Judas est très proche du langage parlé, surtout au début de la question : « *Was wollt ihr mir geben ?* », en recitativo secco. Par rapport au récitatif de Purcell, l'accompagnement est beaucoup plus sommaire. Dans la première intervention de l'évangéliste correspondant à la phrase « *Da ging hin der Zwölfen einer, mit Namen Judas Ischarioth, zu den Hohen Priestern und sprach* », les accords sont plaqués par la basse continue (ici, les cordes) et par l'orgue sur les syllabes qui en délimitent le début et la fin, c'est-à-dire « *Da* » et « *sprach* ». Quant à la syllabe « *rioth* », elle indique une frontière importante correspondant à la fin de la première proposition (voir partition, figure n°38).

11 * [RECITATIVO]

EVANGELIST

Then went one of the di-sci-ples, whose name was Ju-das Is-ca-ri-ot, to the chief. priests and—
 Da ging hin der Zwöl-fen ei-ner, mit Na-men Ju-das Is-cha-ri-oth, zu den Ho-hen prie-ster-n, und

Chorus I

Cont.
Org.

4 JUDAS

said, How much will ye give me if I de-liv-er Him to you? And they of-fered him thir-ty
 sprach: Was wollt ihr mir ge-ben? Ich will ihn euch ver-ra-ten. Und sie bo-ten ihm drei-ßig

EVANGELIST

7 sil-ver pie-ces. And from that time he sought op-por-tun-i-ty that he might be-tray Him.
 Sil-ber-lin-ge. Und von dem an such-te er Ge-le-gen-heit, daß er ihn ver-rie-te.

Figure n°38 : partition de l'extrait, « Da ging hin der Zwölfen einer », tiré des « Passions de Saint Matthieu » (1727, Bach, p : 36).

A partir de l'intervention de Judas, les accords se font plus fréquents et correspondent aux syllabes accentuées « will », « ra » de « verraten », « bo » de « boten » et « drei » de « dreissig ». Cette multiplication correspond à un rythme plus saccadé, plus proche de la parole naturelle. L'accompagnement colle donc totalement à la structure accentuelle du texte et ponctue le mouvement rythmique imprimé par les chanteurs. Par ailleurs, on observe une correspondance frappante (voir figure n°39) entre les syllabes accentuées de l'extrait chanté (**extrait sonore n°9**) et celles de l'extrait parlé par la locutrice allemande (**extrait sonore n°10**). Sur la partition, certaines syllabes atones, comme dans la parole, sont très compressées au niveau temporel (durée d'une double croche, c'est-à-dire un quart de temps), comme c'est le cas des syllabes « ihr » et « mir » dans « Was wollt ihr mir geben ? ». La rythmique est quasiment identique, c'est la rythmique spécifique de l'allemand. En ce qui concerne l'intonation, les pics mélodiques de l'énoncé chanté « Was wollt ↑ ihr mir ge-↑ben ? » sont également en parfaite correspondance avec l'énoncé parlé. Ce passage est justement en discours direct, comme si Bach avait voulu qu'il soit le plus proche possible de la parole. En revanche, dans le reste du texte, on observe des différences de localisation des pics

mélodiques avec le parlé : ils se répartissent à des endroits différents. Au niveau de l'interprétation, les deux chanteurs ne se permettent aucune fioriture, juste une petite note de passage sur « *riot* ». Ils s'en tiennent strictement à une note par syllabe ce qui donne au texte une intelligibilité maximale.

MUSIQUE	PAROLE
Evangelist : Da ging hin der Zwöl-↑ fen einer, mit Namen Judas Ischari-↑ oth, zu den Hohenprie-↑ stern, und sprach ↑↑ . Judas : Was wollt ↑ ihr mir ge-↑ ben ? Ich will ihn euch ↑↑ verraten. Evangelist : Und ↑ sie boten ihm drei-↑ ßig Silber-↑ linge. Und ↑↑ von dem an suchte er Gele-↑ genheit, daß er ihn verrie-↑ te.	Evangelist : Da ging ↑↑ hin der Zwölfen einer, mit Namen Ju-↑ das Ischarioth, zu den Hohenpriestern ↑, und sprach . Judas : Was wollt ↑ ihr mir ge-↑ ben ? Ich will ↑ ihn euch verraten. Evangelist : Und sie bo-↑↑ ten ihm drei-↑ ßig Silberlinge. Und von dem ↑↑ an suchte er Gele- ↑ genheit, daß er ihn ver-↑ riete.

Figure n°39 : accentuation comparée des extraits chantés et parlés de « *Da ging him der Zwölfen einer* », les syllabes accentuées étant indiquées en caractères gras et les pics mélodiques par une flèche montante.

2.2.1.2.2.3. Italien : « *Don Giovanni* » de Mozart

Le dernier extrait est tiré du célèbre opéra composé par Mozart « *Don Giovanni* » (Prague, 1787). C'est un opéra en deux actes, dont le livret a été écrit par Lorenzo da Ponte. L'extrait intitulé « *Ah ah ah ! ah, questa è buona* » prend place dans le deuxième acte, scène XI. Il relate une conversation entre Don Giovanni et son valet, Leporello (**extrait sonore n°11**). Le récitatif employé par Mozart est un Recitativo secco sous la forme parlando. Des trois récitatifs présentés, c'est celui qui restitue le mieux le naturel de la parole. La situation dans laquelle prend place ce dialogue exige une rapidité d'échange entre les deux personnages et une grande spontanéité dans l'expression des émotions (surprise, colère). Le spectateur doit « y croire » et c'est ce que Mozart s'est employé à faire. Le résultat est très réaliste. Le début du récitatif commence par l'interpellation de Leporello par Don Giovanni : « *Oh, Leporello !* », tout en douceur (la scène se passe de nuit et Don Giovanni veut se montrer très discret). La réaction de Leporello exprime la surprise « *Chi mi chiama ?* ». Ces deux phrases sont du parlé pur. La suite consiste en un très subtil dosage entre la déclamation et l'introduction d'un peu de mélodicité dans le phrasé. Dans la version sonore, l'accompagnement est quasiment absent et laisse une totale liberté aux deux interprètes. En effet, le violoncelle et la contrebasse plaquent des accords tous les 8 temps. Sur la partition (figure n°40), ces accords sont tenus pendant 8 temps (deux rondes liées), ce qui est un peu plus « lourd ».

Figure n°40 : partition de l'extrait, « Ah ah ah !ah, questa è buona », tiré de Don Giovanni de Mozart (1787 : 297)

Si on compare maintenant l'accentuation de l'extrait musical (**extrait sonore n°10**) à celle de l'extrait produit en version parlée par une locutrice italienne native (**extrait sonore n°11**), on remarque une très forte correspondance (figure n°41), comme dans le récitatif précédent.

MUSIQUE	PAROLE
Don Giovanni : Oh Leporello ! ¹³	Don Giovanni : Oh Leporello !
Leporello : Chi mi chiama ?	Leporello : Chi mi chiama ?
Don Giovanni : Non conosci il padron ?	Don Giovanni : Non conosci il padron ?
Leporello : Co-sì nol conosc-ci !	Leporello : Co-sì nol conosc-ci !
Don Giovanni : Come ? birbo !	Don Giovanni : Come ? birbo !
Leporello : Ah siete voi ? Scusate !	Leporello : Ah siete voi ? Scusate !
Don Giovanni : Cosa è stato ?	Don Giovanni : Cosa è stato ?
Leporello : Per cagion vostra io fui quasi accoppato.	Leporello : Per cagion vostra io fui quasi accoppato.
Don Giovanni : Ebben, non era questo , un onore per te ?	Don Giovanni : Ebben, non era questo , un onore per te ?
Leporello : Signor, vel dono !	Leporello : Signor, vel dono !
Don Giovanni : Via, via, vien qua , vien qua : che belle cose ti deggio dir !	Don Giovanni : Via, via, vien qua : che belle cose ti deggio dir !

Figure n°41 : accentuation comparée des extraits chantés et parlés de « Ah ah ah !ah, questa è buona » les syllabes accentuées étant indiquées en caractères gras et les pics mélodiques par une flèche montante.

¹³ Traduction en français : Hola, Leporello ! – Qui m'appelle ? – Tu ne connais pas ton maître ? – Poussais-je ne pas le connaître ! – Comment ? Coquin - Ah, c'est vous ? Excusez ! – Qu'y a-t-il eu ? – A cause de vous, j'ai presque été assommé – Eh, bien, ce n'était pas là un honneur pour toi ? – Monsieur, je vous en fait cadeau ! – Allons, allons, viens ici : que de belles choses j'ai à te dire !

Dans l'extrait chanté, l'accentuation de l'italien est donc respectée par l'interprète, mais elle apparaît d'une manière beaucoup plus marquée dans la notation musicale de la partition qu'elle ne l'était dans le récitatif de Bach. En italien, la syllabe pénultième des mots plurisyllabiques est accentuée et notamment marquée par un allongement. Dans la notation, les syllabes accentuées sont presque systématiquement allongées (croches, voir noires), alors que les syllabes atones sont réalisées sur des double-croches, là-aussi quasiment systématiquement. C'est cette alternance entre valeurs longues et valeurs courtes qui permet à la notation musicale de restituer si fidèlement le rythme de l'italien parlé (forme *parlando*). Dans le récitatif de Bach, on peut dire que le chanteur ajoute sa propre connaissance de la prosodie de l'allemand à l'interprétation, il doit donc connaître l'accentuation de l'allemand. Dans le récitatif de Mozart, l'interprète n'a à la limite pas besoin de connaître l'accentuation de l'italien puisque celle-ci est restituée presque parfaitement par la notation musicale.

La rythmicité et le système intonatif de chacune de ces langues ont été respectés lorsque le compositeur les a mis en musique dans le récitatif, ce qui assure au texte une intelligibilité maximale. Les différences prosodiques entre ces langues engendrent donc des mises en musique différentes. Des trois récitatifs, celui de Mozart restitue le mieux la parole et ses changements de débit selon les émotions qui affectent les personnages (rapidité d'élocution sous le coup de la colère par exemple). La notation musicale se trouve donc ici appropriée à la restitution de la prosodie et de l'effet « *parlando* ». Ce n'est plus le texte qui est mis en musique, mais la musique qui est mise en texte. On trouve un seul exemple dans la musique instrumentale « classique », qui ait été traité par le compositeur en **mode récitatif** pour imiter la parole naturelle : c'est le début de l'« *Ode à la Joie* » de Schiller, dans la « *Neuvième Symphonie* » de Beethoven où le violoncelle et la contrebasse préparent l'entrée du chanteur baryton en mode de récitatif (**extrait sonore n°13**).

2.2.1.3. La musicalité des langues dans la mise en musique

2.2.1.3.1. Langues musicales versus langues non musicales ?

Le chant grégorien est fondé sur le rythme et l'intonation de la langue latine. Il en est de même dans les trois récitatifs précédents dont la ligne mélodique et le rythme suivent le « chant » de l'anglais, de l'allemand et de l'italien. Il est donc impossible de traduire ces textes dans une autre langue sans en modifier la musique.

Le problème de la musicalité des langues et de la gestion de cette musicalité dans la composition est central dans la musique vocale. Est-ce que certaines langues se plient plus facilement à la mise en musique ? Est-ce qu'on peut passer d'une langue à une autre sans modifier la musique, ou au contraire, faut-il l'adapter à la nouvelle langue ? Ces questions semblent s'être souvent posées aux compositeurs, au cours des siècles.

La **Querelle des Bouffons** aborde ce problème de manière caricaturale. Elle débuta en 1752 et opposa les partisans de l'opéra italien (ou « *opera buffa* ») à ceux de l'opéra français. Jean-Jacques Rousseau, grand amateur d'opéra italien s'illustra dans cette querelle par l'ensemble des critiques qu'il porta à l'opéra français dans sa « *Lettre sur la Musique Française* » (1753), et notamment contre la musique de Jean-Philippe Rameau (1683-1764). L'une de ses critiques porte sérieusement à discussion. Il reprochait notamment à la langue française de ne pas être une langue musicale, contrairement à l'italien. Compositeur lui-même, il mis sa théorie en pratique dans « *Pygmalion* » (1770), œuvre composée en français, mais où « *les paroles et la musique, au lieu de marcher ensemble, se font entendre successivement* ».

Ce faisant, Rousseau soulève le délicat problème de la musicalité des langues. Il est vrai que chaque langue possède sa propre musique, mais de là à juger que certaines langues en soient totalement dépourvues, il y a un pas qu'un linguiste ne peut franchir. Il n'existe pas de langue « non musicale » ; en revanche, toutes les langues ont en commun de combiner différemment les trois paramètres que sont la hauteur, la durée et l'intensité. L'organisation prosodique qui résulte de cette combinaison peut nous paraître plus ou moins musicale. Par exemple, l'anglais se caractérise par de larges variations de hauteur et une accentuation plus marquée du point de vue acoustique que le français (voir chapitre V, partie 2.1.1.2), ce qui pousse un compositeur comme Claudy Malherbe à adopter cette langue pour la composition de son œuvre « *Locus* » (voir plus bas, partie 2.2.3.3.). Le français serait-elle donc une langue moins musicale que l'anglais ? C'est ce que sous-entend Stricker lorsqu'il reproche à Maurice Ravel de s'être appuyé sur la prosodie du français dans la composition de ses œuvres. Il met notamment en avant le problème de la faiblesse de l'accent tonique :

« Ravel s'est employé à calquer sa ligne mélodique sur celle de la phrase parlée, s'inspirant des faibles accentuations toniques de la langue française au point de subordonner l'ampleur de la ligne mélodique à cet impératif du texte. Là peut-être pourrait-on voir une limite trop définie au génie mélodique ravélien, car cette faiblesse de l'accent tonique qui caractérise le français bride toujours quelque peu une ligne vocale qui s'en inspirerait, en même temps qu'elle défie par là une véritable transposition musicale ».

Mais, on peut se demander à l'inverse si ce « génie ravélien » dont parle Stricker aurait pu exister dans une autre langue que le français ? Prenons une œuvre comme « *Pelléas et Mélisande* » de Claude Debussy (voir plus bas, partie 2.2.1.2.). Il est évident qu'une telle composition puise ses qualités musicales dans celles de la langue française. On peut même dire que la grande valeur artistique de cette œuvre résulte de la fusion parfaite entre la langue et la musique et « *Pelléas et Mélisande* » ne saurait exister en anglais. Non, décidément, les rapports entre le texte et la musique paraissent infiniment plus subtils. En fait, il semble que dans la musique vocale, le problème ne se réduise pas à déterminer si certaines langues sont plus musicales que d'autres, mais plutôt à savoir si le mariage entre paroles et musique est plus ou moins bien assorti. Pour arriver à une cohésion entre les deux systèmes, le compositeur doit prendre en compte les spécificités prosodiques de la langue dans laquelle il compose, car l'accentuation musicale ne doit pas entrer en contradiction avec l'accent de mot ou l'accent de phrase. Mais présenté comme cela, la prosodie paraît ne représenter qu'une source de contrainte pour le compositeur. Il faut souligner qu'elle lui apporte également une grande source d'inspiration pour l'écriture de la partie musicale. Pour mettre en harmonie la langue et la musique, outre ses qualités musicales, le compositeur doit donc aussi avoir l'âme d'un phonéticien.

2.2.1.3.2. Un obstacle à la traduction des œuvres musicales ?

Une œuvre vocale composée dans une langue spécifique n'est donc pas transposable dans une autre langue, puisque celle-ci chante différemment. Le compositeur est de ce fait obligé d'adapter son écriture musicale pour qu'elle soit en correspondance avec le nouveau texte ou alors, de retravailler le texte jusqu'au point où il puisse s'accorder avec la musique. Nous l'avons vu, traduire « *Pelléas et Mélisande* » en allemand ou en anglais implique de réécrire totalement la partition, ce qui exige un travail titanesque de la part du compositeur. C'est pourtant ce que fit Giuseppe Verdi (1813-1901) dans son « *Don Carlos* » (1867), en donnant deux versions de son œuvre, une en italien et l'autre en français.

Annie Prassoloff (mars 1999¹⁴) souligne le « *tour de force technique, une exploration du génie des deux langues, de leurs vertus et de leurs limites* » réalisé par le compositeur. Pour illustrer les différences entre les deux langues, elle prend pour exemple le grand air de Philippe II et le compare dans les deux versions :

« le français donne au grand air de Philippe II une sorte de nudité aggravée par la gaucherie de l'accentuation : "Elle ne m'aime pas...", alors que l'italien magnifie le lyrisme désolé de l'aveu en mettant au premier plan le maître mot et ses voyelles pleines : "Amor per me non ha" ».

Pour Verdi, la traduction du texte impose donc une mise en musique différente.

Pourtant, il se trouve des auteurs qui nient le problème de la musicalité des langues, donc de la traduction. Selon Boris de Schlœzer, les paroles se dissolvent complètement dans la musique :

« Le sens des paroles chantées n'est plus le même que celui qu'avaient ces mêmes paroles avant leur mise en musique... mais celui que leur confère la phrase musicale » (19 ?? : 273).

Il apporte pour preuve que, dans l'histoire de la musique, on trouve des textes identiques mis en musique différemment et inversement, une même musique servie par des textes différents. Mais, si le choix du texte n'a aucune importance, alors « *pourquoi donc les musiciens ont-ils toujours mis tant de soin à choisir leurs textes et leurs livrets, voire à les composer eux-mêmes ?* » C'est la question que se pose Ruwet (1974 : 42-43) en commentant la position radicale de Schlœzer :

« En lisant Boris de Schlœzer, on a souvent l'impression qu'à écrire de la musique vocale, on ne risque que des inconvénients, des malentendus, des limitations. Pourquoi les affronter si on ne doit pas aussi y gagner quelque chose de plus fondamental ? ».

Même dans le récitatif secco sous sa forme parlando, la musique ne s'efface jamais complètement devant la langue et inversement dans le cas extrême de l'aria. Selon Ruwet (1972 : 52),

¹⁴ Texte non publié, disponible sur le web à l'adresse suivante : <http://www.chez.com/annieprassoloff/articles/texte-et-musique.htm>.

« dans la mesure où la voix est, pour l'homme, avant toute chose, l'organe de la parole, dès qu'elle intervient en musique, le langage, comme tel, est présent, et cela même si le chant s'émancipe en de purs mélismes, même si le texte devient totalement incompréhensible. [...] Il semble qu'il soit impossible de voir dans la voix un instrument comme les autres ».

2.2.2. La crise de la mélodie ou le retour à l'intonation

Grâce à son écriture (partie 1.2.2.2.), la musique a progressivement fixé les intervalles et par conséquence, donné naissance à l'harmonie, dont l'intonation mélodique est la forme la plus élémentaire. **Le développement de l'harmonie s'est fondé sur la hiérarchie tonale, c'est-à-dire sur la hiérarchie entre les 7 degrés de la gamme** (partie 1.2.2.4.). Mais, peu à peu, le système tonal classique va pouvoir s'élargir. Pour des raisons techniques (facture des instruments, mise au point du diapason), il n'a pas été possible d'utiliser toutes les tonalités avant le XVIII^e siècle. La mise au point de la gamme tempérée (division de l'octave en 12 demi-tons égaux) a permis au système tonal d'atteindre sa plénitude, mais l'a conduit également à la saturation : l'utilisation permanente des 12 notes de la gamme chromatique et des 24 tonalités possibles (12 tonalités majeures et 12 tonalités mineures) a contribué à faire disparaître toute hiérarchie entre les tons (cf. chromatisme de Wagner). Cette délitescence de la hiérarchie est à l'origine d'une perte progressive de toute conscience de tonalité chez l'auditeur. Cette crise est également celle de la mélodie, dont la perception est fondée sur la hiérarchie tonale. Après tout, la spécialisation sur l'intervalle n'était qu'une « *des facettes parmi bien d'autres de la prosodie du langage* » (Menezes, 1993 : 122). Cette crise était donc irréversible et devait déboucher sur une remise en cause complète des lois tonales. L'horizon des compositeurs se trouvait en effet complètement bouché s'ils persistaient dans la voie sans issue de la tonalité (qui avait déjà connu son apogée, donc n'offrait plus d'exploration possible). Parmi les alternatives possibles, on assiste à la création de nouveaux modes non fondés sur la tonalité (gammes par tons entiers, Debussy) et à la généralisation du chromatisme, qui va déboucher sur un bouleversement du langage musical, l'atonalité. Certains compositeurs veulent renouveler le matériau sonore et exploiter toutes les ressources que permet la voix humaine. Avec Debussy, on assiste à un retour à la musicalité intrinsèque de la langue française, c'est-à-dire à l'intonation. Mais, c'est Schoenberg qui va véritablement bouleverser les règles en inventant le Sprechgesang, nouveau type de déclamation à la frontière entre le chanté et le parlé.

2.2.1.2. Debussy et la musicalité de la langue française

Debussy veut libérer la mélodie de sa soumission à l'harmonie. Pour y parvenir, il va mettre au point une gamme par tons entiers où l'octave est divisée en 6 tons égaux. Dans une telle gamme, il n'y a plus de hiérarchie possible. Par ailleurs, il n'utilise plus les accords en fonction de la progression harmonique de l'œuvre, mais bel et bien pour leur beauté intrinsèque. Du coup, la mélodie ne peut plus être traitée en fonction des schémas traditionnels de tension et de détente. Debussy revient donc à la musique « naturelle » de la langue. Ainsi, dans son opéra « *Pelléas et Mélisande* » (1902), l'écriture musicale respecte le phrasé et l'accentuation de la langue du texte, le français. Le style vocal est très proche du récitatif et la mélodie connaît peu d'écarts d'intervalles. Dans l'extrait intitulé « *Devant le Château* » (Acte I, scène 3), Debussy met en scène la première rencontre entre Pelléas et Mélisande (**extrait sonore n°14**). Les personnages regardent le port et les bateaux s'éloigner, dont celui qui a transporté Mélisande jusqu'au château. La déclamation des chanteurs et l'orchestration suivent le rythme naturel du français : l'allongement final (figure n°42, syllabes accentuées indiquées en gras) et la rythmicité syllabique du français sont respectés. « *La fusion du son et du verbe est telle que texte et musique semblent n'avoir qu'un seul et même rythme* ». Quant à l'orchestration, elle suit même à certains endroits les inflexions de l'intonation : par exemple, pendant l'intervention de Mélisande, « *Je ne sais si nous pourrons le voir, il y a encore une brume sur la mer* ». Debussy fait chanter ses personnages « *comme des personnes naturelles et non pas dans une langue arbitraire, faite de traditions surannées* ». Les voyelles ne sont plus privilégiées comme elles l'étaient dans la musique tonale. La ligne mélodique s'y attardait en faisant fleurir des ornements. Mais, à partir du moment où on revient à la musique « naturelle » de la langue, on retrouve l'équilibre des sons propre à cette langue, notamment les voyelles et les consonnes. L'auditeur est ainsi placé devant la musicalité de sa propre langue (cf. Menezes, 1993 : 89). Debussy conserve cependant quelques petites différences : le e caduc ne chute pas en finale (par exemple « *haut-es* » et « *navi-re* ») comme il devrait le faire dans la langue parlée. Ce maintien engendre un dédoublement de la syllabe et permet au compositeur d'introduire une note supplémentaire, qui « musicalise » un peu plus le phrasé.

M : Quelque-cho-se sort du **port**.
 P : Il faut que ce soit un grand navi-re, les lumiè-res sont très **hau**-tes. Nous le verront tout-à-l'**heure**, quand il entrera dans la voile de clarté.
 M : Je ne sais si nous pourrons le **voir**, il y a encore une **bru**-me sur la **me**-re.
 P : On dirait que la **bru**-me s'élè-ve **lentement**.
 M : Et j'aperçois là **bas** une petite **lumiè**-re que je n'avais pas **vue**.
 P : C'est un **pha**-re. Il y en a **d'au**-tres que nous ne voyons pas **encore**.
 M : Le **navire** est dans la **lumiè**-re ; il est déjà bien **loin**.
 P : Il s'éloigne à toutes **voiles**.

Figure n°42: accentuation de l'extrait « *Devant le Château* », correspondant à l'extrait sonore n°14, « *Pelléas et Mélisande* », de Debussy.

2.2.2.2. Le « Sprechgesang »

Nous l'avons vu, l'évolution naturelle du système tonal le pousse inéluctablement vers la saturation et l'une des possibilités offertes aux compositeurs consiste à généraliser l'utilisation du chromatisme dans l'écriture. C'est ainsi qu'Arnold Schoenberg va créer la **musique dodécaphonique**. Afin d'éviter la formation de centres tonals, il met au point une technique sérielle où chacun des 12 demi-tons ne réapparaîtra pas avant que les 11 autres ne soient apparus (variation continue). Ce principe est très contraignant pour la composition car seule la répétition immédiate d'une même note est permise. Mais, grâce à cette technique sérielle, toutes les tonalités peuvent coexister de manière simultanée. Comme Debussy, Schoenberg doit repenser la mélodie, car celle-ci ne peut plus exister sous sa forme traditionnelle dans un système atonal. Comme lui, il va revenir à la musicalité de la langue, mais il ne va pas l'utiliser de la même façon. Dans son « *Pierrot Lunaire* » (1912), il met au point un mode de déclamation qui se situe à la frontière du chanté et du parlé : le « Sprechgesang ». A cause de son étrange statut, on pourrait le traduire indifféremment par « parole chantée » ou par « chant parlé ». En effet, il est difficile de déterminer s'il tient plus du chanté que du parlé et inversement. Les limites habituelles entre langue et musique sont transgressées d'une manière beaucoup plus forte que dans les formes de récitatif précédentes. Le procédé de notation adopté par Schoenberg traduit cette fluctuation du Sprechgesang entre la langue et la musique : le rythme est donné, mais en revanche, les hauteurs ne sont pas indiquées de manière précises (une croix remplace le cercle de la note, pour suggérer que la hauteur n'est pas fixée, qu'elle est plus haute ou plus basse que celle qui est indiquée par la croix). Par ailleurs, avant Schoenberg, la musique vocale se limitait somme toute à deux types d'émission opposés : le parlé et le chanté, alors qu'il existe toute une palette d'intermédiaires entre ces deux extrêmes.

Schoenberg, lui, ne se limite pas à ces deux modes de production vocale : en plus du parlé et du chanté, il utilise également les rires, les soupirs, les chuchotements, les gémissements, les plaintes... Le Sprechgesang constitue une réelle tentative d'exploration de toutes les possibilités d'expression qu'offre la voix. Non seulement Schoenberg les explore, mais surtout il les mélange. Ainsi, selon Gagnard (1987 : 14),

« il introduit volontiers un son de hauteur précise chanté au milieu d'une phrase en parlando. Toute cette technique de Sprechgesang est basée sur les élans et les retombées, mais ce n'est pas aussi simple, car il mélange constamment les genres d'émission, par exemple le rire (à demi) et le chanté ».

A titre illustratif, nous avons choisi un extrait non pas du « *Pierrot Lunaire* », mais de « *Ode to Napoleon Buonaparte opus 41* », (1942) (**extrait sonore n°15**), car le Sprechgesang y est très affirmé. Le texte en anglais provient d'un texte de Lord Byron (voir texte et traduction en français, figure n°43). La mélodie et le rythme principal sont assurés par la voix du baryton : c'est lui qui conduit l'interprétation. Il n'y a plus de pulsation rythmique. Dans ce passage, la musique n'a plus qu'une fonction d'accompagnement : les cordes suivent, ponctuent et illustrent la « narration ». Le début de l'extrait oscille entre le parlé (« *than littleness* »), le scandé (« *ambitions* ») et le légèrement chanté (« *mortals guess* »). On trouve des phrases entièrement parlées, comme « *it will **teach** to **after-warriors** **more** than **high** philosophy can **preach**, and **vainly** preach'd before* » avec un respect scrupuleux de l'accentuation et de l'intonation anglaise (accent sur la deuxième syllabe de « *philosophy* » par exemple). Un peu plus loin, le rythme est tellement scandé sur la phrase « *Those **pagod** **things** of **sabre** **way*** », qu'une pulsation est réintroduite. Cet extrait est suivi quelques mots plus tard d'une note chantée avec une hauteur fixée sur « *Brass* », puis d'un passage parlé avec « *and feet of clay* ». La suite consiste en une pure déclamation théâtrale « *The triumph and the vanity, the rapture of the strike* », qui englobe le début de la phrase suivante « *the earthquake voice of* », pour basculer brutalement dans du chanté sur le dernier mot « *Victory* », et pas n'importe quel chanté. En effet, le piano a commencé quelques instants plus tôt une citation musicale évoquant le début du premier mouvement de la *Cinquième Symphonie* de Beethoven et la voix le rejoint sur « *Victory* ». On retrouve l'espace d'un mot le statut traditionnel de la voix chantée. Sans arrêt, Schoenberg alterne les modes vocaux, de manière à ne jamais laisser la monotonie s'installer, l'intrusion de passages chantés dans le parlé en est un exemple. L'oreille de l'auditeur doit toujours rester en alerte, sans jamais pouvoir anticiper ce qui va suivre.

Ce mélange continu contribue encore davantage à abolir les limites entre le parlé et le chanté.

Nor till thy fall could mortals guess	C'est depuis ta chute que les mortels connaissent
Ambitions less than littleness !	Que l'Ambition est moindre que la petitesse !
Thanks for that lesson – it will teach	Pour la leçon, merci – car par elles apprendront
To after-warriors more	Les guerriers de l'avenir
Than high philosophy can preach,	Plus que des hauts prêches de la Philosophie,
And vainly preach'd before.	Qu'avant elle prêchait en vain.
That spell upon the minds of men	Ce charme qui pesait sur les esprits des hommes
Breaks never to unite again,	Est brisé pour jamais pouvoir se restaurer,
That led them to adore	Qui les menait à adorer
Those pagod things of sabre way,	Ces êtres fabuleux de l'empire du sabre,
With fronts of brass, and feet of clay.	Dont les fronts sont d'airain, dont les pieds sont d'argile.
The triumph, and the vanity,	Le triomphe et la vanité, dans le combat
The rapture of the strike –	Le ravissement, et la voix
The earthquake voice of Victory,	De la Victoire tel un tremblement de terre,
To thee the breath of life ;	Qui te furent souffle de vie ;
The sword, the sceptre, and that sway	Et le glaive, et le sceptre, et ce pouvoir par quoi
Which man seem'd made but to obey	L'homme ne semble fait rien que pour obéir,
Wherewith renown was rife –	Dont le renom régnait partout,
All quell'd ! –	Tout cela est dompté !

Figure n°43 : texte en anglais et traduction en français de l'extrait de « Ode to Napoleon Buonaparte, opus 41 » correspondant à l'extrait sonore n°15.

2.2.3. Musicalisation de la parole dans la création contemporaine

« Une attitude plus ambitieuse consiste à proposer entre toutes, la musique comme une activité « globalisante », comme une interdiscipline proprement dite, une activité qui, recoupant de multiples disciplines spécifiques, vérifie par synthèse leurs apports partiels, tant sur le plan des faits que sur celui des idées, et se présente au même titre qu'elles comme une activité de découverte, qui vise autant, sinon plus, à fonder une connaissance qu'à créer des œuvres », Pierre Schaeffer (1966 : 31).

2.2.3.1. La langue traitée comme matériau

Selon l'Encyclopedia Universalis, la notion de matériau recouvre en musique

« tout ce qui sert de matière première à l'écriture, c'est-à-dire l'harmonie, l'intervalle, le timbre, le rythme, la forme » (Fréron, 2002).

Or, toute cette matière première est élaborée à partir de la combinaison des paramètres acoustiques que sont la hauteur, la durée, l'amplitude et le timbre. Ces paramètres se trouvent agencés selon une syntaxe spécifique, qui était celle de la tonalité jusqu'au XX^e siècle. La remise en cause de cette syntaxe par des compositeurs comme Schoenberg et Berg entre autres a permis la naissance de l'atonalité. Au milieu du XX^e siècle, une autre révolution va venir complètement modifier le rapport du compositeur au matériau musical, la naissance de l'électro-acoustique. Grâce à elle, de nouveaux instruments voient le jour comme les ondes Martenot par exemple en 1928. Certains instruments traditionnels s'électrifient et sont amplifiés (guitare, violon, etc.). Mais, ce sont surtout les techniques d'enregistrement et d'analyse du son qui connaissent des progrès considérables, avec, pour conséquence, une modification directe du rapport entre le compositeur et le matériau. Celui-ci peut enfin enregistrer, manipuler les sons, les découper et les isoler, les coller et les assembler, les modifier, les créer ex nihilo... Il peut ainsi se transformer en ingénieur du son, ce qui implique un rapport complètement différent à la musique. Alors qu'il avait besoin d'un orchestre pour la mise en musique de ses compositions, il peut dorénavant créer ses œuvres au sein même du studio. Il n'a plus forcément besoin d'un orchestre, mais peut réutiliser des enregistrements ou incorporer un instrument électronique à un orchestre traditionnel (utilisation de potentiomètres par exemple). Ainsi, Pierre Schaeffer, qui est également ingénieur, utilise des sons concrets, c'est-à-dire des sons enregistrés qu'il assemble dans ses compositions ; quand à Edgard Varèse, il sera le premier à coupler des sons « organisés » à un orchestre dans une œuvre qui scandalisera le public, « *Déserts* » (1954).

Grâce à ces techniques, la parole jusque là éphémère, se transforme en objet fixé, donc reproductible, que l'on peut transformer à volonté par le découpage, le collage et la synthèse. Autrement dit, on peut lui faire subir toutes les distorsions que l'on veut. Cette possibilité est en soi une petite révolution, car elle entraîne un changement de statut de la langue. En passant de la catégorie des « intouchables » à la catégorie des « transformables », la langue devient accessible, elle devient un simple matériau. Ce changement de statut entraîne de nouvelles réflexions sur la langue et la musique et spécialement sur les rapports qu'elles entretiennent. Toutes deux peuvent en effet servir de matériau au compositeur qui peut les manipuler comme bon lui semble. En partageant le même statut, elles deviennent solidaires jusqu'au point où dans certaines œuvres comme celles de Pierre Boulez et Luciano Berio, l'opposition musique/langage se voit supprimée.

2.2.3.2. Neutralisation de l'opposition langage/musique chez Berio

Debussy avait largement amorcé le retour à l'intonation dans *Pelléas et Mélisande*, mais pour Menezes (1993 : 89), c'est Luciano Berio qui

« constitue peut être l'exemple le plus singulier de toute sa génération d'une prise de conscience radicale de ce retour à l'origine de la musique en tant qu'intonation de la parole ».

La « *Sequenza III* » pour voix de femme a été créée par Berio en 1966 au Studio di Fonologia de la RAI à Milan. Le texte en anglais de Markus Kutter a perdu toute intelligibilité lors de sa mise en musique. Il est très difficile de l'identifier à moins de l'avoir sous les yeux : « *Give me a few words for a woman to sing a truth allowing us to build a house without worrying before night comes* ». Berio procède en effet à une instrumentalisation de la voix, en découpant les mots en syllabes, en phonèmes. Ainsi, « *la parole est secouée jusqu'à ses racines phonétiques* » (Berio, cité par Gagnard, 1987 : 49). L'**extrait sonore n°16** est une caricature de la parole grâce à la parole elle-même. La voix prononce un discours en charabia, ponctué par des notes aiguës et intenses, puis s'élève dans un simulacre de rire. L'extrait se termine par une suite de trois notes tenues, dont la première et la troisième sont émises bouche ouverte et la deuxième bouche fermée. Le texte est découpé, pour être totalement recomposé au sein d'une œuvre musicale originale.

« Pour contrôler un ensemble aussi vaste de comportements vocaux, il a fallu morceler le texte et le détruire en apparence, afin de récupérer les fragments sur divers plans expressifs et recomposer l'ensemble en unités non plus discursives mais musicales » (Berio, 1998, livret d'accompagnement du CD).

Du coup l'opposition musique/langage n'existe plus, car la parole est complètement musicalisée. Cette absence d'opposition se reflète dans la partition où la notation musicale cohabite avec la transcription phonétique (API) et la transcription orthographique (figure n°44).

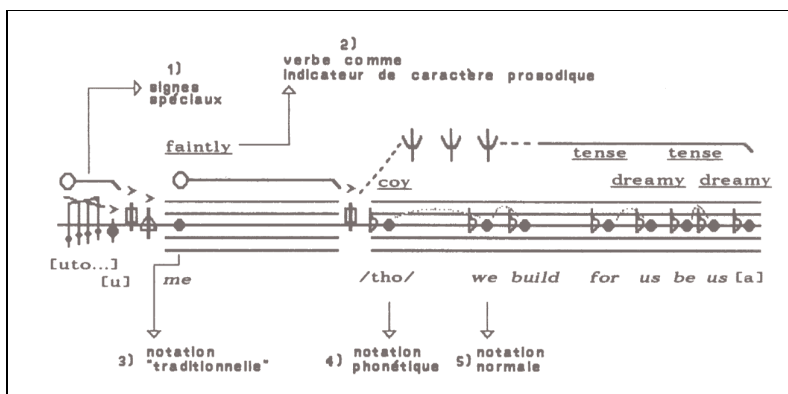


Figure n°44 : extrait d'une transcription « Sequenza III » (page 1) (emprunté à Menezes, 1993 : 124).

Par ailleurs, Berio a recours au pouvoir expressif des mots pour indiquer les caractères prosodiques que le chanteur doit appliquer au texte (figure n°45) : « anxieux », « tendu », « noble » ou encore « extatique ». Bien évidemment, la notation musicale ne peut transmettre de tels états. Le problème, c'est que la notation est surchargée et très difficile à déchiffrer par les interprètes. Ce manque de lisibilité pose les limites de la notation musicale classique, aussi bien que celles de la transcription orthographique et phonétique. Alors que les phonéticiens ont tenté de trouver des systèmes de transcription capables de restituer la prosodie (partie 2.1.1.2.), il est frappant de constater que la musique contemporaine se heurte au même problème, dans la mesure où elle revient à l'intonation et au rythme langagier. Cette problématique commune résulte de l'effacement des limites entre langue et musique.

Exemple 1a: caractères prosodiques dans "Sequenza III"

anxious	joyful
apprehensive	languorous
bewildered	muttering
calm	nervous
coy	nervous L.
distant	noble
distant and dreamy	open L.
dreamy	relieved
dreamy and tense	serene
echoing	subsiding
ecstatic	tender
extremely intense	tense
extremely tense	tense L.
fading	tense muttering
faintly	urgent
frantic	very excited and frantic
frantic L.	very tense
gasping	walking on stage
giddy	whimpering
impassive	whining
increasingly desperate	wistful
intense	witty

Figure n°45 : caractères prosodiques dans la « Sequenza III » de Berio (1966) (emprunté à Menezes, 1993 : 123).

2.2.3.3. Les « musiciens phonéticiens »

Les techniques de traitement du son n'ont pas bénéficié uniquement à la musique, mais également aux sciences phonétiques et il existe dorénavant une sphère commune entre le musicien et le phonéticien, par l'intermédiaire de l'analyse du signal. Par conséquent, il n'est guère étonnant que beaucoup de compositeurs se soient intéressés à la linguistique, par exemple Berio à la phonologie. Dans la création contemporaine, on trouve même des musiciens tels que Claudy Malherbe et Luca Francesconi qui fondent leurs compositions sur une analyse phonétique : ils musicalisent la parole au sein d'une nouvelle discipline que l'on pourrait s'amuser à surnommer la phonémusique. Mais, si beaucoup de musiciens se passionnent pour la langue, la réciproque n'est pas forcément vraie et les travaux réalisés par des phonéticiens (phonéticiens-musiciens) sur les rapports entre langue et musique restent encore peu nombreux (à l'exception notable des travaux de Fonagy sur l'expression des émotions dans la langue et dans la musique).

L'IRCAM¹⁵, fondé par Pierre Boulez en 1975, est un lieu de recherche multidisciplinaire sur les techniques de composition contemporaines. Le programme « *Les noces de la voix et de l'électronique* »¹⁶ regroupe des compositeurs qui travaillent à l'extension des techniques vocales. Parmi eux, Claudy Malherbe compose l'harmonie et la mélodie de « *Locus* » (23 juin 1997, Espace de projection de l'IRCAM) pour quatuor vocal et dispositif électronique, à partir de l'analyse acoustique de phrases en anglais. Il a préféré l'anglais au français à cause de sa musicalité intrinsèque (plus d'accentuation et variations mélodiques plus étendues). Les phrases qui composent le texte sont tirées de manuels de phonétique et d'acoustique,

*« l'ensemble fournissant à la fois un matériau vocal ainsi qu'un commentaire qui donne, au fur et à mesure du déroulement, la description des éléments et des situations en jeu (l'appareil vocal : larynx, gorge, langue, lèvres... ; respirer, parler, manger »*¹⁷.

L'écriture musicale a été élaborée directement à partir de l'analyse acoustique et phonétique de ces phrases. Le compositeur a procédé de la manière suivante : il a commencé par segmenter les phrases en syllabes, puis a réalisé un lissage en moyennant les différents partiels à l'intérieur de chaque syllabe. Il obtient ainsi des bandes de fréquence amplifiées dont l'évolution est statique et les phases de transition effacées.

¹⁵ IRCAM : Institut de Recherche et de Coordination Acoustique/Musique, Centre Georges Pompidou, Paris.

¹⁶ Programme exposé sur le web : <http://sol.ircam.fr/instruments/musicweb.fr>

¹⁷ Commentaires et présentation de l'œuvre de Claudy Malherbe disponibles sur le site web de l'IRCAM.

Toutes ces bandes sont converties en valeurs musicales de manière à former un accord (ou agrégat) correspondant à chaque syllabe. Ces accords sont ensuite reportés sur une double portée musicale (clé de fa/clé de sol). L'harmonisation de « *Locus* » est donc établie à partir de la répartition spectrale de la parole naturelle.

Dans « *Etymo* »¹⁸ (25 novembre 1994, Espace de projection de l'IRCAM, Paris) pour soprano, ensemble et dispositif électronique, Luca Francesconi utilise des extraits de poèmes de Baudelaire (« *Le Voyage* », « *L'Albatros* » et « *Carnets Intimes* », 1862). Pour lui, le mot est musique ; il est donc allé recueillir cette musique en procédant lui aussi comme un phonéticien : il a demandé à la soprano de lire ces textes, puis il a réalisé une analyse spectrale des voyelles et des consonnes, ainsi qu'un relevé de la durée des pauses et des différents segments. « *Etymo* » a été composé à partir de ces mesures. Le texte de l'extrait présenté (**extrait sonore n°17**) est fondé sur le retour des sons [μ] et [ε] : « *Même dans nos sommeils, la curiosité nous tourmente* » (partition, **figure n°46**). Les accords sont composés à partir de la répartition spectrale des voyelles.

The image shows a musical score for a vocal piece. It consists of four staves, each labeled 'Voix' (Voice) on the left. The lyrics are written below the staves. The first staff has the lyrics 'me me me me' with a '3' above the first 'me'. The second staff has the lyrics 'me dans nos som me' with a '3' above the first 'me' and 'à la' above the 'nos'. The third staff has the lyrics 'me dans' with a '3' above the first 'me'. The fourth staff has the lyrics 'nos som meil la cu - ri - o - si - té nous tour-meil-te' with '3' above the first 'nos', 'som', and 'meil'. There are various musical notations including notes, rests, and dynamic markings like 'p' and 'f'.

Figure n°46 : extrait de « *Etymo* » de Luca Francesconi (1994).

2.2.3.4. Typologie des effets : « langue dans la musique »

La musique contemporaine n'est pas un champ clos, dont les expérimentations et les avancées resteraient cantonnées à sa propre sphère d'influence, à ses propres compositeurs et à son propre public.

¹⁸ Commentaires et présentation de l'œuvre de Lucas Francesconi, disponibles sur le site web de l'IRCAM.

Elle exerce son ascendant sur d'autres courants musicaux, plus « populaires ». Ce fut le cas durant les années 60-70 avec la musique pop, qui fut également un terrain d'expérimentation très fertile pendant cette période. Ainsi, un musicien comme Franck Zappa était fasciné par les compositions d'Edgard Varèse et il réalisa beaucoup d'expérimentations, notamment avec son groupe, les *Mothers of Invention*. L'électroacoustique fut largement utilisée et donna naissance à l'instrument phare de la pop : la guitare électrique. Aujourd'hui, la musique électronique fournit également un champ d'investigation fructueux. Il est donc très important de ne pas se cantonner à un seul type de musique, ni bien sûr à une seule culture musicale. Cependant, plutôt que de réaliser une étude qui fasse un inventaire complet sur le sujet et qui serait trop peu en rapport avec notre sujet, il semble plus pertinent d'établir une typologie des différents procédés qui permettent d'incorporer la musique de la langue dans la musique. Cette typologie ne se prétend pas exhaustive, mais jette les bases pour la réalisation d'un catalogue plus complet. Elle est constituée de trois grandes classes : la première englobe toutes les utilisations « naturelles » de la langue dans une composition musicale, imitation ou incorporation sans modifications du donné langagier ; la deuxième comprend toutes les compositions musicales élaborées à partir d'une combinaison spécifique d'éléments prosodiques. La troisième concerne les compositions élaborées à partir d'une modification du signal avec découpage, collage et/ou synthèse (modification par traitement du signal sonore).

Utilisations naturelles de la langue dans une composition musicale :

a) Imitation de la prosodie d'une langue naturelle par un instrument

Les musiciens imitent l'intonation avec leur instrument. Nous en avons déjà donné un exemple avec la guitare parlante de Jimi Hendrix (**extrait sonore n°1**). Cet effet est utilisé par Robert Wyatt lorsqu'il se met à dialoguer avec son saxophone dans « *Alife* » (**extrait sonore n°18**). On le retrouve également dans les improvisations de jazz (Duke Ellington) et de blues (Luther Allison). De la même façon, dans le spectacle « *Con moto* » de La Framboise Frivole, le musicien belge Peter Hens dialogue avec son violoncelle. En musique classique instrumentale, on trouve un passage qui a été traité comme du chant parlé : le violoncelle et la contrebasse entament un récitatif comme le ferait une voix humaine au début de l'Ode à la Joie de Schiller, dans la Neuvième Symphonie de Beethoven dont nous avons déjà parlé plus haut (**extrait sonore n°13**).

Enfin, nous ne pouvions clore ce paragraphe sans parler du phonéticien Kenneth Pike et de sa flûte. La coulisse de cet instrument lui permettait en effet d'imiter les modulations de l'intonation.

b) Incorporation de clichés mélodiques

Bobby McFerrin incorpore le cliché mélodique « *na-na-na-na-nè-re* » dans son improvisation vocale (**extrait sonore n°4**). Les clichés sont souvent utilisés dans les improvisations, en raison de leur musicalité, mais surtout parce qu'ils sont immédiatement reconnus par les membres d'une même communauté linguistique. Ce faisant, le musicien adresse une sorte de clin d'œil musical à l'auditoire.

c) Utilisation de la musicalité naturelle de la langue sans modifications

La parole possède une musicalité intrinsèque, une mélodie et un rythme spécifique que les musiciens peuvent réutiliser telle quelle, sans la modifier. Au début de « *Kids can dance* », le groupe Hexstatic utilise l'intonation d'un énoncé en anglais comme ligne mélodique : « *I know you like music, Here's again, It is yours, I canno't accepted, I want you to have it, or before you go, please, play that music for me. Yes master* » (**extrait sonore n°19**). Le rythme d'une batterie vient se superposer au rythme et à la mélodie « naturelle » de cette phrase, sans les modifier.

d) Composition fondée sur la musicalité naturelle de la langue

La composition peut s'inspirer de la musicalité naturelle de la langue. Ainsi, nous avons vu que le chant grégorien était fondé sur la musicalité du latin (**extrait sonore n°5**). De même, le récitatif donne la priorité au texte, donc à la musicalité de la langue de composition (**extraits sonores n°7, 9 et 11**). Dans *Pelléas et Mélisande*, Debussy a élaboré son écriture à partir de la musicalité de la langue française (**extrait sonore n°14**).

Incorporation de la prosodie à l'écriture musicale et élaboration d'un matériel original**e) Mode de déclamation à la frontière du parlé et du chanté**

Le Sprechgesang est une technique de déclamation qui exploite l'ambiguïté entre la langue et la musique, qui mélange le parlé et le chanté d'une manière telle qu'il devient difficile de déterminer ce qui appartient à la langue et ce qui relève de la musique (**extrait sonore n°15**).

f) Polyphonie d'un petit groupe de voix, structurées autour du retour de certains événements phonétiques (Crystal, 1987 : 173)

Dans le morceau « *It can't happen here* », Franck Zappa réalise avec les autres membres de son groupe une polyphonie structurée sur le retour de faits phonétiques et prosodiques (**extrait sonore n°20**). La phrase introductive de l'extrait est chantée par Zappa lui-même : « *But they will freak out in Wahinsgton D.C.* ». Une polyphonie rythmique lui succède qui se structure de la manière suivante : l'accompagnement est composé de phrases scandées par les différents membres du groupe. Ces phrases sont répétées et alternées sans interruption de manière à former une rythmique stable, dont on peut extraire la pulsation : : « *DC, DC, DC* » + « *they can't happen here* » + « *ba ba ba* »... Un peu plus loin vient se greffer la « mélodie principale », déclamée par Zappa : « *everybody say that can't happen here ; no freak for us, can't happen here ; everybody's clean and can't happen here, no, no, it won't happen here, I tell you ...* ». Dans un autre passage (0 mn 48), l'énoncé suivant est scandé par un chanteur : « *And they thought it couldn't happen here ; and knew it couldn't happen here. They were so sure it couldn't happen here, but...* ». Les autres voix l'accompagnent en interprétant des parodies de l'intonation de cette phrase sur la suite déphonématisée « *mamama...* ».

g) Instrumentalisation de la voix

Dans la Sequenza III (**Extrait sonore n°16**), Luciano Berio instrumentalise la voix en découpant les mots en syllabes et en phonèmes, qu'il recompose ensuite à sa guise pour créer une musique originale. C'est une sorte de collage au niveau de la composition, sans avoir recours à une manipulation du signal sonore.

Traitement du signal sonore (montage- collage – analyse)**h) Reconstitution d'une mélodie originale à partir du collage de plusieurs fragments langagiers**

Dans « *Ninja tune* », le groupe Hexstatic¹⁹ crée une mélodie à partir du collage de plusieurs mots : « *Media* », « *one* », « *solo* » et « *deadly* », prélevés au sein de différents énoncés et prononcés par différentes voix (**extrait sonore n°21**). L'assemblage de ces mots au sein d'un nouvel énoncé « *Media, one, one, one solo, media, one, one, one, deadly* » crée une nouvelle mélodie. La rythmique est générée par l'agencement spécifique de ces mots (première phrase de l'extrait) auquel vient s'ajouter à la phrase suivante un rythme de batterie.

i) Effet cocktail : polyphonie de grands groupes (Crystal, 1987 : 173)

Dans « *Ninja tune* », le groupe Hexstatic réalise une polyphonie en exploitant l'effet cocktail. Différentes voix d'hommes et de femmes, s'exprimant dans différentes langues (japonais, anglais, espagnol, allemand...), ont été prélevées dans des journaux radiophoniques. Ces voix ont ensuite été réagencées de manière à créer une polyphonie (**extrait sonore n°22**). L'extrait commence par une voix d'homme qui s'exprime en japonais. Une autre voix d'homme en anglais vient s'y superposer, suivie quelques instants plus tard d'une voix de femme en espagnol. A mesure que le nombre de voix se rajoutent, les paroles deviennent inintelligibles et l'oreille ne perçoit plus qu'un brouhaha, une bouillie sonore, de laquelle émerge progressivement à la fin de l'extrait une rythmique introduite par une basse, à laquelle vient s'ajouter un « montage » rythmique réalisé à partir du collage de différents fragments prélevés dans les phrases (voir suite, h, **extrait sonore n°21**). Grâce au collage, on peut imaginer un autre type de polyphonie, l'autopolyphonie où le montage est réalisé à partir de différents enregistrements d'une seule et même voix (Crystal, 1987 : 173).

¹⁹ Site web du groupe Hexstatic : <http://hexstatic.co.uk/>

j) Écriture musicale réalisée à partir d'une analyse acoustique du donné langagier

Des compositeurs comme Luca Francesconi (**extrait sonore n°17**) et Claudy Malherbe fondent leur écriture musicale sur une analyse acoustique de la parole. Les accords musicaux sont par exemple déduits de la répartition spectrale des voyelles chez Francesconi.

3. Similitudes de formes étendues à des similitudes structurales

3.1. Universaux

Langue et musique ont de fortes similitudes de formes, à tel point que dans certaines situations, nous avons vu qu'il était difficile de les distinguer l'une de l'autre. Il n'est donc pas étonnant que ces similitudes soient exploitées par les spécialistes de la matière sonore que sont les phonéticiens et les musiciens. Nous ne serions toutefois pas complets si nous n'abordions pas le travail des chercheurs d'autres disciplines. Il est en effet très séduisant d'étendre ces similitudes de forme à des similitudes de structure et de chercher à savoir si langue et musique ne partagent pas des « structures profondes » similaires, voire identiques. Ainsi, beaucoup de théoriciens en linguistique ou en musicologie ont tenté de définir des universaux communs à la langue et à la musique. Ils est vrai qu'elles ont toutes les deux un statut particulier dans l'histoire de l'humanité : elles se sont développées dans toutes les sociétés humaines et cela, quelle que soit l'avancée technologique atteinte par chacune de ces sociétés. Tous les êtres humains sont capables de parler et de faire de la musique, sans apprentissage préalable. L'acquisition de la langue, comme celui de la musique se fait de manière « naturelle », par simple confrontation avec le milieu (cf. chapitre II). Et la variation entre les différents milieux est à l'origine aussi bien de la diversité linguistique, que de la diversité musicale. La langue et la musique constituent donc des universaux au sein des pratiques humaines. Une connaissance approfondie de ce qui rapproche langue et musique permet de mieux cerner leurs différences, nous l'avons vu dans la partie 1.2. pour le rythme et la mélodie. Et peut-être de relever le défi lancé par Mc Adams, « *comprendre en quoi la musique est complètement différente de tout autre chose* », défi qu'on peut lancer pour la langue également. Quelles qu'en soient les conclusions, une telle approche comparative ne peut que se révéler riche d'enseignement.

Cette approche se montrera d'autant plus riche que le chercheur aura pris soin d'éviter certains pièges, dont le plus grave serait de caricaturer l'un des systèmes à travers le filtre de l'autre et de lui appliquer des concepts qui ne lui conviennent pas.

3.2. Similitudes structurales

Les messages verbaux et musicaux ont en commun d'être construits à partir de la **combinaison d'un certain nombre d'unités sans signification**, les phonèmes pour la langue (26 pour le français) et les notes pour la musique (12 pour la gamme chromatique). Dans les deux systèmes, ce **répertoire d'unité** est **limité**. Mais ces unités de bas niveau, sans signification propre peuvent se combiner entre elles pour produire des unités plus complexes à un niveau plus élevé, qui elles-mêmes vont se combiner, pour atteindre un niveau encore plus élevé. Ainsi, selon Lerdahl et Jackendoff (1983), un énoncé se divise en phrases, ces phrases se divisent en propositions, les propositions en mots, les mots en syllabes et les syllabes en phonèmes ; une pièce musicale se divise en sections, ces sections se divisent en phrases musicales, les phrases en figures mélodiques et les figures rythmiques en notes. Ces unités de différents niveaux répondent donc à une **organisation hiérarchique**. Cette structuration donne la possibilité à la langue et à la musique de créer une infinité d'énoncés porteurs de signification, alors qu'elles partent toutes les deux d'un répertoire très restreint d'unités de bas niveaux. Elles sont donc de **nature générative**. Cette structuration hiérarchique de la langue et de la musique n'est pas seulement une abstraction, mais elle prend une forme très concrète grâce au marquage acoustique de certaines unités (voir chapitre II, partie 2.3.1.). Ainsi, la prosodie contient de nombreux indices sur les frontières linguistiques importantes. Par exemple, la frontière entre les propositions est marquée par une augmentation ou une diminution de la Fo ou par un allongement de la syllabe précédant la frontière de la proposition (Klatt et Cooper, 1975). De la même manière, les frontières musicales sont indiquées par une chute de la hauteur et un allongement de la durée des notes (Juszyk et Krumhansl, 1993). Ce marquage acoustique des constituants va se révéler d'une importance fondamentale dans l'acquisition de la langue et du système musical maternel par l'enfant au cours de la première année de sa vie (voir chapitre II).

3.3. Peut-on parler de grammaire et de langage à propos de musique ?

La linguistique générative définit la grammaire comme un ensemble fini de règles qui rendent compte d'un nombre illimité d'énoncés différents. S'il est possible d'établir la grammaire d'une langue grâce à la description de ces éléments constitutifs et des règles qui permettent de les combiner, il devrait donc être possible d'établir la grammaire d'une musique spécifique, d'autant que nous venons de voir que la musique, comme la langue obéit à une organisation hiérarchique d'unités. Pour un auteur comme Chomsky, la grammaticalité de la musique ne fait aucun doute. À vrai dire, cette idée n'est pas neuve et il existe déjà des grammaires de la musique modale, de la musique tonale ou de la musique sérielle par exemple. Les musiciens doivent apprendre et maîtriser ces systèmes de règles s'ils veulent pouvoir composer dans ces « langages musicaux ». Mais selon l'approche générative, la grammaire ne consiste pas seulement en une description des unités et des règles couchée sur le papier, elle est surtout et avant tout un système intériorisé par l'auditeur, qui lui permet d'interpréter les messages qu'il reçoit et d'en générer par lui-même. Cette grammaire intériorisée lui donne la capacité de juger si une séquence est correcte ou non, autrement dit, si elle est grammaticale ou non. Cette « compétence » a été décrite pour la langue par la grammaire générative et transformationnelle de Chomsky. Mais, existe-t-il une compétence similaire pour la musique ? Pour le savoir, Sloboda (1976) a demandé à des pianistes de déchiffrer des pièces musicales inconnues. Ces pièces respectent scrupuleusement les règles du langage musical dans lequel elles ont été composées (musique tonale), mais comportent des erreurs de notation. Les règles sont tellement intériorisées par les pianistes, qu'ils se montrent totalement incapables de détecter ces erreurs, car ils les corrigent inconsciemment au moment de la lecture. La grammaire de la musique tonale semble donc bien être intériorisée chez l'auditeur, qui développe ainsi, comme pour sa langue maternelle, une « compétence » à juger si les énoncés musicaux sont corrects ou non, s'ils sont grammaticaux ou agrammaticaux.

Si la musique possède une grammaire, alors, peut-on parler de langage à propos de la musique ? En fait, tout dépend de la définition du langage qu'on adopte. Pour Cooke (1959), ce terme désigne simplement

« un système de communication d'un état émotionnel entre le compositeur et les auditeurs, lequel système incorpore une sorte de lexique informel d'unités musicales expressives » (Clarke, 1989 : 24).

Si l'auditeur maîtrise le langage de la musique, il pourra comprendre le message que le compositeur lui adresse par le biais du canal que représente son œuvre ; s'il échoue, il sera considéré comme peu musicien. Cette transmission de l'intention du compositeur à l'auditeur est rendue possible selon Cooke grâce aux correspondances très stables qui existent entre la configuration mélodique et le contenu affectif dans la musique occidentale tonale. Osmond-Smith (1989 : 145) critique vertement ce type de théorie où l'auditeur est censé comprendre le message que le compositeur veut lui faire passer :

« cette sorte de modèle communicatif s'applique assez bien au morse, mais je ne suis pas vraiment convaincu qu'il décrive de façon adéquate le processus impliqué dans un médium artistique ».

Selon Ruwet (1972 : 44), il ne faut tenir compte de ce que nous a appris la linguistique et ne pas s'en tenir à la conception « langage-signal ».

Pour Pierre Schaeffer (1966), si la musique est un langage, ce n'est pas au même titre que le Langage proprement dit car la structure musicale se trouve liée indissolublement aux qualités sensibles de son matériau :

« on peut mettre en doute un parallélisme étroit entre langue et musique, en raison de l'arbitraire qui reste attaché au choix du sens, de la relation libre du signifiant et du signifié, qui fait du mot un signe, alors que la note de musique a toujours paru s'imposer en dehors de tout arbitraire comme une donnée du monde physique, à la quelle nous serions sensibles » (Schaeffer, 1966 : 35).

D'après lui, dans toute l'histoire de la musique, c'est la « musique pure » qui se rapproche le plus de la langue, langue dans le sens que lui donne Saussure, c'est-à-dire dégagée de la parole. Il prend pour exemple de musique pure l'« *Art de la Fugue* », composée en 1749 par Jean-Sébastien Bach. Cette œuvre a été écrite sans prévoir aucune instrumentation et Chion souligne que

« la musicalité est entièrement dégagée de la sonorité et l'objet sonore y est entièrement utilisé comme signe ; il n'y a que des relations et la musique n'est plus faite que d'un jeu de valeurs, de traits distinctifs, de hauteurs et de durée » (Chion, 1983 : 83).

Seule la hauteur peut constituer selon Schaeffer la valeur de base d'une musique pure, grâce à sa haute capacité d'abstraction et notamment sa capacité d'être abstraite des objets qui la supportent :

« la musique traditionnelle tend à s'abstraire peu à peu de la sonorité pour se constituer en langue, c'est-à-dire en système de valeurs pures [...] et de la sorte, à ne retenir des objets sonores que tel ou tel trait distinctif, compromis entre une acoulogie naturelle et sociale. Tel est le point où musique et langage se rapprochent le plus et en donnent des preuves évidentes » (Schaeffer, 1966 : 309).

Seulement, cette proximité ne peut être atteinte que si le langage musical respecte le cadre naturel des rapports simples de quinte et d'octave, donc qu'il se fonde sur des critères de perception musicale. La musique pure n'échappe pas à la loi qui veut que toute musique soit faite *« pour être entendue »* et c'est justement cette caractéristique qui constitue la différence entre langue et musique : alors que la signification du langage n'a qu'un lien arbitraire avec son support acoustique, les valeurs musicales, mêmes abstraites, sont reliées aux propriétés de l'objet. La musique est en effet bâtie sur un système où le signe n'est pas arbitraire et où il fonde ses valeurs sur les propriétés effectives de l'objet perçu, c'est-à-dire les rapports simples de hauteurs.

Seulement, lorsque Schaeffer parle d'arbitrarité du signe, il se réfère à la relation entre le signifiant et le signifié exposée par Saussure et cette définition ne tient justement pas compte de la prosodie, elle s'applique seulement au mot. Elle a d'ailleurs contribué à marginaliser la prosodie au sein des descriptions linguistiques en la faisant considérer comme accessoire, voire paralinguistique. Or, du point de vue de la forme, c'est bien la prosodie qui entretient le rapport de proximité maximale avec la musique. Et sa réalisation acoustique n'entretient pas de rapport arbitraire avec son signifié. Il faudrait donc prendre la précaution de comparer ce qui est comparable et il est étonnant qu'un musicien ne se soit pas montré plus sensible à la musique de la langue. Par ailleurs, les notions de signifiant et de signifié reposent directement sur la distinction que Saussure établit entre langue et parole. Mais une distinction de ce type n'a jamais été établie pour la musique :

« dans la science de la musique, on n'a pas établi de distinction terminologique correspondante [à la distinction langue / parole pour le langage]. Dans cet article, nous utiliserons le terme d'exécution de la musique pour désigner la réalisation acoustique de la musique » (Carlson et al., 1989 : 551).

Il reste donc très difficile de s'affranchir de son domaine de spécialité : un linguiste aura tendance à imposer des modèles créés au départ pour la linguistique, sans respecter les spécificités de la musique.

Ainsi, celle-ci est souvent caricaturée à l'extrême et il est hasardeux de parler de la musique en général, alors qu'en réalité, on ne se réfère qu'à une culture musicale circonscrite géographiquement et historiquement : il devient facile de généraliser et de conclure que les limitations propres à cette musique sont celles de la musique en général.

3.4. Capacité cognitive commune

Dans ce contexte, une approche comme celle de Lerdahl et Jackendoff (1983) se révèle intéressante car ces chercheurs essaient justement de respecter les propriétés de la langue et celles de la musique sans caricaturer l'une en fonction de l'autre. Ils reconnaissent les similitudes structurelles entre la langue et la musique, notamment entre la structure prosodique de la langue et la structure des trames temporelles en musique, mais ne veulent pas les assimiler l'une à l'autre. Ils font l'hypothèse que la langue et la musique sont des

« manifestations différentes d'une capacité cognitive plus générale et abstraite, à savoir la segmentation et l'organisation temporelles, qui caractérisent aussi d'autres activités, comme la danse » (Clarke, 1989 : 31).

Dans ce sens, langue et musique seraient *« deux caractéristiques analogues, ayant une certaine forme d'origine cognitive commune »*, qui se seraient développées *« dans leur domaine respectif, selon les principes spécifiques du domaine »* (Clarke, 1989 : 31). Cette hypothèse rejoint d'une certaine manière celle de Condillac, Rousseau et Assafjew sur la spécialisation de chacun des deux systèmes, spécialisation qui serait responsable de leur différenciation au cours de l'histoire de l'humanité (cf. partie 1.4.). Cette perspective d'une origine cognitive commune de la langue et de la musique prend une dimension tout-à-fait captivante lorsque l'on se place non plus du point de vue de la phylogenèse, mais de l'ontogenèse. En effet, pendant la première année de la vie, langue et musique sont dans une certaine mesure indifférenciées (cf. Chapitre II). Elles pourraient donc dériver toutes deux de certaines propriétés fonctionnelles propres à notre système auditif nerveux (cf. Chapitre VII, 3. discussion).

Bibliographie chapitre I

- Awake ! (1997). Do african drums really talk ? Watchtower Bible and Tract Society of New-York, July 22, 21-23.
- Barraud, Henry (2002). La mélodie. Dans *Encyclopedia Universalis* en ligne.
- Beier, U. (1976). The talking drum of the Yoruba. In Umiker and Sebeok, 690-692.
- Benveniste, Emile (1951). *La Notion de « Rythme » dans son expression linguistique*. Paris : P.U.F., 402 p.
- Berio, Luciano (1998). *Sequenzas*. Livret d'accompagnement du CD, Deutsche Grammophon (457 038-2). Ensemble Intercontemporain.
- Bigand, Emmanuel (1994). Contributions de la musique aux recherches sur la cognition auditive humaine. Dans Mac Adams et Bigand, 249-298.
- Bolinger, Dwight L. (1998). Intonation in American English. In Hirst and Di Cristo, 45-55.
- Boulez, Pierre (1963). *Dire, Jouer, Chanter. La Musique et ses Problèmes Contemporains*. Paris : Juillard.
- Brelet, Gisèle (1949). *Le Temps Musical : Essai d'une Esthétique Nouvelle de la Musique*. Paris : P.U.F., 842 p.
- Burnet, James (1776). Réédité en 1974. *Of the Origine and Progress of Language*. New-York: Olms, vol. III.
- Busnel, René Guy and Classe, André (1976). *Whistled Language*. New-York: Springer, 117 p.
- Carlson, Rolf, Friberg, Anders, Frydén, Lars, Granström, Björn et Sundberg, Johan (1989). La parole et l'exécution de la musique : parallèles et contrastes. Dans Mac Adams et Deliège, 551-571.
- Chion, Michel (1983). *Guide des Objets Sonores : Pierre Schaeffer et la Recherche Musicale*. Paris : Buchet-Chasnel, 187 p.
- Clarke, F. E. (1989). Considérations sur le langage et la musique. Dans Mac Adams et Deliège, 23-43.
- Cohen, Antonie and Nootboom, Sieb G. (1975). *Structure and Process in Speech Perception*. Berlin: Springer, 351 p.
- Condillac (1746). *Essai sur l'Origine des Connaissances Humaines*. Paris : Armand Colin, ed. de 1924, 225 p.
- Cooke, Deryck (1959). *The Language of Music*. Londres: Oxford University Press.
- Cooper, G. and Meyer, L. B. (1960). *The Rhythmic Structure of Music*. University of Chicago Press.
- Cowan, Georges M. (1948). Mazateco whistled speech. In *Language*, 24/3, 280-286.
- Crystal, David (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press, 381 p.
- Crystal, David (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press, 472 p.
- Dauer, Rebecca M. (1983). Stress-timing and syllable-timing reanalysed. In *Journal of Phonetics*, 11, 51-62.
- Delattre, Pierre (1969). L'intonation par les oppositions. Dans *Le Français dans le Monde*, 64, 6-13.
- Delcloque, Philippe (1999). *Speech Technology Applications in CALL*. Dundee: University of Albertay Dundee, 87 p.
- Di-Cristo, Albert (1975). *Soixante et Dix Ans de Recherches en Prosodie : Bibliographie Alphabétique, Thématique et Chronologique*. Aix-en-Provence : Editions de l'Université de Provence, Travaux de l'Institut de Phonétique d'Aix-en-Provence, 351 p.
- Di-Cristo, Albert (1998). Intonation in French. In Hirst et Di Cristo, 195-218.
- Di Cristo, Albert (1999). Le cadre accentuel du français contemporain : essai de modélisation. Dans *Langues*, 3/2, 184-205.
- Didier, Béatrice (1982). Condillac et la Musique. Dans Sgard, 395-405.

- Dodane, Christelle, Konopczynski, Gabrielle, Santi, Serge et Espesser Robert (1999). Fo declination line: more evidence from infant's and children's speech (9-48 months). *Actes du Colloque ICPH's*, San Francisco, 2161-2164.
- Dom de Malherbe (1946). *Le Chant Grégorien*. Paris : Lyra Dei.
- Dom Gajard, Joseph (1951). *La Méthode de Solesmes : ses Principes Constitutifs, ses Règles Pratiques d'Interprétation*. Paris : Desclée et Cie, 94 p.
- Dowling, Jay W. and Harwood, Dane L. (1986). *Music Cognition*. Orlando: Academic Press, 258 p.
- Drake, Carolyn (2002). Temporal organisation of sound sequences: Same processes in music, environmental scenes and speech ? ISCA Workshop "Temporal Integration in the Perception of Speech", 8-10 avril, Aix en Provence.
- Drake, Carolyn, Penel, A. and Bigand, Emmanuel (2000a). Tapping in time with mechanically and expressively performed music. In *Music Perception*, 18, 1-24.
- Drake, Carolyn, Jones, Riess, Mari and Baruch, Clarisse (2000b). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. In *Cognition*, 77, 251-288.
- Encyclopædia Universalis (2002). *Base de données Universalis en ligne* (consultée en février 2002) : <http://www.universalis-edu.com>.
- Essien, Jimmy Akpan (2000). *Contribution à la Recherche sur la Perception des Tons du Yoruba : Evidences Expérimentales à partir des Tambours, des Signaux de la Parole et de la Synthèse*. Thèse de Doctorat, Université Paris III, non publiée, 426 p.
- Faure, Georges (1962). *Recherches sur les Caractères et le Rôle des Éléments Musicaux dans la Prononciation Anglaise (Essai de Description Phonologique)*. Thèse de Doctorat d'Etat. Paris : Didier, 380 p.
- Fonagy, Yvan and Magdics, Klara (1963). Emotional patterns in intonation and music. In *Zeitschrift für Phonetik, Sprach-Wissenschaft und Kommunikations-Forschung*, 16 (1-3), 293-326.
- Fonagy, Yvan, Guzman, Marina et Bérard, Eva (1976). Comment mesurer l'accent d'intonation ? Dans *Travaux de l'Institut d'Etudes Linguistiques et Phonétiques*, 2, 41-61.
- Fonagy, Yvan, Bérard, Eva et Fonagy, Judith (1983). Clichés mélodiques. Dans *Folia Linguistica*, XVII, 153-185.
- Fraisse, Paul (1974). *La Psychologie du Rythme*. Paris : PUF, 244 p.
- Fréron, Alain (2002). Matériau. *Encyclopedia Universalis* en ligne.
- Gagnard, Madeleine (1987). *La Voix dans la Musique Contemporaine et Extra-Européenne*. Van de Velde, 117 p.
- Gajewski, Boreslas (1902). *Grammaire du Solrèsol ou Langue Universelle de François Sudre*. Paris, 44 p. (disponible à l'adresse suivante : <http://www.ptialaska.net/~srice/solresol/intro.htm>).
- Grosjean, François (1983). How long is the sentence ? Prediction and prosody in the on-line processing of language ? In *Linguistics*, 21, 501-529.
- Grosjean, François et Hirt, Cendrine (1996). Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects. In *Language and Cognitive Processes*, 11 (1-2), 107-134.
- Guaïtella, Isabelle (2000). Les systèmes de notation de la prosodie et du geste : pourquoi et comment les utiliser ? Dans Guimbretière, 165-184.
- Guimbretière, Elizabeth (2000). *Apprendre, Enseigner, Acquérir : la Prosodie au Cœur du Débat*. Rouen : Publications de l'Université de Rouen, 312 p.
- Guimbretière, Elisabeth (2000). Entretien avec Ivan Fonagy : Mes rencontres avec l'intonation. Dans Guimbretière, 151-163.
- Handel, Stephen (1989). *Listening: an Introduction to the Perception of Auditory Events*. Cambridge: MIT Press, 597 p.
- Hardcastle, William J. and Laver, John (1997). *The Handbook of Phonetic Sciences*. Padstow: Blackwell Publishers, 904 p.
- Herzog, G. (1945). Drum signalling in a West African Tribe. In *Word*, 1, 217-238.

- Hirsh-Pasek, Kathryn, Kemler, Nelson, Deborah, Jusczyk, Peter W., Wright, Cassidy, K., Druss, B. and Kennedy, L. (1987). Clauses are perceptual units for young infants. In *Cognition*, 26, 269-286.
- Hirst, Daniel (1998). Intonation in British English. In Hirst et Di Cristo, 56-78.
- Hirst, Daniel and Di Cristo, Albert (1998). *Intonation Systems: a Survey of Twenty Languages*. Cambridge: University Press, 487 p.
- Hirst, Daniel and Di Cristo, Albert (1998). A survey of intonation systems. Dans Hirst et Di Cristo, 1-44.
- Jackendoff, Ray S. (1991). Musical parsing and musical affect. In *Music Perception*, 9, 199-230.
- Jakobson, Roman et Waught, Linda, R. (1980). *La Charpente Phonique du Langage*. Paris : les Editions de Minuit, 336 p.
- Jespersen, Otto (1922). *Nature, Evolution et Origine du Langage* (Language: It's nature, development and origin). Paris : Payot, 436 p. Première édition, George Allen and Unwin, Ltd, London.
- Jones, Daniel (1909). *Intonation Curves*.
- Jones, Riess Marie et Yee, William (1994). L'attention aux événements auditifs : le rôle de l'organisation temporelle. Dans Mc Adams et Deliège, 75-121.
- Jones, Riess Marie and Boltz, M. (1989). Dynamic attending and responses to time. In *Psychological Review*, 96, 459-491.
- Jusczyk, Peter W. (1989). Perception of cues to clausal units in native and non-native languages. Paper presented at the *Biennial Meeting of the Society for Research in Child Development*, Kansas City, Mo., April.
- Kaiser, L. (1957). *Manual of Phonetics*. Amsterdam: North-Holland publishing company, 460 p.
- Klatt, Dennis K. and Cooper, William E. (1975). Perception of segment duration in sentence contexts. In Cohen and Nooteboom, 69-86.
- Konopczynski, Gabrielle (1978). Un essai original de transcription musicale de la prosodie. Wilhelm Wundt : Völkerpsychologie (1900-1912). Dans *Travaux de l'Institut de Phonétique de Strasbourg*, 10, 82-97.
- Krumhansl, Carol L. (1983). Perceptual structures for tonal music. In *Music Perception*, 1, 28-62.
- Krumhansl, Carol L. Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions in music. In *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.
- Krumhansl, Carol L. and Jusczyk, Peter W. (1990). Infant's perception of phrase structure in music. In *Psychological Science*, 1, 70-73.
- Lacheret-Dujour, Anne et Beaugendre, Frédéric (1999). *La Prosodie du Français*. Paris : C.N.R.S. Langage, 353 p.
- Leipp, Emile (1971). *Acoustique et Musique*. Paris : Masson. 3^{ème} éd. (1980), 352 p.
- Léon, Pierre et Léon, Monique (1964). *Introduction à la Phonétique Corrective : à l'Usage des Professeurs de Français à l'Etranger*. Paris : Hachette, 98 p.
- Léon, Pierre et Martin, Philippe (1969). *Prolégomènes à l'Etude des Structures Intonatives*. Dans *Studia Phonetica*, n°2. Paris : Didier, 225 p.
- Léon, Pierre (1992). *Phonétisme et Prononciations du Français*. Paris : Nathan Université, 192 p.
- Lerdahl, Fred (1985). Théorie générative de la musique et composition musicale. Quoi ? Quand ? Comment ? Dans *La Recherche Musicale*. Paris : Bourgeois/IRCAM, 101-120.
- Lerdahl, Fred and Jackendoff, Ray S. (1985). *A Generative Theory of Tonal Music*. Cambridge: M.I.T. Press, 368 p.
- Lieberman, Philip (1963). Some effects of semantic and grammatical context on the production and perception of speech. In *Language and Speech*, 6, 172-187.
- Lindblom, Björn and Sundberg, J. (1970). Towards a generative theory of melody. In *Svensk Tidskrift för Musikforskning, Swedish Journal of Musicology*, 52, 171-881.
- Lipsitt, Lewis P. and Rovee-Collier, Carolyn K. (1981). *Advances in Infancy Research*. Vol I : XXII. Norwood: Ablex, 267 p.

- Mac Adams, Stéphane et Deliège, Irène (1989). *La Musique et les Sciences Cognitives*. Liège : Pierre Mardaga, 649 p.
- Mac Adams, Stéphane et Bigand, Emmanuel (1994). *Penser les Sons : Psychologie Cognitive de l'Audition*. Paris : PUF, 402 p.
- Marler, Peter R. (1969). Animals and Man: Communication and its Development. In Roslansky, 25-62.
- Marler, Peter R. (1977). The evolution of communication. In Sebeok, 45-70.
- Massin, Jean et Massin, Brigitte (1983). *Histoire de la Musique Occidentale*. Messidor-Temps Actuels. Ed. utilisée, 1998. Fayard, 1312 p.
- Meigret, Louis (1550). *Le Tretté de la Gramèze Françoeze*. Nach der einzigen Pariser Ausgabe. Herausgegeben von Wendelin Foerster, Heilbronn, 1988.
- Menezes, Flo (1993). *Luciano Berio et la Phonologie : une Approche Jakobsonienne de son Oeuvre*. Berlin : Peter Lang, Publications Universitaires Européennes, 278 p.
- Mettas, Odette (1971). *Les Techniques de la Phonétique Instrumentale et l'Intonation*. Paris : P.U.F., 166 p.
- Michon, John A. (1967). *Timing in Temporal Tracking*. Soesterberg: Institute for Perception RVO-TNO, 127 p.
- Morgan, James L., Meier, R. P. and Newport, Elissa L. (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking to the acquisition of language. In *Cognitive Psychology*, 19, 498-550.
- Nazzi, Thierry, Kemler Nelson, Deborah, Jusczyk, Peter and Jusczyk, A.M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. In *Infancy*, 1/1, 123-147.
- Osmond-Smith, David (1989). Entre la musique et le langage : vue depuis le pont. La parole et l'exécution de la musique : parallèles et contrastes. Dans Mac Adams et Deliège, 137-148.
- Palmer, C. and Krumhansl, Carol L. (1990). Mental representation for musical meter. In *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.
- Papousek, Mechthild and Papousek, Hanus (1981). Musical elements in the infants vocalizations: Their significance for communication, cognition and creativity. In Lipsitt, 163-224.
- Passy, Paul (1890). *Etudes sur les Changements Phonétiques et leurs Caractères Généraux*. Paris : Firmin-Didot, 270 p.
- Patel, A. and al. (1998) Processing prosodic and musical patterns: A neuropsychological investigation. In *Brain Language*, 61, 123-144.
- Perrot, Jean (1997). *Polyphonie pour Yvan Fonagy*. Paris : L'Harmattan, 530 p.
- Pierrehumbert, Janet (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT, Indiana University Linguistic Club.
- Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press, 203 p.
- Povel, D. J. (1981). Internal representations of simple temporal patterns. In *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3-18.
- Rivière-Raverlat, Jacqueline (1997). *Développer les Capacités d'Ecoute à l'Ecole*. Paris : PUF, 207 p.
- Roslansky, J. D. (1969). *Communication. A Discussion at the Nobel Conference*. Amsterdam: North Holland Publishing Company.
- Rossi, Mario (1999). *L'Intonation, le Système du Français : Description et Modélisation*. Paris : Ophrys, 237 p.
- Roudet, Léonce (1899). Méthode expérimentale pour l'étude de l'accent. Dans *La Parole*, 5, 321-344.
- Roudet, Léonce (1910). *Eléments de Phonétique Générale*. Paris : Welter, 363 p.
- Rousseau, Jean-Jacques (1763). *Essai sur l'Origine des Langues où il est parlé de la Mélodie et de l'Imitation Musicale*. Ed. de 1990. Paris : Gallimard, 284 p.
- Rousselot, l'abbé (1901-1908). *Principes de Phonétique Expérimentale*. Paris : Welter, 2ème vol., 1252 p.
- Ruwet, Nicolas (1972). *Langage, Musique, Poésie*. Paris : Seuil, 247 p.
- Sauvanet, Pierre (2002). Le rythme. Dans *Encyclopedia Universalis* en ligne.

- Schaeffer, Pierre (1966). *Traité des Objets Musicaux*. Paris : Seuil, 712 p.
- Schaffer, L. H. (1982). Rhythm and timing in skill. In *Psychological Review*, 83, 109-122.
- Schuyesmans (2002). *Le Chant Grégorien*. Site Internet disponible à l'adresse suivante : <http://www.schuyesmans.be/gregoriaans/>
- Scripture, Edward, W. (1902). *The Elements of Experimental Phonetics*. Londre: Edward Arnold. Présente édition (1973), New-York, AMS Press Inc, 627 p.
- Sebeok, Thomas (1977). *How Animals Communicate*. London: Indiana University Press.
- Sgard, Jean (1982). *Condillac et les Problèmes du Langage*. Paris : Slatkine, 462 p.
- Shadle, Christine H. (1997). The aerodynamics of speech. In Hardcastle and Laver, 33-64.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992). ToBI: A standard for labelling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, 867-870.
- Sloboda, J. A. (1976). The effect of item position on the likelihood of identification by influence in prose reading and music reading. In *Canadian Journal of Psychology*, 30, 228-236.
- Soulez, Antonia (1997). L'essence de la phônè (à propos du Cratyle 426C-427E). Dans Perrot, 411-423.
- Steele, Joshua (1775). *Prosodia Rationalis or an Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols*. London: Nichols, J.
- Todd, N. (1985). A model of expressive timing in tonal music. In *Music Perception*, 3, 33-57.
- Umiker, Jean and Sebeok, Thomas A. (1976). *Speech Surrogates*. Mouton. The Hague.
- Van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences*. PhD. The Netherlands, Eindhoven University of Technology.
- Van Waesberghe, J. Smits (1957). Phonetics in its relation to musicology. In Kaiser, 372-384.
- Viret, Jacques (1987). *La Modalité Grégorienne : un Langage pour quel Message ?* Lyon : A Cœur Joie, 167 p.
- Vos, P. and Troost, J. (1989) Ascending and descending melodic intervals: statistical findings and their perceptual relevance. In *Music Perception*, 6, 383-396.
- Wing, A. M. and Kristofferson, A. B. (1973a). The timing of inter-response intervals. In *Perception and Psychophysics*, 13, 455-460.
- Wing, A. M. et Kristofferson, A. B. (1973b). Response delays and the timing of discrete motor responses. In *Perception and Psychophysics*, 14, 5-12.
- Wundt, Wilhelm (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Leipzig : Engelmann, W., 3ème édition, 1912.
- Zwanenburg, Wiecher (1965). *Recherches sur la Prosodie de la Phrase Française*. Leiden : Universitaires Pers., 135 p.

Chapitre II

Indifférenciation de la langue et de la musique au début de la vie

« The vocal production of musical elements like most other prerequisites of human language do occur isolated in one or another group of animals, but in none do they coincide to permit the explosive development of human language » (Marler, 1969, in Papousek & Papousek, 1981 : 169).

« La parole et la musique sont deux exemples d'une communication humaine formalisée qui passe par des signaux acoustiques. Tous deux doivent être destinés aux mêmes systèmes perceptifs et cognitifs. Les limitations et les capacités de ces systèmes doivent contribuer fortement, à la fois, au développement de la parole et de la musique » (Carlson et al., 1989 : 569).

Table du second chapitre

1. Préalables à la description du développement linguistique et musical de l'enfant

- 1.1. Perception et production
 - 1.1.1. La perception précède la production
 - 1.1.2. La perception et la production entretiennent des rapports étroits au moment de l'acquisition
- 1.2. La perception est globale avant d'être analytique
 - 1.2.1. Les principes gestaltistes et l'analyse de scènes
 - 1.2.2. Les processus de groupement sont opérants au début de la vie
 - 1.2.3. Traitement global versus traitement local

2. Les éléments musicaux, point d'entrée dans la LM et dans le SMM

- 2.1. Une oreille mature et ouverte sur la musique de l'environnement bien avant la naissance
 - 2.1.1. Prédominance des éléments musicaux dans le monde sonore intra-utérin
 - 2.1.2. Les fœtus et leur LM
 - 2.1.2.1. Les voix de son environnement
 - 2.1.2.2. Les éléments musicaux de la langue
 - 2.1.3. Des fœtus mélomanes
 - 2.1.3.1. Des réactions différentes selon le type de musique
 - 2.1.3.2. Une mémoire musicale réactivable après la naissance
 - 2.1.4. La « culture prénatale » du fœtus
- 2.2. Sensibilité du bébé aux éléments musicaux : le « trésor inépuisable » des ressources enfantines
 - 2.2.1. Sensibilité au rythme
 - 2.2.2. Sensibilité aux contours mélodiques
 - 2.2.2.1. Les contours mélodiques de la musique
 - 2.2.2.2. Les contours d'intonation de la langue
 - 2.2.3. Sensibilité à la prosodie affectant les mots
 - 2.2.4. Capacités commune de segmentation pour la LM et le SMM
- 2.3. L'amorçage de l'attention par les éléments musicaux
 - 2.3.1. Richesse et redondance du signal
 - 2.3.2. Les signaux d'amorçage dans la langue
 - 2.3.3. Une théorie de l'amorçage valable pour la musique ?
- 2.4. Ajustements vocaux inconscients adaptés à la sensibilité musicale du bébé
 - 2.4.1. Préférence de l'enfant pour le langage et la musique adressés
 - 2.4.1.1. Le langage adressé
 - 2.4.1.2. Le chant adressé
 - 2.4.2. Caractéristiques du langage adressé à l'enfant
 - 2.4.2.1. Une pédagogie « inconsciente »...
 - 2.4.2.2. ...qui évolue avec l'âge de l'enfant

- 2.4.3. Caractéristiques du chant adressé à l'enfant
 - 2.4.3.1. Les berceuses
 - 2.4.3.2. Les chants de jeux
- 2.4.4. Points communs entre le LAE et le CAE

3. D'une base commune à une différenciation progressive

- 3.1. Perception
 - 3.1.1. Jusqu'à 6 mois, l'enfant est un « super-discriminateur »
 - 3.1.2. Spécialisation progressive sur la LM aux dépens de la richesse perceptive de l'enfant
 - 3.1.2.1. Focalisation sur la structure vocalique de la LM à partir de 6 mois
 - 3.1.2.2. Focalisation sur la structure consonantique de la LM à partir de 10/12 mois
 - 3.1.3. Spécialisation progressive sur le SMM
 - 3.1.3.1. Définition de l'acculturation musicale
 - 3.1.3.2. Bons patterns ou intégration des régularités de l'environnement
 - 3.1.3.3. Evolution de la perception des gammes non natives
- 3.2. Production
 - 3.2.1. Maîtrise progressive de l'appareil phonatoire
 - 3.2.2. Le babillage canonique, étape de différenciation entre la langue et la musique ?
 - 3.2.3. Le langage émergent
 - 3.2.4. Le chant émergent

Bibliographie du chapitre II

Pendant la première année de la vie, l'acquisition de la langue maternelle est avant tout perceptive. Cette imprégnation progressive avec les sons de l'environnement commence de manière très précoce, puisque l'oreille est mature bien avant la naissance. Et il se trouve que l'information qui réussit à franchir les obstacles de la paroi abdominale et du liquide amniotique de la mère est avant tout... musicale. Donc, non seulement, la perception est première, mais les éléments musicaux sont perçus en premier par le fœtus. Il n'est donc guère étonnant que, après sa naissance, le bébé développe une forte sensibilité à la musique et aux caractéristiques prosodiques de sa langue maternelle. Comme s'ils devinaient cette sensibilité, les adultes qui entourent l'enfant se mettent à lui parler d'une façon spéciale, en exagérant les paramètres prosodiques. Cette musicalité du langage adressé à l'enfant le rapproche d'une certaine manière des « langues musicales » dont nous avons parlées dans le chapitre précédent. Et lorsque l'on découvre que l'entourage du bébé adapte également sa façon de chanter pour faire ressortir la mélodie et le rythme, on en vient à se demander si des processus communs ne sous-tendent pas l'acquisition de la langue et du système musical maternels, voire si le traitement de la langue et celui de la musique ne sont pas indifférenciés au début de la vie. Comme le souligne Fassbender (1995 : 98-99),

« il est légitime de supposer que [la] sensibilité [du bébé] se fonde partiellement sur la capacité à grouper et à segmenter l'information auditive sur la base des principes de la Gestalt. Au départ, la perception de la langue et de la musique font appel à une base identique. Ce qui n'exclut pas qu'elles puissent suivre des voies développementales distinctes dès l'instant où une signification est attachée à une information acoustique particulière à la faveur de l'interaction sociale découlant de l'intuition parentale ».

Dans les premières vocalisations infantiles, il est par ailleurs très difficile de faire la part entre ce qui relève du musical et du linguistique. Pendant les quelques pages qui vont suivre, empruntons au bébé ses oreilles et sa voix, pour découvrir le monde essentiellement musical qui est le sien pendant la première année de sa vie.

1. Préalables à la description du développement linguistique et musical de l'enfant

1.1. Perception et production

1.1.1. La perception précède la production

« C'est que, en définitive, la langue n'est ni dans l'onde sonore, ni dans les organes phonateurs mais dans la discrimination auditive, à savoir dans les habitudes de discrimination propres à chaque communauté. De façon imagée et paradoxalement, on peut dire que les êtres humains parlent avant tout avec leurs oreilles » (Martin, 1996 : 103).

Dans l'acquisition de la langue maternelle (désormais LM) et du système musical maternel (désormais SMM), la perception précède la production pour une raison avant tout physiologique : à la naissance, le tractus vocal du nourrisson n'est pas assez développé pour pouvoir produire autre chose que des vagissements et des cris et les capacités de l'enfant à produire des sons sont contraintes par l'état de développement de son organisme (configuration du conduit vocal en diagonale plus proche de celle du primate que celle de l'adulte, en coude, liée à la position debout ; la configuration enfantine ne permet pas la maîtrise des articulateurs avant la seconde moitié de la première année). En revanche, les oreilles du nourrisson sont grandes ouvertes sur le monde puisque **l'oreille du fœtus est fonctionnellement mature dès le cinquième mois de gestation** (cf. 2.1.1.). En effet, chez l'être humain, le développement de la cochlée se fait in utero, contrairement à beaucoup d'autres espèces (Herbinet et Busnel, 1983). Elle s'enroule dès la 6^{ème} semaine de gestation et atteint son complet développement morphologique vers la 10^{ème} semaine (3 mm et enroulée sur 2 tours ½, comme une cochlée adulte). Elle entre en fonction vers 18 à 20 semaines, en dépit d'une immaturité cellulaire qui s'achèvera vers le huitième mois (Lecanuet, 1997 : 22). L'oreille interne est donc prête à fonctionner dès 4 mois ½, mais Shahidullah et Hepper (1992) ont observé une écoute réactive à des sons purs (250-500 Hz) encore plus précoce, dès la 16^{ème} semaine. Les fœtus de 30 à 36 semaines réagissent lorsqu'ils sont exposés à des sons intenses (< 90 dB de 2 à 5 secondes) (Hepper et Shaidullah, 1994). On observe les mêmes effets avec la musique (extraits de 5 minutes), bien que les fœtus âgés de 32-34 semaines réagissent moins que les fœtus de 35 à 37 semaines (Smith et al., 2000), révélant une différence de maturation auditive.

Certes, les seuils auditifs sont au départ élevés (100 dB), mais ils s'abaissent très rapidement pour permettre la discrimination des fréquences.

A sa naissance, le nourrisson est donc bien loin d'être une « ardoise vierge », totalement libre dans ses premiers jours de toute influence de son environnement linguistique et musical. La période de latence entre la perception et la production, rendue nécessaire par l'immatunité physiologique de son appareil vocal, joue un rôle primordial dans l'acquisition. En effet, pendant les premiers mois de vie, l'enfant s'imprègne de son environnement et notamment de sa LM et de son SMM ; il apprend progressivement à repérer les caractéristiques sonores saillantes au sein du flux sonore, et une fois qu'elles sont mémorisées, à les reconnaître. Selon Papousek et Papousek (1981 : 168), le décalage entre perception et production révèle la nécessité d'une mémoire à long terme au cours de l'acquisition.

1.1.2. La perception et la production entretiennent des relations étroites au moment de l'acquisition

Ce décalage ne signifie pas pour autant que perception et production soient deux processus totalement indépendants. Car si les activités de perception sont entravées par un système auditif déficient, la production de la LM et du SMM est gravement perturbée. Ainsi, on n'observe pas de babillage canonique chez les bébés sourds profonds (Vinter, 1992). Comme les autres sons de son environnement, les premières productions de l'enfant deviennent à leur tour un « input » pour son système perceptif, constituant une sorte de boucle où perception et production interagissent. On trouve de fortes empreintes de cette interaction dans le proto-langage¹ du bébé, qui possède déjà les caractéristiques prosodiques de base de la LM. Ces caractéristiques sont suffisamment saillantes pour que la LM d'un bébé de 8 mois soit identifiable par des auditeurs adultes appartenant à la même communauté linguistique (De Boysson Bardies, 1993). L'acculturation avec l'environnement va ainsi dans le sens du lien entre perception et production, puisque la production est influencée par la perception. Selon Jusczyk, perception et production sont très étroitement liées au moment de l'apparition des premiers mots.

¹ Proto-Langage ou pseudo langage : « *En situation d'interaction, l'enfant produit du « protolangage », discours construit comme une parole dépourvue de mots* » (Konopczynski et Vinter, 1989).

Alors que, dès 8-10 mois, le bébé est capable de reconnaître des mots en contexte (Benedict, 1979) et vers 11-12 mois, des mots hors contexte (Hallé et De Boysson-Bardies, 1994), il est remarquable qu'au niveau de la production, ses premiers mots n'émergent qu'aux alentours de 12 mois au sein du babillage mixte (cf. tableau récapitulatif « Les mots de la LM : reconnaissance, compréhension et production », Volume II, annexe 1.5.).

Cependant, dans les études portant sur l'acquisition du langage, perception et production sont souvent étudiées de façon isolée et la tendance reste majoritairement à la séparation, notamment à cause de la spécialisation des chercheurs dans l'un ou l'autre domaine. Pour Jusczyk (1997), ce « divorce » a été favorisé, parce que les informations collectées sur le développement de la perception et de la production l'ont été à des âges différents. Ainsi, beaucoup de travaux sur la perception se sont focalisés sur des enfants de 6 mois, voire plus jeunes, tandis qu'en production, les travaux se sont intéressés majoritairement aux enfants de plus d'un an. Cependant, les choses ont évolué depuis une vingtaine d'années et les productions précoces du bébé sont maintenant très bien décrites.

L'étude des rapports entre perception et production est d'une importance fondamentale pour comprendre comment les unités de représentation sont extraites du signal sonore. Cette conversion implique le passage d'un niveau acoustique, à un niveau abstrait, linguistique. La théorie qui pourrait décrire un tel processus de conversion serait capable non seulement d'éclairer la façon dont un enfant apprend sa LM, mais également d'unifier les deux niveaux. L'intérêt d'une théorie « unifiée » n'a bien sûr pas échappé aux chercheurs et sa nécessité a encore été soulignée lors des conférences plénières qui ont clôturé le dernier Congrès International des Sciences Phonétiques de San Francisco (1999). Il existe un certain nombre de théories sur l'acquisition qui tentent de faire interagir perception et production. La théorie motrice de la perception est l'une des plus anciennes et des plus connues (Liberman et al., 1967). Selon cette théorie, les sons sont perçus en fonction de leurs lieux d'articulation. Au moment du babillage, les bébés établiraient progressivement des relations entre les diverses articulations qu'ils produisent et les sons qui en résultent. Lindblom et al. (1984) proposent un modèle d'interaction entre les contraintes perceptives et articulatoires. Selon eux, la « discriminabilité » et la « prononcabilité » seraient des contraintes tout aussi importantes qu'écouter et parler. Selon leur modèle, les représentations perceptives des mots et leurs représentations articulatoires se combineraient en une généralisation plus abstraite et applicable aux deux systèmes.

Cette généralisation correspondrait au niveau phonologique. Une autre théorie, celle de Gathercole et Baddeley (1993) posent l'hypothèse que perception et production seraient coordonnées par l'intermédiaire d'une boucle phonologique, active chez l'enfant tout au long de sa première année. Cette boucle, en prise directe avec le flux sonore, permettrait de retenir l'information phonologique pendant une à deux secondes. Grâce à cette mémoire à court terme, l'enfant pourrait mettre constamment à jour son répertoire phonologique (capacité de base pour imiter).

1.2. La perception est globale avant d'être analytique

1.2.1. Les principes gestaltistes et l'analyse de scène

Au début du siècle dernier, les psychologues gestaltistes ont décrit un certain nombre de lois qui permettent au sujet, à partir d'un ensemble d'informations visuelles, de grouper les données provenant d'une même source (Wertheimer, 1923). Ces lois interviennent dans une étape préattentive (Neisser, 1967) et permettent la formation d'unités qui optimisent le traitement cognitif (Bregman, 1990). Elles permettent d'organiser au niveau perceptif le flot d'informations. Par exemple, si des éléments se déplacent en même temps et dans la même direction, ils seront groupés en une seule forme, obéissant ainsi à la loi de destin commun. Parmi ces formes, certaines sont plus prégnantes que d'autres ; on parle alors de « bonnes formes ». Une bonne forme satisfait à plusieurs de ces lois de groupement en raison de sa simplicité, de sa régularité et de sa symétrie. Le système perceptif « construit » donc une interprétation de son environnement en projetant des formes sur la réalité sensorielle. Initialement décrits pour la modalité visuelle, ces principes de groupement fonctionnent également pour l'audition (Deutsch, 1982 ; Bregman, 1990, 1994 ; Bigand, 1994). Lorsqu'un individu se trouve confronté pour la première fois à une mixture sonore, son système auditif doit extraire de cet ensemble des régularités acoustiques pour construire des formes. Par exemple, il regroupera les sons ayant une même évolution temporelle, avec un début et une fin simultanés. De la même manière, il groupera les éléments possédant le même timbre ou encore, proches du point de vue de leur fréquence. Pour désigner ce processus, Bregman (1990 : 38) parle d'« *analyse primitive de scènes auditives* ». L'analyse de scène est une méthode générale de partition d'une mixture en sources acoustiques distinctes à partir des régularités acoustiques de l'environnement sonore :

« L'analyse de scène désigne la stratégie par laquelle l'ordinateur regroupe l'ensemble des propriétés visibles – pourtours, textures de surface, couleurs, distances, etc. – d'un même objet. Ce processus semble déterminer la forme globale et les propriétés correctes d'un objet. Par analogie, l'analyse des scènes auditives est le processus réunissant en une unité perceptive l'ensemble des signaux provenant, dans une période de temps donnée, d'une seule source sonore de l'environnement » (Bregman, 1994 : 112).

Dans l'analyse de scène, l'attente et l'attention jouent un rôle primordial. En effet, la perception de l'auditeur sera différente s'il est prévenu qu'il va écouter un extrait de musique ou des sons de parole. Par exemple, il ne reconnaîtra pas l'objet « voyelle » à l'écoute des trois premiers formants synthétisés d'une voyelle, s'il n'a pas été prévenu au préalable que l'objet à percevoir était une voyelle. Dans un test de perception, les consignes sont donc d'une importance fondamentale, puisqu'elles vont « orienter » l'attention de l'auditeur et conditionner sa perception. Peu à peu, l'auditeur va mémoriser des formes dans son environnement et selon Bregman (1994), les activer soit de manière purement automatique (reconnaissance automatique), soit de manière volontaire en « cherchant à entendre » lorsqu'il sera confronté à une nouvelle mixture sonore (attention volontaire).

1.2.2. Les processus de groupement sont opérants au début de la vie

Si les processus de groupement décrits par les gestaltistes et l'analyse de scène sont opérants au début de la vie, ils devraient permettre à l'enfant, pendant la période de latence où il ne peut pas produire, de « rentrer » dans sa LM et dans son SMM, c'est-à-dire de détecter des formes à partir d'un tout vaste et indifférencié. Et il est effectivement apparu que certains principes de groupement sont déjà effectifs de façon très précoce chez le bébé et lui permettent un pré traitement du flux sonore, qu'il soit langagier ou musical.

Parmi ces principes, la **loi de proximité** permet au système perceptif de regrouper des éléments proches les uns des autres. Ainsi, pour la hauteur, si des notes successives se répartissent sur deux registres différents, l'auditeur adulte percevra deux lignes mélodiques distinctes évoluant parallèlement. Ce phénomène a été baptisé par Bregman « séparation des flux auditifs » (Bregman, 1978 ; Bregman, 1981). Les bébés sont capables de faire le même type de groupement, et ce de manière très précoce. Demany (1982) a testé 44 **bébés âgés de 1 à 3 mois et demi, qui ont regroupé les stimuli auditifs en fonction de leur proximité de fréquence**. Cette capacité a également été relevée chez des bébés de 2 à 5 mois (Fassbender, 1993).

Demany (1977) a montré que **la loi de proximité fonctionnait également au niveau temporel, et ce, à partir de 2 mois ½**. Les bébés sont en effet capables d'organiser des sons successifs qui ne diffèrent pas par la hauteur en fonction des intervalles de temps qui séparent ces sons, les intervalles de grande longueur étant interprétés comme des pauses. **La loi de proximité s'applique également pour le contenu spectral**. En revanche, **les différences d'intensité ne constituent pas un indice auditif suffisant** pour permettre un groupement chez des bébés de 7 à 11 mois (Thorpe et Trehub, 1989).

Un autre principe, la **loi de similitude** permet au système perceptif de regrouper des éléments semblables plutôt que dissemblables. Dans une série d'études, Thorpe et ses collaborateurs (Thorpe et Trehub, 1989 ; Thorpe et al., 1988) utilisent un pattern de 6 tons séparés par un même intervalle de temps. Les trois derniers tons sont différents des trois premiers du point de vue de la Fo, de la structure spectrale ou de l'amplitude (XXXOOO). Une pause est ensuite introduite entre les deux groupes ainsi formés (XXX OOO) ou à l'intérieur d'un des deux groupes (XXXO OO). **Les bébés de 6 à 9 mois réagissent à la dernière séquence, car elle viole le principe de regroupement par similitude**. Si le changement respecte la structure de la séquence standard, il n'est pas remarqué. Ce faisant, **les bébés se comportent exactement comme les auditeurs adultes** (Thorpe, 1985). Le phénomène d'illusion de durée résulte de la loi de similitude. En effet, les auditeurs repèrent une pause entre les groupes, même si celle-ci n'existe pas matériellement dans le signal sonore. Selon Trehub (1989 : 308) :

« des discontinuités au niveau de la hauteur, du timbre ou de la sonie indiquent des aspects importants de la structure d'un pattern et déclenchent une illusion de durée qui guide les processus primitifs de segmentation ou de ségrégation. Plus la discontinuité ou le changement d'un de ces paramètres est grand, moins la détection des pauses entre les groupes est précise. De ce fait, les auditeurs ne tiennent pas compte de l'information temporelle disponible et imposent une structure temporelle correspondant aux aspects de la configuration générale ».

Cette illusion de durée fonctionne également dans le traitement de la langue, puisque l'auditeur regroupe les événements en fonction de leurs similarités. Selon Studdert-Kennedy (1975), cette illusion ne pourrait pas fonctionner dans une langue étrangère.

Les principes de groupements sont donc opérants pour la langue et pour la musique et cela, de manière très précoce. Les bébés sont capables d'organiser l'espace acoustique en regroupant les éléments qui sont proches les uns des autres ou qui partagent des caractéristiques communes (fréquentielles, temporelles, de timbre).

Ces aptitudes sont d'une importance capitale pour la segmentation du flux sonore, car elles relèvent d'une capacité générale à grouper et à segmenter, qui fonctionne pour la langue et la musique.

1.2.3. Traitement global vs traitement local

L'analyse de scène et les lois de type gestaltiste impliquent un traitement de type global ou holistique des données sensorielles. Dans un traitement de ce type, c'est la forme, l'ensemble qui prime sur les éléments. En effet, si on change un seul élément, c'est l'ensemble qui se modifie. En revanche, la forme ne sera pas modifiée si on change l'ensemble des éléments, tout en maintenant constants les rapports entre ces éléments. Dans la perception visuelle, l'information globale précède l'information locale (Kimchi, 1992 pour une revue) et elle est identifiée plus rapidement que l'information locale (Navon, 1977 ; Proverbio et al., 1998). Le traitement global est en effet nécessaire avant un traitement de l'information plus analytique qui permette de décomposer une forme en ses éléments constitutifs. Dans la perception des contours mélodiques, il est mis en jeu, que ce soit dans la langue ou dans la musique (traitement des patterns globaux de changement de direction de hauteur, forme des contours mélodiques). On le retrouve également dans la perception du registre de hauteur et dans l'approximation des intervalles mélodiques et harmoniques. Il permet la discrimination de deux patterns et la reconnaissance de ces patterns, mais en aucun cas leur catégorisation. Il est donc limité, car les signaux sont traités comme des tous qui ne sont pas décomposables et donc pas recombinaux. C'est une capacité prélinguistique qui ne serait pas spécifique aux êtres humains, mais que l'on rencontrerait chez certains mammifères, les primates notamment.

En revanche, le traitement local qui est un traitement analytique, ne se rencontre que chez les êtres humains. Il permet d'isoler les éléments qui composent une forme, puis de comparer ces éléments pour ensuite les classer dans des catégories et leur attribuer une étiquette (processus de catégorisation). Selon Deutsch (1969), il intervient en musique dans la perception des phonèmes, des accords et des classes de hauteur des notes. Dans la langue, il intervient dans la perception des phonèmes. Le traitement analytique ne peut pas intervenir avant le traitement global. Cependant, le passage d'un type de traitement à l'autre est fortement conditionné par la difficulté relative de la tâche à réaliser par l'auditeur. Une tâche difficile conduit à un mode d'écoute synthétique (global) où l'auditeur manifeste peu de sensibilité pour les composantes individuelles du pattern.

En revanche, une tâche facile conduit à un traitement analytique, où l'auditeur se montre au contraire sensible aux composantes du pattern. Par ailleurs, il existe des contextes « facilitants » qui permettent d'orienter l'auditeur vers un traitement plus analytique. Si le contexte dans lequel apparaît un pattern donné respecte les principes utilisés dans l'environnement de l'auditeur (« bon contexte »), le pattern sera traité plus facilement, même s'il n'a jamais été rencontré auparavant.

Pour comprendre le passage d'un traitement global à un traitement analytique, prenons l'exemple de la reconnaissance d'un air familier. Chez les adultes, la reconnaissance de la mélodie d'un extrait musical est relative, elle dépend de la relation entre les différents sons successifs, c'est-à-dire les intervalles entre ces sons. Les intervalles sont très importants dans la reconnaissance des mélodies familières. Par exemple, c'est une tierce majeure qui marque le début de « *A la Claire Fontaine* ». Si les intervalles sont maintenus, mais que la hauteur de départ est changée, nous percevrons les mélodies comme équivalentes. Le traitement par intervalles caractérise la perception d'une mélodie familière (traitement de type analytique). En revanche, le traitement est de type global lorsque nous percevons pour la première fois une mélodie. Dans ce cas, c'est le contour mélodique qui prime et notamment, les changements de sa direction (Bartlett et Dowling, 1980 ; Dowling, 1978). Cependant, si ces mélodies non familières ont été composées selon les principes musicaux familiers à l'auditeur (par exemple, la musique tonale), elles seront codées et mémorisées plus facilement (Francès, 1958 ; Zenatti, 1969). Selon Bartlett et Dowling (1980), si une oreille connaît les principes musicaux qui gouvernent la musique tonale, le traitement sera plus analytique, car le contexte est facilitant (stratégie de traitement par intervalles et non par contours). Les auditeurs musiciens qui ont acquis une meilleure connaissance de ces principes grâce au solfège et à une écoute intensive de la musique pourraient manifester un traitement plus analytique (voir chapitre III, 4.2.).

Au niveau du développement, l'apparition des capacités de traitement analytique dépend de la maturation corticale du cerveau de l'enfant. D'après LeCours (1975), la myélogénèse du cortex auditif n'est pas complètement achevée avant l'âge de 5 ans et avant cette maturation, le jeune enfant utiliserait un traitement auditif subcortical, plus adapté au traitement global (Whitaker, 1976). Jusqu'à 5 ans, la perception de l'enfant serait donc dominée par un traitement de type global qui expliquerait ses difficultés à intégrer les détails en un ensemble structuré.

Cette domination du global sur le local ne veut pas pour autant dire que l'enfant n'a pas de capacités de traitement analytique. Sans elles, il ne pourrait d'ailleurs pas distinguer et mémoriser les unités de la langue et de la musique. Il semblerait que des changements importants interviennent au niveau cognitif dès la fin de la première année de sa vie, qui lui permettent de développer des capacités plus analytiques. Ces « nouvelles » capacités lui permettraient de repérer et de mémoriser les contrastes qui constituent le stock d'unités de sa LM et de son SMM. Cette réorganisation expliquerait la perte des capacités de discrimination constatées à l'âge de 10-12 mois et l'ajustement à certaines catégories de sons (cf. partie 3.1.2.). Comme c'est le cas pour l'oreille musicienne, cette réorganisation pourrait intervenir à un moment où l'enfant a emmagasiné suffisamment de « connaissances » sur sa LM et son SMM pour que le traitement devienne plus analytique dans les contextes facilitants.

« Thus, in the early months, the infant auditory system may be better equipped to process relatively simple auditory patterns such as the melodic pitch patterns of motherese than the more complex auditory patterns of normal adult speech [...] Such holistic pattern recognition constitutes a prelinguistic auditory processing mode, for which the exaggerated pitch contours of mothers' speech may be a particularly appropriate form of acoustic stimulation » (Fernald, 1984 : 14).

Ces fameuses capacités dont le bébé fait preuve pour discriminer les contrastes de nombreuses langues (cf. 3.1.1.) mettraient en jeu un traitement de type global et non pas analytique, le traitement global permettant de discriminer deux patterns, mais non de les extraire et de les mémoriser.

2. Les éléments musicaux, point d'entrée dans la LM et le SMM

2.1. Une oreille mature et ouverte sur la musique de l'environnement bien avant la naissance

2.1.1. Prédominance des éléments musicaux dans le monde sonore intra-utérin

Les sons qui parviennent à la cochlée du fœtus sont altérés et transformés par un environnement bruyant. Des enregistrements intra-abdominaux (hydrophone placé dans le vagin après la rupture des membranes avant ou après l'accouchement) permettent de se placer dans les mêmes conditions d'audition que le fœtus et révèlent un bruit de fond dans les fréquences basses (500-700 Hz, Lecanuet, 1997 : 24) constitué des bruits internes de l'organisme maternel (battements de cœur, respiration, digestion et pour ce qui nous intéresse, mouvements laryngés). Cependant, ce bruit de fond est moins important que ce que l'on pensait au départ (72 à 96 dB SPL, Lecanuet, 1997 : 24) et en l'absence de bruits gastriques, l'utérus est relativement tranquille (Lecanuet, 1995). Les pressions sont comparables à celles du milieu extérieur dans un environnement silencieux, notamment pour les basses fréquences (<300-500 Hz, de longueur d'onde supérieure au diamètre maternel ; Lecanuet, 1997 : 25). A cause de la paroi abdominale et du liquide amniotique qui se comportent comme un filtre passe-grave, les sons arrivent de manière atténuée. Cette atténuation croît avec la fréquence du signal, mais **préserve les fréquences basses correspondant aux caractéristiques prosodiques**. Ainsi, chez la brebis, Armitage et al. (1980) ont observé peu d'atténuation de la Fo entre 175 et 500 Hz. Toutes les informations prosodiques (mélodie, rythme, pauses) atteignent donc la cochlée du fœtus, ainsi que quelques mots isolés et bien articulés, mais qui restent peu intelligibles (« *papa, maman, bonjour...* »). **Les voix, bien que déformées, émergent clairement du bruit de fond et, en particulier, la voix maternelle**. Celle-ci est moins atténuée (8 dB) que les voix externes (20 dB, Querleu et al., 1988) et selon Lecanuet (1997 : 25), elle est même supérieure de 5 dB au niveau d'émission externe, à cause de la double transmission des sons (conduction osseuse et transmission aérienne). **Ce sont donc les éléments musicaux qui sont favorisés dans la perception que le fœtus a de son environnement sonore.**

2.1.2. Les fœtus et leur LM

2.1.2.1. Les voix de son environnement

Grâce au cardiographe qui permet d'observer les modifications du rythme cardiaque du fœtus et à l'échographe, qui permet de visualiser ses mouvements, les chercheurs peuvent développer des protocoles d'habituation pour étudier les préférences du fœtus pour un type de stimulation sonore et ses capacités à discriminer différents stimuli. Ainsi, une fois qu'il s'est habitué à un stimulus, on lui fait écouter un nouveau stimulus. Si son rythme cardiaque se modifie ou s'il se met à bouger, c'est qu'il distingue le nouveau stimulus du précédent (la mémoire et l'apprentissage sont donc mis en jeu). C'est avec cette technique qu'il a été possible de démontrer la sensibilité des fœtus proches du terme à la **variation inter-locuteurs**. Ils sont en effet capables de distinguer les caractéristiques vocales de deux personnes différentes, mais également les **voix de femmes des voix d'hommes**. La conversation exerce un effet calmant sur eux, mais ils deviennent particulièrement attentifs lorsqu'ils entendent la **voix de leur mère** (Gentil-Neymon, 1993). Après la naissance, le nouveau-né manifeste un certain nombre de préférences qui témoignent de son exposition prénatale : **il préfère les voix humaines par rapport aux autres sons** (DeCasper et Fifer, 1980). **Il préfère les voix féminines aux voix masculines** (Mehler et al., 1978) et, en particulier, **la voix de sa mère** (DeCasper et Fifer, 1980), même si celle-ci est filtrée et seulement après 12 heures de contact post-natal (Busnel, 1997 : 37). Si son père lui a parlé pendant la grossesse, il marquera une préférence pour sa voix (Pugin, 1983), sinon, il ne la reconnaîtra pas, même après un contact post-natal de 4 à 10 heures (DeCasper et Prescott, 1984). Selon Lecanuet (1995 : 34-35), l'expérience prénatale de l'enfant avec la voix de sa mère pourrait expliquer les résultats de Hepper et al. (1993), qui montrent que les nouveau-nés discriminent une voix produisant des contours prosodiques exagérés d'une voix normale parlant sans exagération, uniquement si cette voix est celle de leur mère.

2.1.2.2. Les éléments musicaux de la langue

En ce qui concerne les paramètres prosodiques, Nowik-Stern et al. (1998) ont montré que les prématurés regardent plus longtemps le haut parleur lorsqu'il diffuse une voix s'adressant à un bébé plutôt qu'à un adulte. **Ils préfèrent donc le langage adressé (LAE) dont les caractéristiques prosodiques sont fortement exagérées** (voir 2.4.1.1.).

Cette préférence n'est pas étonnante étant donné que la transmission de la prosodie est privilégiée pendant la vie fœtale. Une autre étude vient confirmer cette préférence : Busnel (1997) a montré, dans un travail concernant 200 mères dialoguant avec leur enfant avant et après la naissance, que la majorité des enfants se montrait plus sensible au langage leur étant adressé qu'au langage adulte. Dès 2 jours, des nourrissons préfèrent le LAE (Fernald, 1985 ; Cooper et Aslin, 1990). Par ailleurs, l'unité prosodique de base, la **syllabe est reconnue in utéro**. Dans une étude réalisée par Lecanuet et al. (1993), les fœtus se sont montrés capables de distinguer l'inversion « *biba* » / « *babi* ». Enfin, à 4 jours, ils ont pu **repérer les frontières prosodiques** qui différencient le groupe bi-syllabique « *mati* » dans « *panorama typique* » et « *mathématicien* » (Christophe et al., 1994).

S'ils font preuve d'une sensibilité envers les caractéristiques prosodiques de la parole, est-ce que les fœtus sont capables de mémoriser des extraits de parole auxquels ils ont été régulièrement exposés ? Mehler et son équipe ont mené une série d'expérimentation à Baudelocque entre 1986 et 1990. Ils ont montré que des bébés russes, français, anglais et italiens âgés de 4 jours auxquels leur mère avaient lu régulièrement une histoire dans leur LM pendant les cinq mois précédant l'accouchement s'arrêtaient de pleurer en entendant à nouveau l'histoire après leur naissance. Par ailleurs, l'histoire est reconnue même lorsqu'on lui applique un filtre passe-bas (Spencer et al., 1987). Et, lorsqu'elle est lue par une autre personne que la mère, elle est tout de même reconnue, ce qui implique que **le bébé se fie à des indices acoustiques et probablement prosodiques** (DeCasper et Spence, 1986). Par ailleurs, dans les expérimentations de Mehler, les bébés français ont réagi à leur LM pendant les phases d'habituation (alors que la voix n'était pas celle de la mère). Ils reconnaissent donc non seulement les caractéristiques sonores de l'histoire qui leur a été lue avant la naissance, **mais également leur LM**. Cette familiarité des nouveau-nés avec leur LM est confirmée avec des bébés « anglophones » capables de distinguer l'anglais de l'espagnol à 2 jours (Moon et al., 1993 ; Bertoni, 1993). Ils se montrent même capables de **distinguer les grands types de langues sur la base de la prosodie** (anglais-hollandais / espagnol-italien). A deux jours, le nourrisson est également **capable de reconnaître les différentes émotions** (joie, tristesse, colère) qui affectent une voix féminine enregistrée, **avec toutefois une préférence pour la joie**. Ces émotions ne sont reconnues que dans la LM de la mère et dans les deux langues pour les bébés bilingues.

2.1.3. Des fœtus mélomanes

2.1.3.1. Des réactions différentes selon le type de musique

Comme ils préféraient les voix aux autres types de stimuli, **les nouveau-nés préfèrent les tons musicaux par rapport au bruit** (Levarie et Rudolph, 1978). Tout comme la parole, **la musique exerce un effet calmant sur les fœtus** (Lecanuet et al., 2000) : les mouvements de 43 fœtus à terme (37 semaines de gestation) deviennent moins variables à l'écoute des berceuses de Brahms (n°40). Chapman (1975) a fait écouter les mêmes berceuses à des prématurés pendant 5 minutes lors de séances intervenant six fois par jour et il a constaté chez eux une **prise de poids plus importante avec la musique qu'avec la voix**. Mosser (1990) a montré la même chose avec 104 fœtus de 36 semaines, mais avec un apaisement plus rapide avec les chants grégoriens et les chants traditionnels qu'avec les berceuses. Une étude en cours de Kisilevsky, avec l'équipe de Lecanuet confirme ces résultats mitigés pour les berceuses avec des fœtus âgés de 38 semaines en sommeil actif. A l'écoute des berceuses de Brahms, ils observent quelques effets sur les fœtus au début et à la fin de l'extrait musical, mais peu pendant les extraits. **Les fœtus ne réagissent donc pas de la même manière selon le type de musique.**

2.1.3.2. Une mémoire musicale réactivable après la naissance

En ce qui concerne la mémorisation d'extraits musicaux, **on constate le même effet d'apprentissage que pour la LM**. Les nourrissons sont **capables de reconnaître une mélodie** qui leur a été chantée par la mère pendant sa grossesse d'une autre mélodie, et ce, même sans les paroles, sur « *la, la, la* » (Panneton, 1985²). Puisque l'information segmentale ne change pas, ils ne peuvent s'appuyer que sur des indices mélodiques. Hepper (1991) a montré que les nourrissons de 2 à 4 jours **reconnaissent le générique de la série de télévision « Neighbours », série que leur mère regardait régulièrement pendant la grossesse** : à son écoute, ils s'arrêtent de bouger et leur rythme cardiaque décroît. Dans des études plus anciennes, Feijoo (1975, 1981) avait fait écouter à 23 femmes enceintes (24 séances de relaxation, intervenant de la 14^{ème} à la 36^{ème} semaine de gestation) une phrase musicale de 12 secondes, extraite de « *Pierre et le Loup* » de Prokofiev.

² Cité par Lecanuet (1995 : 35).

Cette phrase est jouée par le basson, instrument dont la bande de fréquence est inférieure à 2000 Hz. Feijoo avait ensuite réalisé deux tests avec 15 enfants, le premier étant prénatal et intervenant une semaine après la dernière stimulation, le second intervenant immédiatement après la naissance (6 mn au plus tard). **A l'écoute de la phrase musicale interprétée par le basson, les fœtus se sont mis à bouger.** Quant aux nouveau-nés, ceux qui criaient s'arrêtèrent de pleurer (9 sur 15), ceux qui avaient les yeux fermés les ont ouverts (6 sur 15) et ils ont effectué moins de mouvements cloniques (de 14 à 4). En revanche, lorsque la phrase était passée à l'envers ou qu'elle était interprétée par un violon, elle n'était pas reconnue. Un an après la naissance, Feijoo (1983) a testé 7 des 15 bébés qui avaient participé à l'expérimentation initiale. Sur les 7, 6 enfants se sont calmés à l'écoute de la phrase de Pierre et le Loup.

Les travaux de Lamont³ à l'Université de Leicester confirment que cette mémorisation prénatale est réactivable près d'une année après l'exposition musicale. Ainsi, **des bébés âgés de un an sont capables de reconnaître la musique à laquelle ils avaient été exposés pendant le dernier trimestre précédant leur naissance, et ce, quel que soit le type de musique** (musique classique, musique du monde, musique pop ou musique reggae) :

« After the babies were a year old, they heard the prenatal music and other music that was matched for style, key, pace and loudness. For example, a baby who was exposed to UB40's Many River to Cross before birth heard this piece with another slow reggae track, Freddie McGregor's Stop Lovin' You [...] » (Lamont³).

Les sujets contrôles par comparaison ne manifestent aucune préférence particulière pour l'une ou l'autre musique. **Le tempo de la musique semble jouer un rôle important** puisque, selon l'auteur, **les bébés qui ont été habitués à des musiques aux rythmes rapides montrent des préférences plus marquées que les bébés qui ont été habitués à des musiques « lentes » :**

« the style of the music is not important – the babies recognize UB40 just as much as they do Mozart. But the pace of the music seems to be influential --the babies with faster music like Five's If Y a Getting' Down or the start of Vivaldi's Four Seasons show stronger preferences than the babies with slower music like Mozart's Adagio for Wind » (Lamont³).

³ Voir site Web : <http://news.bbc.co.uk/1/hi/health/1432495.stm>.

Cette « mémoire musicale » expliquerait le sentiment de familiarité qu'éprouvent certains musiciens professionnels, eux-mêmes fils de musiciens à l'écoute d'une pièce spécifique et qui

« déclarent avoir éprouvé un sentiment de familiarité en entendant ou en commençant une pièce nouvelle... qui s'avérait – après enquête – avoir été maintes fois jouée par leur mère pendant qu'elle les portait » (Lecanuet, 1995 : 37).

Lecanuet souligne que

« quoi qu'il en soit, on peut dire que l'expérience musicale prénatale contribue – de la même façon, et certains souhaiteront dire « plus que » n'importe quelle exposition à une stimulation acoustique structurée – à modeler les capacités auditives, à développer des préférences à long terme ou une sensibilité générale au type de sons rencontrés in utero » (Lecanuet, 1995 : 37).

2.1.4. La culture prénatale du fœtus

De Casper et al. (1994) ont montré que la reconnaissance prénatale pouvait intervenir de manière très précoce, entre la 30 et la 37^{ème} semaine de gestation. La reconnaissance d'un extrait sonore implique au préalable la mémorisation de cet extrait. Le fœtus fait donc preuve de capacités très précoces de mémorisation et d'apprentissage. L'imprégnation de l'enfant avec son environnement commence quatre mois avant sa naissance et il y a une continuité entre la vie fœtale et la vie post-natale. Pour Busnel (1994), la période de gestation constitue pour le fœtus une véritable « culture prénatale » pendant laquelle il va apprendre et se familiariser avec son futur environnement sonore et pour ce qui nous intéresse, avec sa LM et son SMM. La durée d'exposition au stimulus est fondamentale dans la formation de cette « culture ». Le nouveau-né semble particulièrement sensible aux caractéristiques prosodiques de sa LM et à la musique, ce qui n'est finalement guère étonnant, puisque cette information lui parvient de manière privilégiée pendant la fin de la vie fœtale. Mais, cette sensibilité aux éléments musicaux se prolonge-t-elle après la naissance ?

2.2. Sensibilité du bébé aux éléments musicaux : le « trésor inépuisable » des ressources enfantines

« Il semble que le nourrisson se fonde sur l'information signalant les unités syntaxiques importantes pour décomposer les stimulations acoustiques verbales et musicales. L'environnement acoustique intra-utérin et l'expérience post-natale du discours adressé au bébé fournissent les contextes d'apprentissage des indices nécessaires à la segmentation et à la structuration du langage et de la musique. Mais, le nourrisson est aussi porteur d'importantes prédispositions à percevoir et à acquérir ces indices. Dès les tous premiers jours, il est extrêmement sensible aux variations mélodiques et temporelles du langage et de la musique. (Fassbender, 1995 : 98).

2.2.1. Sensibilité au rythme

Vers l'âge de 2 à 3 mois, **le bébé est capable de différencier des séquences ayant des structures rythmiques différentes** (Demany et al., 1977). Cependant, la perception du rythme chez les adultes est plus élaborée, puisqu'ils sont capables, à partir d'une séquence, d'abstraire une structure rythmique qui restera identique malgré les variations de tempo. Est-ce que les bébés sont capables de la même capacité de généralisation ? Trehub et Thorpe (1989) montrent qu'ils sont capables de différencier des séquences de 3 sons (X XX ou XX X) et de 4 sons (XX XX et XXX X), et ce, en dépit des variations de tempo et des variations de hauteur. Les résultats montrent cependant que cette tâche leur est difficile car le niveau de performance, bien que significatif, reste faible. Néanmoins, ils suggèrent que **chez le nourrisson, le traitement est relationnel au niveau temporel**. Cette capacité est centrale dans la perception de la parole, car elle permet à l'auditeur d'extraire le même contenu verbal en dépit des variations de débit. **Cette « normalisation » fonctionne également dans la perception des séquences musicales**. Cependant, ce type de traitement n'est pas spécifique à l'être humain, puisque les étourneaux (Hulse et al., 1984) et les pigeons (Miller et Liberman, 1979) sont également capables de normaliser des séquences rythmiques.

2.2.2 Sensibilité aux contours mélodiques

2.2.2.1. Les contours mélodiques de la musique

Nous l'avons vu dans la partie n° 1.2.2., le type de traitement appliqué à la mélodie par le système perceptif dépend fortement de l'expérience que le sujet a contractée au contact de son milieu sonore, qu'il soit langagier ou musical (traitement par contours ou par intervalles). Qu'en est-il des capacités du bébé à traiter les mélodies ? A quel type de stratégie a-t-il

recours ? Pour le savoir, Trehub et al. (1984) ont étudié la capacité de bébés de 6 à 7 mois et d'adultes à détecter des mélodies modifiées par rapport à une mélodie standard composée de 6 sons. Les trois variantes débutent et se terminent sur les mêmes notes, mais la première est une transposition de la mélodie standard dans une autre tonalité, la seconde comprend des changements de notes qui préservent le contour et la dernière, des changements qui ne préservent pas le contour (cf. figure n°47).

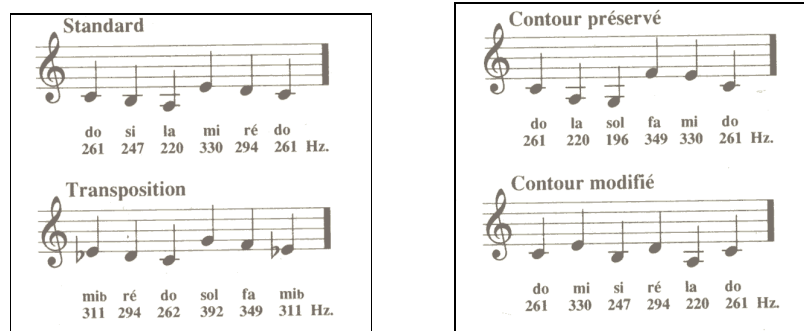


Figure n°47 : une mélodie standard, suivie des trois types de changement à détecter : transposition, changement préservant le contour et changement modifiant le contour (emprunté à Trehub et Trainor, 1994 : 313).

Comme les auditeurs adultes, les bébés se montrent aptes à détecter les changements affectant le contour, mais échouent à détecter les transpositions de la mélodie initiale et les changements préservant le contour. Ces deux types de modification passent inaperçus, comme si c'était la mélodie standard qui était reprise. Leur stratégie est donc similaire à celle des adultes pour le traitement des mélodies non familières. Ils traitent les contours de manière globale, en ignorant les détails de ces contours. Ces résultats montrent que **les bébés utilisent une stratégie de traitement global où l'information sur le contour domine celle qui concerne la hauteur absolue et les intervalles**. Pour Trehub et Trainor,

« ces résultats suggèrent que le traitement de contour de hauteurs fonctionne peut-être chez les nourrissons comme un important mécanisme d'organisation perceptif qui dirige leur segmentation des patterns auditifs complexes » (Trehub et Trainor, 1994 : 317).

2.2.2.2. Les contours d'intonation de la langue

Dès l'âge de 2 à 3 mois, **les bébés sont capables de discriminer des syllabes au contenu segmental identique, mais affectées de contours d'intonation différents**, montants et descendants (Morse, 1972 ; Kuhl et Miller, 1982). Menyuk (1994) obtient des résultats similaires avec des bébés plus jeunes, âgés de 6 à 7 semaines. Best (1993) montre que **les bébés anglophones âgés de 2 mois discriminent certes les contrastes prosodiques de l'anglais, mais seulement lorsque la parole est adressée aux enfants (LAE)**. Ils sont donc capables de différencier les contours prototypiques de leur LM et de répondre de manière adéquate aux messages qu'ils véhiculent, tels que l'incitation au réveil, à l'attention et au contact visuel (Papousek et al., 1990).

Plus tard, ces capacités s'étendent à d'autres langues : **les bébés de 6 à 8 mois sont capables de discriminer les contrastes prosodiques affectant la parole adressée à l'enfant dans deux langues différentes**, leur LM, l'anglais et une langue étrangère, l'espagnol (Best, 1993). Cependant, **vers 10-12 mois, cette capacité semble s'altérer**, puisque les bébés anglophones ne se montrent plus capables de discriminer les contrastes prosodiques de l'anglais, ni ceux de l'espagnol (Best, 1993). Ces résultats pourraient s'expliquer par le fait que le LAE utilisé dans l'expérimentation ne correspond pas au niveau des bébés de 10 à 12 mois, mais à celui de bébés plus jeunes. En effet, les caractéristiques du LAE évoluent avec l'âge et les capacités de l'enfant (Snow, 1977 ; voir 2.4.2.2.), et il eut peut-être fallu utiliser un autre type de LAE. Une autre étude confirme cette hypothèse, puisque, lorsque le LAE utilisé est adressé à un bébé de 10-12 mois, les bébés du même âge sont capables de discriminer les contours marquant l'affirmation et la question en anglais et en mandarin. Contrairement au niveau segmental (voir 3.1.), **il ne semble pas se produire de réorganisation perceptive du système prosodique de la LM au niveau des intonèmes**. Ces effets commenceraient à apparaître plus tard, vers 18 mois, au moment où l'enfant utilise la syntaxe de manière plus approfondie (au niveau de la phrase).

2.2.3. Sensibilité à la prosodie affectant les mots

Un certain nombre d'études ont mis en relief une sensibilité particulière des bébés à la prosodie affectant le mot, comme l'accentuation lexicale en anglais par exemple. Une étude récente a montré que des nouveau-nés étaient **capables de discriminer des stimuli multisyllabiques différant uniquement par la place de l'accent**, correspondant à des mots italiens (Sansavini et al., 1997). Cette capacité avait déjà été démontrée chez des bébés anglophones de 1 à 2 mois (Spring et Dale, 1977). Dès le premier mois, les bébés se montrent sensibles aux mots fortement accentués et à la terminaison des phrases (Menyuk, 1977). A 2 et 3 mois, ils ne font d'ailleurs **pas attention aux mots prononcés sans exagération de la prosodie** (Kagan et Lewis, 1965 ; Eimas, 1975). Cette sensibilité à la prosodie des mots va progressivement évoluer vers un traitement plus phonétique, comme le montre une étude de Jusczyk et al. (1993a). A 6 mois, les bébés américains sont capables de distinguer des mots anglais de mots néerlandais, et ce, même avec un filtrage passe-bas, mais uniquement sur la base de l'information prosodique. Par contre, à 9 mois, ils ont recours également aux différences phonétiques entre les mots. A 9 mois, les bébés anglophones marquent une préférence pour les mots anglais accentués sur la syllabe initiale (Jusczyk et al., 1993b). Ce type d'accentuation représente le pattern accentuel dominant dans cette langue (voir chapitre V, 2.1.1.2.). Parallèlement à cette sensibilité prosodique, la compréhension du sens des mots se développe, puisque les bébés sont capables vers 10-12 mois de comprendre une trentaine de mots en contexte. Et au même moment, apparaissent les premiers mots au sein du babillage mixte. Selon Bertoncini et De Boysson-Bardies (2000 : 104),

« au cours du développement perceptif, la prosodie peut aider à indexer différents niveaux d'unités fonctionnelles de la langue. En association avec les composants syllabiques, elle fournirait le format des mots, en offrant les premiers indices pour la segmentation et le groupement des syllabes en unités-mots ».

C'est donc la prosodie qui permet à l'enfant de se sensibiliser aux mots de sa LM, puis d'accéder progressivement à l'information segmentale. Elle joue le rôle d'un guide qui l'incite à porter son attention sur l'information importante.

2.2.4. Capacités communes de segmentation pour la LM et le SMM

Dans le premier chapitre, nous avons évoqué les préférences du bébé pour les extraits de parole segmentés correctement : ils écoutent plus longtemps les extraits interrompus aux limites des phrases, plutôt qu'à l'intérieur des phrases. Cette capacité a été relevée dès l'âge de 4 mois et demi (Juszyk, 1989) et elle a également été attestée à 6 mois (Hirsh-Pasek et al., 1987 ; Kelmer Nelson et al., 1989 ; Nazzi et al., 2000). Dans une étude réalisée par Stern et al. (1983), des bébés de 7 à 10 mois marquent une **préférence pour le langage adressé à une fillette de 19 mois, lorsque les pauses sont insérées aux frontières des propositions**. Il est très important de souligner que cette préférence ne fonctionne que pour le LAE, car dans une autre étude concernant des bébés du même âge, les auteurs ne relèvent aucune préférence si le matériel utilisé concerne le langage adressé à l'adulte (Kelmer Nelson et al., 1989). Le LAE semble donc posséder des propriétés apparemment très attractives pour le petit enfant, et conserve ce pouvoir pendant un temps relativement long. Une autre étude réalisée par Juszyk et al. (1992) montre que la préférence pour des extraits segmentés aux frontières linguistiques persiste chez les bébés de 9 mois et ce, même si les extraits sont filtrés pour ne laisser passer que l'information prosodique (à l'aide d'un filtre passe-bas).

De manière troublante, cette capacité de segmentation est similaire pour la musique et apparaît également de manière très précoce : **dès 4 mois et demi, les bébés préfèrent des extraits musicaux** tirés des menuets de Mozart, **segmentés aux limites des phrases musicales**, plutôt qu'à l'intérieur de ces phrases (Juszyk et Krumhansl, 1993). Si l'extrait musical est inversé, la complexité du contour est conservée, ainsi que les hauteurs, mais la direction du contour et la durée des notes sont modifiées. Les bébés préfèrent la segmentation naturelle. Les auteurs en concluent donc que les changements de contour, et notamment son abaissement, et l'allongement des sons précédant les frontières sont des éléments d'importance pour diriger la segmentation chez le bébé.

2.3. L'amorçage de l'attention par les éléments musicaux

2.3.1. Richesse et redondance du signal

Nous avons vu dans la partie 1.2.2. que les bébés manifestaient des capacités générales de structuration des ensembles acoustiques. Ces capacités leur permettent de repérer des régularités structurelles à l'intérieur de la variabilité qui caractérise le flux sonore.

Or, le signal comporte lui-même un grand nombre d'informations qui vont permettre à l'enfant de guider son attention sur des événements importants. La langue et la musique sont organisées hiérarchiquement (chapitre I, partie n°3.2.) et les signaux saillants dans le signal correspondent à des frontières linguistiques et musicales importantes, qui délimitent notamment les différents constituants (voir paragraphe précédent, 2.2.4.). Dans le flux sonore, ces signaux constituent donc des éléments proéminents sur lesquels l'attention de l'enfant va pouvoir se fixer, un peu comme des prises sur un mur d'escalade, avant que son système perceptif procède à un traitement plus en profondeur. Et c'est justement grâce à la prosodie que ces signaux sont mis en proéminence. En effet, **la prosodie contient de nombreux indices sur les frontières linguistiques importantes** (Morgan et al., 1987). D'après la théorie dite du « bootstrapping prosodique », appelée également « bootstrapping du signal » (Gleitman et Wanner, 1982 ; Peters, 1983 ; Gleitman et al., 1988), **la prosodie guiderait ainsi l'attention sur les unités des différents niveaux d'organisation** (phrases, propositions, différents syntagmes, mots), ce qui permettrait à l'enfant d'élaborer son savoir sur l'organisation syntaxique de sa LM :

« an infant who is intentionally circumscribed utterance segment as potential syntactic constituent would be at considerable advantage in learning the syntactic rules of his language »
(Read et Schreiber, 1982 : 26).

Cette hypothèse suppose que l'enfant ait des capacités perceptives suffisantes pour détecter les signaux prosodiques, mais les travaux évoqués dans les parties précédentes prouvent que la sensibilité du bébé à ce type de signaux est très forte. Elle suppose également que le signal de parole contienne des informations et plus celles-ci seront riches et nombreuses, moins les arguments en faveur de capacités « programmées » pour le traitement de la langue seront justifiés. Quels sont ces signaux qui orientent l'attention du bébé à l'intérieur du flux sonore ?

2.3.2. Les signaux d'amorçage dans la langue

Gleitman et Wanner (1982) pensent qu'il s'agit des **syllabes accentuées**, car pour les petits anglophones, l'apprentissage des syllabes atones est beaucoup plus long. Pye (1983) a découvert que les enfants Quiché Maya apprennent les inflexions accentuées avant les racines des verbes, qui sont inaccentuées. Mais, les syllabes accentuées ne se présentent pas acoustiquement de la même manière dans les différentes langues. En français par exemple, les syllabes atones ne subissent pas de réduction, contrairement à l'anglais (cf. chapitre V, différences d'accentuation entre le français et l'anglais).

Les signaux qui vont guider l'attention de l'enfant vont donc varier selon les langues et les spécificités de leur organisation prosodique. Ainsi, en anglais, la **frontière entre les propositions est marquée par une pause, une augmentation ou diminution de la fréquence fondamentale ou un allongement de la syllabe précédant la frontière de la proposition** (Klatt et Cooper, 1975 ; Luce et Charles-Luce, 1983 ; Kakatani et Dukes, 1977). Cooper et Paccia-Cooper (1980) ont relevé des **pauses plus longues aux frontières des propositions**, tandis que **les groupes intonatifs ont tendance à correspondre aux unités syntaxiques majeures telles que les propositions** (Bolinger, 1978 ; Cruttenden, 1986 ; Selkirk, 1984). Bernstein Ratner (1986) a relevé des allongements plus prononcés au niveau segmental aux frontières des propositions. Cependant, Fisher et Tokura (1996) n'ont pas identifié de marqueurs prosodiques plus petits que ceux relevés pour la proposition. Enfin, Jusczyk et al. (1992) ont remarqué une corrélation entre hauteur et durée dans la frontière entre le sujet et le prédicat. L'information sur les unités semble non pas portée par un seul type de signal, mais par une véritable constellation de signaux (intégration de signaux multiples mis en jeu dans l'apprentissage des catégories syntaxiques).

2.3.3. Une théorie de l'amorçage valable pour la musique ?

La théorie du bootstrapping a été élaborée pour la langue, mais les similitudes de segmentation constatées pour la langue et la musique (voir 2.2.4.) et la sensibilité de l'enfant aux éléments musicaux en général suggèrent la possibilité de l'étendre au traitement de la musique. On remarque en effet une forte analogie entre les signaux utilisés pour marquer la structure de la langue et celle de la musique : la chute de la hauteur avant les frontières musicales est similaire à celle que l'on retrouve à la fin des énoncés dans la parole. De la même manière, l'allongement de la durée des notes précédant les frontières musicales (Jusczyk et Krumhansl, 1993) se retrouve dans la parole (Scott, 1982). Cette analogie suggère des mécanismes communs de traitement de la langue et de la musique.

2.4. Ajustements vocaux inconscients adaptés à la sensibilité « musicale » du bébé

2.4.1. Préférences de l'enfant pour le langage et le chant adressé

2.4.1.1. Le langage adressé

Lorsqu'un adulte s'adresse à un enfant, et en particulier, la mère à son enfant, il adopte une façon de parler particulière, extrêmement simplifiée au niveau de la syntaxe, du sens et du répertoire phonologique : le langage adressé à l'enfant, également appelé « mamanaï » (motherese ou baby talk en anglais). Or, selon la théorie de l'amorçage prosodique, la segmentation est opérée sur la base de l'information prosodique. Le LAE se caractérise justement par une remarquable musicalité (Papousek et Papousek, 1981 ; Fernald, 1989), qui s'exprime par l'intermédiaire de paramètres prosodiques fortement exagérés. Selon Papousek et Papousek (1981 : 171), le langage adressé constitue l'expérience musicale la plus précoce de l'enfant. Papousek ajoute dans un autre article (1995 : 104), que :

« les parents et l'entourage du nourrisson font preuve d'une propension intuitive à parler au nouveau-né et à lui prodiguer ce que l'on peut assimiler aux premières formes de l'éducation musicale ».

Nous savons que les fœtus préfère déjà le LAE dans le ventre de leur mère. Cette préférence persiste pendant la première année de vie. Les bébés de 2 jours préfèrent ce type de discours (Fernald, 1985 ; Cooper et Aslin, 1990), même si les contours sont synthétisés en ondes sinusoïdales (Fernald et Kuhl, 1987). Fernald (1985) a testé 48 bébés de 4 mois ; 33 tournent plus souvent la tête en direction du haut parleur qui diffuse du LAE. Papousek et Papousek (1981 : 173) soulignent les manifestations de joie et de plaisir suscitées chez le bébé à l'écoute du LAE. Fernald et Kuhl (1981) ont isolé les paramètres acoustiques qui permettent de faire la différence entre le LAE et la parole adressée à un adulte. Ils ont fait ainsi apparaître que la préférence pour le LAE était conservée uniquement pour la Fo, paramètre probablement hautement saillant pour les bébés, plus saillant que le rythme, les formants et l'amplitude (Fernald, 1984 ; Fernald et Kuhl, 1987 ; Katz et al., 1996). Ils préfèrent d'ailleurs un discours où la Fo varie (Fernald, 1985 ; Fernald et Kuhl, 1987), et les contours dont la Fo monte rapidement (Fernald, 1989).

A 8 mois ½, les bébés continuent à préférer le LAE, alors que les adultes ne manifestent aucune préférence pour ce type de langage (Kelmer Nelson et al., 1989). Cette préférence se prolonge jusqu'à 19 mois, âge où la progression de l'enfant dans l'acquisition de sa LM ne semble plus dépendre de ce style de parole.

2.4.1.2. Le chant adressé

De la même manière que pour le langage, les adultes adaptent leur façon de chanter aux petits enfants. Lorsqu'une mère chante à son enfant (« chants de nourrice » ou « nursery songs »), elle le fait en effet d'une manière particulière, très différente du chant « adulte ». Par analogie avec le LAE, on parle de chant adressé à l'enfant (désormais CAE). Il existe plusieurs types de chants adressés à l'enfant. Les **berceuses** sont destinées à endormir l'enfant et pour arriver à ce résultat, les contours descendants y sont prédominants. Elles peuvent se répéter infiniment jusqu'à ce que le bébé s'endorme. Un autre type de chant possède une fonction inverse à celle des berceuses : les **chansons de jeux** qui sollicitent l'activité motrice de l'enfant. Pour parvenir à ce résultat, les contours montants y sont majoritairement utilisés. Le chant adressé a beaucoup intéressé les chercheurs qui travaillaient sur le LAE, à cause des similitudes entre les deux types d'adaptation vocales. Le CAE semble en effet jouer un rôle aussi important dans l'acquisition du SMM, que le LAE pour la LM. L'étude de leurs similitudes pourrait contribuer à mettre à jour des processus généraux mis en œuvre dans l'acquisition valables pour la langue et pour la musique. Comme ils préfèrent le LAE, les enfants marquent une préférence pour le chant adressé à l'enfant (CAE) plutôt que pour le chant non adressé (Trainor, 1996) et notamment, ils préfèrent les berceuses aux chants de jeux et aux chants d'adultes, ainsi que le chant de la mère plutôt que celui du père (Trehub et Henderson, 1994). Là aussi, la hauteur paraît jouer un rôle fondamental dans cette préférence : entre deux versions de la même chanson, les enfants préfèrent la version chantée avec une hauteur plus élevée (Trainor et Zacharias, 1999).

2.4.2. Caractéristiques du langage adressé à l'enfant

Les caractéristiques acoustiques du LAE ont été principalement décrites par Papousek et Papousek (1981), Stern et al. (1983) et Fernald (1989). Dans le LAE, la **hauteur de la voix** est **plus élevée** de 3 à 4 demi-tons. La mère parle comme si sa cavité vocale était de la même taille que celle de son bébé, avec une voix palatale.

On observe une grande **continuité dans l'évolution de la mélodie**, comparé à sa variabilité dans la parole adulte (Fernald et Simon, 1984) : il y a **beaucoup de glissandi** et peu de changements brusques. Les **contours sont très exagérés, mais leur configuration reste simple et unidirectionnelle** (de type haut-bas ou bas-haut) et ils sont **constamment répétés par la mère**. Cette continuité et cette simplicité les rendent particulièrement faciles à traiter par l'enfant. Ils lui permettent notamment de **localiser facilement la source**, c'est-à-dire la personne qui lui parle et de maintenir son attention fixée sur cette source :

« The exaggerated intonation of mothers' speech, with its greatly expanded pitch range and high degree of pitch continuity, thus maximizes both perceptual contrast, necessary for engaging and maintaining infant attention, and perceptual coherence, facilitating the task of following the voice of a single speaker » (Fernald, 1984 : 13).

Ces contours sont composés de « bons intervalles », c'est-à-dire **d'intervalles aux rapports fréquentiels simples**, et donc facilement décodables par le système perceptif (Burns et Ward, 1982 ; Rakowski, 1990) : les tierces, les quarts, les quintes et les octaves (Fernald, 1976) y sont prédominants (pour les non musiciens, se reporter à la grille d'intervalles). Le **degré de mélodicité du LAE** est **élevé**, ce qui lui confère une grande musicalité (voir chapitre 1, partie n°1.2.4.). Cinq ou six contours prototypiques reviennent régulièrement dans le LAE. Les contours montants servent à attirer l'attention du bébé, tandis que les contours en cloche, dont la tessiture est plus large, servent à maintenir son attention (voir figure n°48). Ils sont parfois utilisés dans des questions, même si les phrases ne sont pas interrogatives (Ferguson, 1977). La tessiture tend à être plus étroite dans les contours descendants, de manière à calmer l'enfant (Papousek et Papousek, 1981).

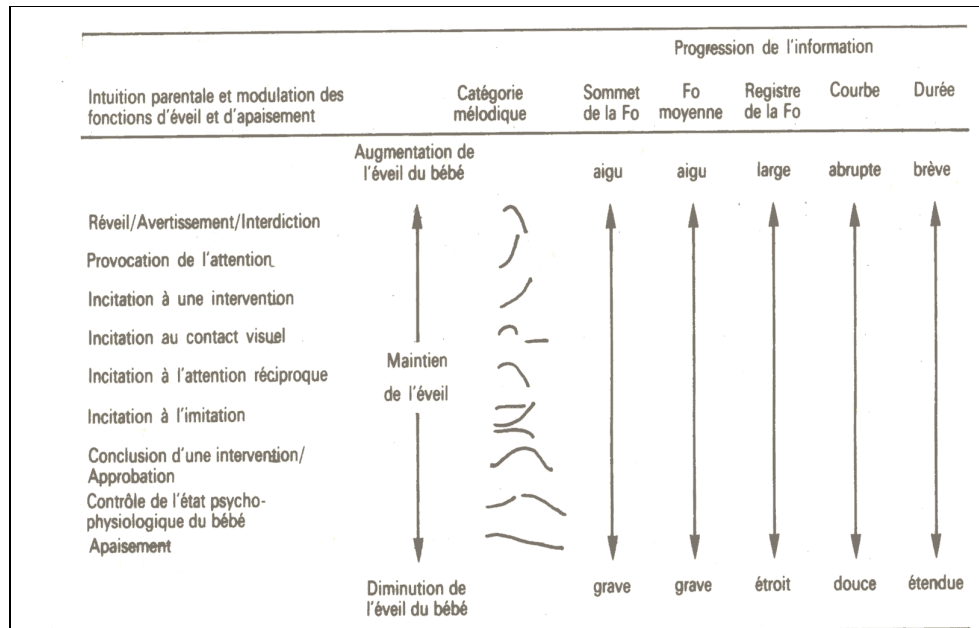


Figure n°48 : les mélodies du LAE (emprunté à Papousek, 1995 : 112)

Ces contours correspondent de façon remarquable aux principes de base de la perception de la forme, en particulier aux lois de proximité et de similitude (Fernald, 1984) et à la loi de destin commun (Fassbender, 1995). Dans le LAE, la plupart des énoncés sont courts, dépourvus de signification et incomplets d'un point de vue grammatical, mais ils tendent à être complets du point de vue intonatif (Broen, 1972). En ce qui concerne les autres paramètres prosodiques, les énoncés sont produits avec plus de régularité et plus de rythmicité. Le débit s'adapte à l'état du bébé : s'il est attentif, le débit sera plus lent et les énoncés plus rythmiques ; en revanche, le débit s'accélère lorsque le bébé n'est pas attentif. Il ralentit lorsque l'enfant est en train de s'endormir (Papousek et Papousek, 1981). Non seulement, on constate plus de pauses entre les différents segments, mais de plus, ces pauses sont prolongées. Enfin, le LAE se caractérise par de grands changements d'amplitude.

Le niveau segmental est complètement déformé au profit des exagérations produites au niveau de la prosodie. Les mots sont transformés et remplacés par des diminutifs, comme « *doggie* » pour « *dog* » en anglais et par des reduplications (« *chachat* » pour « *chat* » en français). Tout se passe comme si le niveau segmental était délaissé, car secondaire au début de l'acquisition du langage. En revanche, les voyelles sont prolongées comme elles le seraient dans une chanson. A l'intérieur de la syllabe, c'est la voyelle qui porte le contour mélodique. Or, dans le LAE, l'exagération des contours contribue à mettre les voyelles en relief. Par ailleurs, le grand nombre de monosyllabes de type V ou CV suggère une importance de ces sons dans l'acquisition (voir ce chapitre, partie n° 3.1.2.1.).

Kuhl et al. (1997) ont étudié les voyelles du langage adressé à des bébés américains, russes et suédois âgés de 2 à 5 mois. Il est apparu que les mères utilisaient la prosodie pour attirer l'attention de leur bébé sur les trois voyelles [i] [a] et [u] et consacrent du temps à obtenir une réponse. Hormis les travaux ponctuels de Robin, nous ne disposons pas encore de description complète du LAE pour le français.

2.4.2.1. Une pédagogie « inconsciente »...

L'ensemble de ces ajustements vocaux censé s'adapter aux capacités de traitement de l'enfant dans un but d'acquisition peut être qualifié de « pédagogie » à la différence près, qu'une pédagogie est un processus conscient alors que ces adaptations sont totalement inconscientes de la part des parents. Ces ajustements se font en effet de manière intuitive et naturelle et ils sont très rarement utilisés en dehors du contexte d'interaction avec un bébé (Fernald et Simon, 1984). Ce type de langage est également utilisé par des enfants âgés de 2 à 4 ans lorsqu'ils s'adressent à des enfants plus jeunes ou à des objets ersatz d'animés, tels que poupées ou peluches. Dans un contexte différent, on retrouve des ajustements vocaux similaires dans ce que Giacomi et Hérédia (1986) appellent la « coopération exolingue ». Cette coopération se met en place entre l'enseignant de langue et l'apprenant étranger. Comme la mère lorsqu'elle parle à son bébé, l'enseignant exagère l'intonation et marque le rythme lorsqu'il s'adresse à l'apprenant, sans doute pour faciliter sa compréhension et son apprentissage. Dans les deux cas, le locuteur adapte sa manière de parler aux capacités de son interlocuteur pour le « tirer » progressivement vers un savoir plus complexe. C'est donc un procédé pédagogique par excellence. En ce qui concerne le LAE, les adultes sont tout-à-fait capables de reconnaître ce type de langage d'après ses caractéristiques prosodiques amplifiées ; ils sont même capables de différencier un LAE produit en présence de l'enfant, d'un LAE produit en l'absence de l'enfant (langage adressé à une poupée par exemple). Et pourtant, lorsqu'à leur tour, ils produisent ce type de discours, ils n'ont pas conscience de produire ces ajustements. La mère en particulier, rencontre des difficultés à les reproduire en l'absence de l'enfant (Fernald, 1976) : « *more than a melody. It's a real tutorial and they didn't even know they were doing it* » (Kuhl et al., 1997).

2.4.2.2. ...qui évolue avec l'âge de l'enfant

Les caractéristiques du LAE ne restent pas statiques et la mère adapte son parler à mesure que l'enfant grandit : elle insère de plus en plus de mots isolés, de pauses et elle interrompt ses phrases beaucoup plus souvent (Broen, 1972). Ces caractéristiques se renforcent jusqu'à ce que le bébé atteigne l'âge de 4 mois, pour ensuite s'atténuer. C'est à ce moment justement que la dyade mère-enfant connaît une période de face à face très intense. La mère utilise les contours de hauteur montants pour attirer l'attention de son bébé lorsqu'il regarde ailleurs, puis utilise des contours en « cloche » pour maintenir le contact oculaire avec lui (Stern et al., 1982), et ce, jusque vers 6 mois. Plus tard, lorsque le bébé atteint 14 mois, elle accentue le prédicat, c'est-à-dire l'information nouvelle véhiculée par l'énoncé, grâce à une hauteur plus élevée et une plus grande intensité. Elle place les mots-clés en position finale (Fernald et Mazzie, 1991), de manière à ce que les enfants les repèrent plus facilement. Au même moment, les capacités de compréhension du bébé concernant les mots de sa LM se développent rapidement. Depuis l'âge de 8 mois, il est en effet capable de reconnaître les mots en dehors du contexte et à 11 mois, son attention se fixe plus longtemps sur les mots familiers, même s'ils sont altérés au niveau phonologique. Enfin, c'est vers 12 mois qu'apparaissent les premiers mots isolés, au sein du babillage mixte. Selon Kuhl et al. (1997), les voyelles et les consonnes sont prononcées par la mère de manière plus distincte lorsque l'enfant commence à parler (hyperarticulation), que lorsqu'il est plus jeune ou au contraire plus âgé (Malsheen, 1990 ; Ratner, 1984). Le parler « bébé » s'adapte donc intuitivement aux capacités perceptives du bébé et doit évoluer en fonction du développement :

« to be effective, the communicative signals of the mother must be well matched to the perceptual capabilities and limitations of the young » (Fernald, 1989 : 10).

Il se crée donc un jeu de miroir entre la mère et son enfant : la mère rend accessible à son enfant des informations sur les unités linguistiques importantes dans le signal de parole et continue à le faire jusqu'au moment où elle « ressent » que le bébé est prêt à passer à une étape plus difficile dans l'acquisition de sa langue.

2.4.3. Caractéristiques du chant adressé à l'enfant

Lorsque la mère chante à son enfant, elle le fait d'une manière beaucoup plus expressive (Trehub et al., 1997). Le ton est plus aimant (Trainor, 1996) et la qualité de la voix est émotionnellement plus engageante (Trehub et al., 1997). Cette différence d'interprétation est à ce point importante que des auditeurs adultes sont capables d'identifier le chant produit en présence de l'enfant de celui produit en son absence, tout comme ils étaient capables de le faire pour le LAE. Ainsi, dans une étude réalisée par Trehub et al. (1993a), des adultes américains sont capables d'identifier les chansons américaines interprétées en présence de l'enfant, avec un taux de réussite de 90 %. Ils sont également capables de la même identification pour une culture musicale et une langue différente, l'hindi, avec certes un taux de réussite moins élevé, mais toutefois supérieur au hasard (Trehub et al., 1993b). Même si les parents tentent de simuler la présence de l'enfant, la version produite reste identifiable par les auditeurs, même dans une culture étrangère (Trehub et al., 1997). Enfin, le CAE semble s'adapter au développement perceptif de l'enfant puisqu'une comparaison entre des chansons adressées à des bébés et des chansons adressées à des enfants d'âge préscolaire relève que les chansons pour les pré-scolaires sont articulées plus clairement (Bergeson et Trehub, 1999).

2.4.3.1. Les berceuses

La hauteur des berceuses est plus élevée que celle des autres chansons (Trainor et al., 1997). Bergeson et Trehub (1999) ont réalisé une étude acoustique comparative de chants adressés à des bébés et des chants adressés à des enfants d'âge préscolaire. Ils n'ont relevé aucune différence significative entre les deux types de chant concernant le tempo, la durée, les pauses entre les phrases et la durée des voyelles. En revanche, la différence de hauteur était très significative, les chants adressés aux bébés étant plus hauts de 3 à 5 demi-tons, avec une différence de hauteur modale d'un demi-ton. Ce registre semble s'accorder à la sensibilité des bébés qui préfèrent les chants dont le registre est plus élevé (Trainor et Zacharias, 1999). En revanche, la tessiture est étroite, la tonalité douce et les contours répétitifs. La principale fonction des berceuses étant d'endormir l'enfant, elles se caractérisent par une prédominance de contours descendants, un tempo lent et un rythme répétitif. Par ailleurs, les berceuses sont chantées pianissimo. Au niveau segmental, la mère privilégie l'effet sonore sur la signification et sa prononciation est relativement indistincte. Cependant, les voyelles sont prolongées. La mère redouble des mots et répète des séquences de mots familiers.

Est-ce que les adultes sont capables de reconnaître les berceuses d'autres chansons ayant un tempo et un style musical similaire ? Trehub et al. (1992) ont apparié 30 berceuses provenant de différentes cultures à 30 chansons équivalentes. Les adultes ont été capables de reconnaître les berceuses avec un taux de réussite de 63 %, et ce, même après l'adjonction d'un filtre passe-bas avec une fréquence de coupure de 500 Hz et même après la resynthèse de la ligne mélodique. D'après Unyk et al. (1992), la simplicité des contours et leur configuration majoritairement descendante constitueraient les caractéristiques responsables de leur identification.

2.4.3.2. Les chants de jeux

Les chants de jeux ont une autre fonction que les berceuses, puisqu'ils doivent susciter l'activité motrice du bébé. Pour cette raison, ils sont composés en majorité de contours montants. Par rapport au chant non adressé à l'enfant, la hauteur des chants de jeux est plus variable. Les syllabes accentuées sont plus longues et plus intenses et on constate plus d'énergie dans les basses fréquences. En ce qui concerne le rythme, le tempo est plus lent et les pauses plus longues (Trainor et al., 1997 ; Trehub et Trainor, 1998). Ces caractéristiques acoustiques permettent à l'auditeur de les identifier comme plus brillantes, plus rythmées et plus staccato (interprétées en détachant les notes les unes des autres), avec des consonnes plus proéminentes et chantées avec un ton de voix souriant (Trainor et Rock, 1997). Les auditeurs adultes sont capables de faire la différence entre les berceuses et les chants de jeux sans aucune difficulté. Ils reconnaissent en effet l'intention de la chanson avec un taux de réussite de 100 % (Trainor et Rock, 1997). Ces résultats ne sont guère surprenants car les différences formelles sont très grandes entre les deux types de chants. Ces différences ne sont d'ailleurs pas seulement acoustiques, mais également visuelles, car ils sont capables de les différencier seulement à partir de l'image de la mère (alors que le son est coupé), avec un taux supérieur au hasard (60 %).

2.4.4. Points communs entre le LAE et le CAE

A la croisée du langage adressé et du chant adressé, on trouve des caractéristiques acoustiques communes qui pourraient bien se révéler fondamentales dans le processus d'acquisition. Il semble que la hauteur joue le rôle le plus important. Outre leur hauteur plus élevée, le LAE et le CAE se caractérisent par des contours simples et unidirectionnels, continuellement répétés par l'adulte.

Dans les deux cas, ces contours sont utilisés pour les mêmes fonctions : soit solliciter l'activité motrice / l'attention du bébé, avec des contours montants, soit calmer / endormir le bébé, avec des contours descendants. Pour Fernald et al. (1989), ces contours sont universellement liés à des contextes de soins maternels et il est troublant de rencontrer des ajustements vocaux similaires au LAE dans un autre contexte de soin, celui des infirmières avec leur malades. En effet, celles-ci utilisent un registre plus élevé et une tessiture plus large lorsqu'elles s'adressent à leurs patients et ce type de parole est reconnu par les auditeurs (Journiac, 2002). Au niveau du rythme, le débit du LAE et le tempo du CAE sont plus lents. Quant au niveau segmental, il est déformé au profit de l'effet sonore. Enfin, les voyelles sont prolongées et mises en valeur par les contours mélodiques et on rencontre beaucoup de répétition au niveau des mots. Les enfants préfèrent le langage et le chant qui leur sont adressés. Les caractéristiques du LAE et du CAE sont tellement différentes lorsqu'ils sont produits en présence ou en l'absence de l'enfant, que les adultes reconnaissent la situation immédiatement. Enfin, dans les deux cas, les ajustements vocaux s'adaptent à l'âge de l'enfant, évoluant en même temps que lui. Les adultes identifient le chant et la parole adressés même dans une langue et dans une culture qui ne leur sont pas familières. Cette pédagogie « inconsciente » se retrouve en effet dans beaucoup de cultures différentes et de langues différentes (chinois mandarin, japonais, suédois, italien, allemand, anglais, français...). Est-ce que ces adaptations constitueraient des universaux dans le développement de l'être humain ?

3) D'une base commune à une différenciation progressive

3.1. Perception

Nous avons vu dans les parties 2.1. et 2.2. que le bébé manifestait de façon très précoce une grande sensibilité au rythme, aux contours mélodiques, à la prosodie affectant les mots de sa LM et enfin, aux indices « musicaux » correspondant aux frontières linguistiques et musicales majeures de sa LM et de son SMM. Ces étonnantes capacités de traitement se limitent-elles aux éléments musicaux ou concernent-elles également le niveau segmental ?

3.1.1. Jusqu'à 6 mois, l'enfant est un « superdiscriminateur »

Les sons de parole sont constitués par des gestes qui sont des actions coordonnées des mouvements du canal vocal. Best (1993) parle à leur sujet de gestes articulatoires (à la suite des travaux du physiologiste russe Bernstein). Il est impossible de reproduire le même son exactement de la même manière à deux occasions différentes.

Il subsiste toujours des différences acoustiques entre deux sons parlés de la même catégorie ; ces différences sont dues aux mouvements d'ajustement des différents articulateurs vocaux. Mais elles sont inutiles au locuteur pour identifier les segments phonétiques. Il a en effet besoin de se concentrer sur un type de distinctions efficaces. Son système perceptif va donc recoder le signal de manière à préserver les seules distinctions qui sont utiles pour distinguer les mots, c'est-à-dire les oppositions phonologiques. Ce recodage, très rapide et très efficace, a été baptisé **perception catégorielle** par les chercheurs des Laboratoires Haskins (Liberman et al., 1967). Il consiste à identifier les événements sonores selon leur appartenance à des catégories, mais en ignorant les variations acoustiques au sein d'une même catégorie. En conséquence, l'auditeur ne peut discriminer deux sons appartenant à la même catégorie perceptive ; il ne peut le faire que s'ils appartiennent à deux catégories distinctes.

Or, en matière de discrimination, le nourrisson fait preuve de capacités bien supérieures à celles de l'adulte. Il est en effet capable, non seulement de discriminer les contrastes de sa LM, mais également une grande variété de contrastes appartenant à des langues qui lui sont inconnues, et pour certaines d'entre elles, rares et très éloignées de la LM (langues kikuyu, hindi, ntlakapmx, zoulou...). Il existe d'ailleurs très peu d'études où les enfants ont échoué (pour un récapitulatif des contrastes discriminés, voir Volume II, annexe 1.3.), ce qui a conduit Trehub (1981) à qualifier le bébé de « superdiscriminateur ». Ces études ont porté en majorité sur les consonnes, mais on sait que les bébés sont capables même plus tôt de discriminer les voyelles (Trehub, 1973), même non natives, dès 1 mois (Polka et Werker, 1994). Ces capacités ont stupéfié les chercheurs, à tel point que certains ont pensé qu'elles ne pouvaient résulter que d'une prédisposition de l'enfant au traitement des langues naturelles (Eimas, 1974). En effet, comment expliquer autrement le fait que les nouveau-nés montrent des capacités de discrimination supérieures à celles des adultes ? Par ailleurs, si cette prédisposition pour le traitement de la langue était avérée, on pourrait également supposer l'existence d'un traitement spécialisé pour la musique, ce qui militerait en faveur d'une séparation très précoce du traitement de la langue et de la musique.

Une étude de Kuhl et Miller (1975) est venue contredire cette vision « innéiste » des choses en montrant que les chinchillas perçoivent le voisement des sons de parole comme les êtres humains. Des résultats similaires ont été observés chez des macaques rhésus (Kuhl et Padden, 1982) et des cailles japonaises qui avaient été entraînées au préalable à percevoir le lieu d'articulation (Kluender et al., 1987).

Ce mode de traitement n'est donc pas réservé aux seuls êtres humains, mais partagé par les espèce du groupe des mammifères. Les recherches entreprises vers la fin des années 70 (Cutting et Rosner, 1974 ; Pisoni, 1977) sur les stimuli non-verbaux amènent un autre type de découverte : la perception catégorielle ne serait pas seulement un mode de traitement du mode langagier, mais plutôt un trait général de la manière dont tous les signaux complexes sont traités. Elle a été observée notamment pour la vision des couleurs (Bornstein et al., 1976) et pour la musique (Juszyk et al., 1980). **La perception catégorielle est donc une capacité générale du système perceptif, valable entre autres pour la langue et la musique et commune à plusieurs espèces, parmi lesquelles figure l'être humain.**

3.1.2. Spécialisation progressive sur la LM aux dépens de sa richesse perceptive

Au contact de sa LM, l'attention de l'enfant va progressivement se fixer sur les représentations qui lui seront utiles pour son acquisition, augmentant sa performance et sa rapidité de traitement. Ce processus est nécessaire à l'acquisition de la LM, car il permet à l'enfant de se spécialiser sur les contrastes natifs à assimiler. Pour décrire ce phénomène de focalisation sur la LM, nous pourrions reprendre le terme d'acculturation, utilisé en psychologie sociale avec le sens de :

« processus d'apprentissage par lequel l'enfant reçoit la culture et l'ethnie ou le milieu auquel il appartient » (Fraisie, 1958 : 108-109).

En ce qui concerne la langue, l'acculturation désignerait le processus d'apprentissage par lequel l'enfant « recevrait » sa LM de son environnement. Mais en se spécialisant, l'enfant perd une partie de ses capacités de discrimination initiales. Ainsi, alors qu'il pouvait encore discriminer un grand nombre de contrastes, à partir de 6 mois, on voit sa perception se restreindre progressivement aux seules distinctions utiles dans sa LM. Cette période de transition correspond à une baisse de la sensibilité du bébé aux contrastes non-natifs et à une augmentation simultanée de sa sensibilité aux contrastes natifs.

3.1.2.1. Focalisation sur la structure vocalique de la LM à partir de 6 mois

Les voyelles, des sons musicaux

Vers l'âge de 6 mois, l'enfant connaît une réorganisation perceptive graduelle en faveur des voyelles de sa LM. Au niveau segmental, pourquoi son attention commence-t-elle par se fixer sur les voyelles ?

Il semblerait que ces sons jouent un rôle particulièrement important au moment de l'acquisition. Les voyelles occupent une place privilégiée au sein de la syllabe dont elles forment le noyau. Au niveau perceptif, elles sont facilement repérables puisqu'elles portent l'information prosodique. Elles sont en outre majoritairement composées d'intervalles de quintes justes, intervalles considérés comme prototypiques, c'est-à-dire faciles à traiter par le système perceptif (voir partie n°3.1.3.). Le rapport qui correspond à un intervalle de quinte est un rapport simple (2/3), fortement présent dans les composantes des sons complexes de l'environnement naturel. Selon Best (1993), les propriétés acoustiques des voyelles les rendent plus faciles à traiter que les consonnes. Elles constituent des points de repère faciles à coder et à mémoriser, des **prototypes**. L'idée que certains objets pourraient fonctionner comme des points de référence cognitifs n'est pas nouvelle en psycholinguistique (Rosch, 1975). On pourrait définir le prototype comme une **représentation idéale formée par l'expérience et qui possède des caractéristiques invariantes. Il constitue le meilleur exemplaire ou l'instance centrale d'une catégorie**. L'idée fondamentale, c'est que les catégories ne sont pas composées de membres aux statuts équivalents, mais qu'elles comportent des membres qui sont de meilleurs exemplaires que d'autres :

« La notion de prototype est donc originellement reliée aux individus : le prototype est l'exemplaire qui est reconnu comme étant le meilleur par les sujets. Cette origine psychologique pose directement le problème de la variation individuelle. Puisque étant d'abord la meilleure instance que les individus ont à l'esprit, le prototype pourrait en théorie varier d'un individu à un autre.[...] Le point définitoire nouveau est que le prototype n'est vraiment considéré comme le meilleur exemplaire d'une catégorie que s'il apparaît comme étant celui qui est le plus fréquemment donné comme tel » (Kleiber, 1990 : 48-49).

Les stimuli prototypiques sont ceux qui incluent le maximum de caractéristiques de la catégorie (Grieser et Kuhl, 1989). Le prototype fonctionne donc comme référence à l'assimilation des structures d'événements perceptifs similaires.

Dans sa théorie dite « *native language magnet* » (1992, 1993, 1994), Kuhl pose l'hypothèse que l'espace acoustique initial est divisé en frontières psycho-acoustiques correspondant aux caractéristiques phonémiques de la LM. Kuhl appelle cet espace « carte perceptive ». Nous pouvons en voir trois exemples sur la figure n°49. Celle de gauche correspond à la carte perceptive du suédois, celle du centre, à la carte perceptive de l'anglais et celle de droite, à la carte perceptive du japonais. Les frontières délimitent les différentes catégories constituant le système vocalique de la langue concernée.

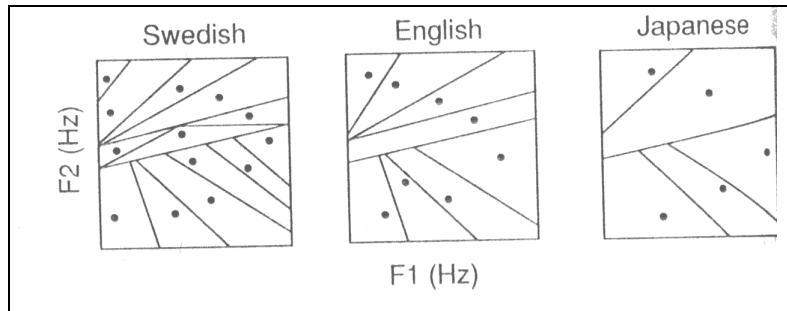


Figure n°49 : cartes perceptives du suédois (à gauche), de l'anglais (au centre) et du japonais (à droite) (emprunté à Kuhl, 1994 : 36).

Ces représentations internes, constamment générées par le cerveau, comparent leur propre structure aux formes qui se présentent dans l'environnement linguistique. L'enfant retient celles qui sont compatibles avec les contrastes, les structures syllabiques et les traits prosodiques de sa LM. Par exemple, le système japonais est composé de cinq voyelles, qui forment cinq catégories (cf. figure n°49). Le bon exemplaire de chaque voyelle (le prototype) se trouve au centre de ces cinq divisions ; sur les cartes, il est représenté par un point. Si un bébé japonais est confronté à une langue étrangère comme le suédois (13 catégories, cf. figure n°49), son système maternel va jouer le rôle d'un filtre. Le prototype fonctionne en effet comme un aimant pour les autres stimuli appartenant à la même catégorie. Lorsqu'un certain nombre de sons se trouvent autour d'un prototype dans une catégorie (figure n°50, au dessus), ils sont perceptivement attirés par ce prototype (« *perceptual magnet effect* », figure n°50, en bas). Les études sur les adultes suggèrent que ces limites ne disparaissent jamais réellement, même si on peut augmenter la performance de discrimination des contrastes NN par un apprentissage approprié (Flege 1989). Les catégories natives interfèrent donc avec la capacité à percevoir les distinctions phonétiques d'une nouvelle langue (cf. chapitre V, 1.2., apprentissage de l'anglais par des locuteurs francophones ; phénomène de crible phonologique).

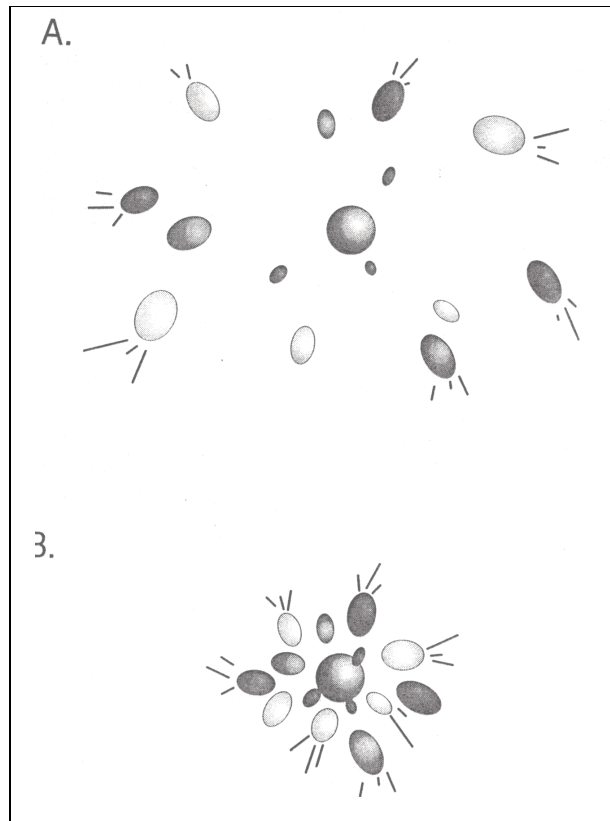


Figure n°50 : effet « magnétique » des prototypes de la LM (emprunté à Kuhl, 1994 : 124).

Kuhl et al. (1992) ont testé la discrimination de la voyelle [I] par des bébés américains et suédois afin de déterminer les effets potentiels des prototypes de cette voyelle en anglais et en suédois. Dans chacune de ces deux langues, le prototype de la voyelle n'est pas localisé au même endroit et la voyelle anglaise n'est pas perçue comme un prototype par un adulte suédois. Si la langue a déjà organisé à 6 mois une représentation de ces voyelles, il y aura une différence d'assimilation. Les résultats montrent qu'il existe bien une représentation de l'espace vocalique adaptée à la LM. En effet, les bébés américains se montrent sensibles uniquement à l'effet attractif de la voyelle anglaise et réciproquement pour les bébés suédois. Kuhl en conclut que « *by 6 months of age, exposure to the ambient language alters infants' perception of the phonetic units of language. This is the earliest age at which linguistic experience has been found to affect perception of the building blocks of language* » (Kuhl, 1994 : 126).

3.1.2.2. Focalisation sur la structure consonantique de la LM à partir de 10-12 mois

Après les voyelles, l'enfant se spécialise sur des constructions plus complexes que sont les consonnes de sa LM. Ces sons sont en effet produits plus rapidement que les voyelles et demandent plus de précision dans les gestes articulatoires. La focalisation sur les voyelles de la LM avait entraîné une perte de la sensibilité pour les voyelles non natives. Le même phénomène se produit vers l'âge de 10-12 mois pour les consonnes où l'attention de l'enfant se focalise sur les contrastes consonantiques de sa LM. Cette « spécialisation » entraîne une forte altération de ses capacités de discrimination pour les consonnes non natives, qu'il percevait pourtant bien à 2 mois, 4 mois et 6 mois. Cette **seconde réorganisation perceptive** le rapproche des capacités de discrimination d'un auditeur adulte pour le même type de contrastes. Elle serait liée à un traitement plus analytique qui permet à l'enfant, après un contact suffisant avec sa LM, de se spécialiser sur des constructions plus complexes et de les mémoriser. Ce type de traitement nécessite un traitement cognitif plus « lourd » de la part de l'enfant et ses capacités initiales à discriminer une grande variété de contrastes pourraient provenir d'un traitement global et donc, plus superficiel (cf. 1.2.2.).

L'étude menée par Werker et Tees (1984) constitue l'une des recherches les plus connues révélant les effets de cette réorganisation perceptive. Les auteurs ont testé les capacités de discrimination de bébés anglophones, à trois âges différents, 6-8 mois, 8-10 mois et 10-12 mois. Les contrastes non natifs testés appartenaient à deux langues très éloignées de la LM des bébés, l'hindi et le salish (langue thomson). Il est apparu que les bébés les plus jeunes discriminaient sans aucune difficulté les différents contrastes alors que les enfants les plus âgés échouaient dans cette tâche. Quant aux bébés de 8-10 mois, leurs résultats étaient mitigés. Cette étude révèle que l'acculturation à la LM est progressive. Best (1993) propose un modèle d'acquisition du langage, qui prend en compte cette réorganisation. Jusqu'à 6 mois, le bébé observe et perçoit les gestes articulatoires (« phonologie gestuelle »), sans être influencé par sa LM. A partir de 6 à 10 mois et jusqu'à 18 à 24 mois, il commence à reconnaître les gestes articulatoires natifs, mais pas encore leur valeur fonctionnelle. La transition s'effectue vers 10-12 mois pour les consonnes et de manière plus précoce (6 mois) pour les voyelles. C'est seulement à partir de 18 mois qu'il apprendra la fonctionnalité, la syntaxe et la morphologie (pour une description de son modèle d'assimilation perceptive, se reporter au chapitre V, partie 1.2.).

3.1.3. Spécialisation progressive sur le système musical

3.1.3.1. Définition de l'acculturation musicale

Si l'environnement linguistique modèle à ce point les capacités perceptives du bébé, il devrait se produire la même « acculturation » avec le SMM. Pour les petits occidentaux, ce SMM est celui de la musique tonale. Cela ne veut pas dire qu'ils ne peuvent pas être exposés à d'autres systèmes, mais que, dès la naissance, les auditeurs occidentaux sont majoritairement exposés à la musique tonale, notamment à cause de la très large diffusion de ce type de musique par les différents médias (musique de variétés notamment). Francès (1958 : 108-109) définit l'acculturation musicale comme

« un système d'habitudes perceptives qui s'enchaînent les unes aux autres, sans que les termes objectifs dont dépend leur acquisition soient consciemment posés par les sujets ».

Quelles sont ces habitudes perceptives et comment se mettent-elles en place ?

3.1.3.2. Bons patterns ou intégration des régularités de l'environnement ?

Les notes distantes d'une octave (rapport de fréquences 2 : 1) sont perçues comme équivalentes et portent d'ailleurs le même nom dans la plupart des cultures musicales. Ce principe, appelé **équivalence d'octave**, constitue une caractéristique musicale presque universelle selon Dowling et Harwood, (1986). Demany (1982) a montré que cette équivalence est perçue de façon très précoce, par des bébés de 2 mois et demi. Elle pourrait relever d'une contrainte de traitement du système perceptif. Dans la plupart des systèmes musicaux, l'octave est divisée en un nombre restreint de hauteurs distinctes (en général entre 5 et 7), qui composent les **gammes**. Dans la musique occidentale, la gamme la plus utilisée est la **gamme majeure** et les adultes sans formation musicale ont une connaissance tacite des différentes fonctions des notes de cette gamme (Krumhansl et Jusczyk, 1990 ; pour une description, voir chapitre I, 1.2.2.4.). La **triade majeure** occupe une place centrale au sein de la gamme majeure. Ainsi, Cohen et al. (1987) ont montré que, si les auditeurs adultes appliquent une stratégie de traitement par contours, habituelle pour une mélodie non familière, ils devraient échouer à détecter le changement d'intervalle si le contour est préservé (voir 1.2.2). Effectivement, les résultats montrent que les performances dans la tâche de discrimination sont beaucoup plus faibles pour les modifications affectant les intervalles, que pour les modifications affectant les contours.

Cependant, lorsque c'est la triade majeure qui sert de standard, les performances sont supérieures que pour la triade augmentée. Il est donc possible que la triade majeure soit un **bon pattern**, puisqu'elle permet aux auditeurs de dépasser leur stratégie habituelle de traitement par contours. Dans une autre étude, Trehub et al. (1990) ont montré que les mélodies occidentales conventionnelles, composées essentiellement à partir de triades majeures, facilitaient le traitement des mélodies chez des bébés de 6 mois. Les bébés seraient donc non seulement sensibles aux relations entre les notes à l'intérieur de la mélodie (hiérarchie intra-tonale, voir chapitre I, 1.2.2.4.), mais particulièrement sensibles à la triade majeure.

Les notes extrêmes de la triade sont en outre reliées par une **quinte juste** (do et sol, dans l'accord do-mi-sol en do majeur). Piston (1969) souligne le rôle central de cet intervalle, qui servirait de bonne forme autant pour les sons musicaux que pour les sons de la parole (voir ce chapitre, 3.1.2.1.). Cuddy et al. (1979) ont montré que les auditeurs adultes discriminent mieux les changements mélodiques lorsque la mélodie standard et la mélodie comparative sont reliées par un intervalle de quinte juste. Une autre étude révèle que la détection de changements d'un demi-ton dans la mélodie est plus aisée dans les mélodies composées à partir de gammes majeures, plutôt qu'à partir de gammes par tons, et ce, même pour des bébés âgés de 6 mois (Trehub, 1994).

Il reste cependant à déterminer si le statut particulier de la quinte juste, de la triade majeure et de la gamme majeure viennent du fait qu'elles soient des formes « intrinsèquement » bonnes, c'est-à-dire naturellement faciles à traiter ou du fait qu'elles apparaissent très souvent dans l'environnement musical du bébé. Pour Schenker (1906⁴), la triade majeure serait une forme consonante, stable et peut-être prototypique, alors que la triade augmentée serait dissonante et instable. En ce qui concerne la quinte juste, son rapport de fréquence simple (3 : 2) la rendrait particulièrement consonante (Burns et Ward, 1982 ; Rakowski, 1990). Kolinski (1967) relève que, dans la plupart des cultures, on trouve des adultes qui se mettent à entonner une chanson à la quinte ou à l'octave alors même qu'ils pensent être à l'unisson avec le groupe. Quant à la gamme majeure, elle posséderait, selon Balzano (1980), une certaine élégance structurelle. De là à dire que ces patterns sont « meilleurs » et plus « naturels » par rapport aux patterns d'autres cultures musicales, il y a un pas qu'il nous semble dangereux de franchir.

⁴ Cité par Trehub et Trainor (1994 : 324).

Il est d'autant plus dangereux que certains chercheurs comme Trehub par exemple (1994 : 336), considèrent les bébés de 6 mois comme des auditeurs naïfs de toute expérience avec leur SMM. Or la sensibilité du fœtus à la musique nous laisse entrevoir que l'acculturation est un long processus débutant bien avant la naissance (voir 2.1.3.). Le bébé de 6 mois est un auditeur déjà « aguerri », puisque bénéficiant d'une expérience longue de 10 mois avec son SMM. Ce point de vue oriente l'interprétation de leurs résultats et les amène à rejeter l'hypothèse d'une intégration progressive par l'enfant des régularités de son environnement. Pourtant, cette hypothèse n'entre pas en contradiction avec la présence de « bons patterns », qui faciliteraient le traitement de la musique (les fameuses « bonnes formes » de la Gestalt Theorie). Pour nuancer ce débat, il faudrait d'abord déterminer s'il se produit une acculturation en faveur du SMM, du même type que celle qui se produit pour la LM.

3.1.3.3. Evolution de la perception des gammes non natives

Il existe beaucoup moins de travaux sur l'acculturation musicale que sur l'acculturation avec la LM. Néanmoins, ces travaux suggèrent des différences de traitement entre les auditeurs adultes et les bébés. Alors que les bébés sont capables de détecter une note changée en dehors et à l'intérieur de la tonalité (Trainor et Trehub, 1992), les adultes détectent plus facilement ce changement en dehors de la tonalité (Cuddy et al., 1979 ; Trainor et Trehub, 1992). Ces résultats suggèrent que les bébés n'ont pas encore été assez exposés à leur SMM. Si la perception des bébés de 8 mois est différente de celles des adultes, il doit se produire à un moment donné une réorganisation perceptive en faveur du SMM.

Dans la musique tonale, l'intervalle de base est le demi-ton (plus petit dénominateur commun, rapport de 1.0595 ou 15/16). Tous les autres intervalles sont constitués à partir de lui et de ses multiples. Les gammes javanaises (appelées Pelog Scales) sont composées d'intervalles légèrement plus grands ou plus petits que les intervalles utilisés dans la musique tonale (cf. figure n°51). Ils ne sont pas fondés sur un rapport de fréquence constant et on ne peut pas les décomposer en unités stables. Ces intervalles possèdent donc une structure plus complexe, ce qui devrait les rendre plus difficiles à percevoir, même en faisant abstraction du poids de la culture musicale.

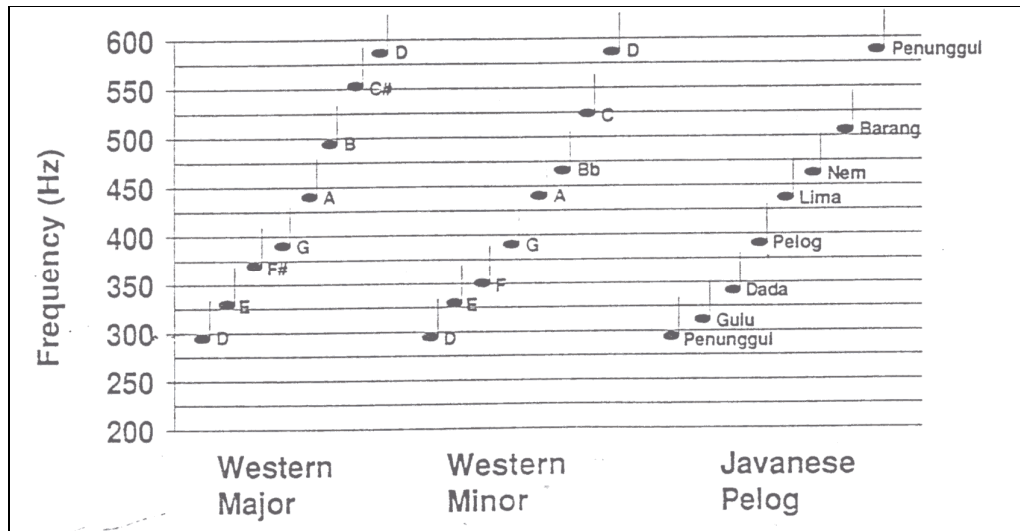


Figure n°51 : structuration des gammes majeures, mineures et javanaises (emprunté à Lynch et al., 1991 : 969).

Lynch et al. (1990) ont construit plusieurs mélodies de 7 notes, à partir de gammes majeures, de gammes mineures et de gammes « pelog ». La tâche consiste à discriminer la mélodie originale d'une mélodie dont la cinquième note est modifiée. Les résultats montrent que les bébés de 6 mois sont capables de discriminer les mélodies et leurs versions modifiées à partir des gammes majeures, mais également dans les mélodies construites à partir de gammes « pelog », ce qui n'est pas le cas chez les adultes. La performance similaire des bébés pour les gammes natives et non natives suggère que les gammes occidentales ne sont pas plus faciles à traiter que les gammes javanaises. En revanche, le poids d'une acculturation musicale est révélé par le fait que les adultes manifestent une meilleure performance pour les gammes natives que pour les gammes non natives. Selon Francès (1958), les auditeurs adultes ont intégré les intervalles spécifiques à leur culture. Ainsi, lorsque les auditeurs occidentaux écoutent une musique organisée selon une division différente de la gamme diatonique, comme c'est le cas des gammes javanaises par exemple, ils auront tendance, par un

« processus d'abstraction », à « réduire en notes une courbe vivante, et en intervalles inscrits sur une portée des intervalles qui en sont souvent fort écartés » (Francès, 1958 : 48-49⁵).

C'est le même phénomène que celui que nous avons appelé « crible tonal » dans le chapitre I (2.1.1.2.). Dans une autre étude de Lynch et Eilers (1992), les bébés de 6 mois se montrent meilleurs pour discriminer des changements intervenant dans des gammes majeures et gammes augmentées sans qu'il y ait beaucoup de différences entre les deux.

⁵ Cité par Zenatti (1994 : 330).

Ils ont également de bonnes performances pour les gammes pelog. En revanche, les performances des bébés de 10-12 mois sont mauvaises pour les gammes augmentées et les gammes pelog, alors qu'ils réussissent mieux avec les gammes majeures. Leurs résultats sont similaires à ceux des auditeurs adultes.

L'ensemble de ces résultats suggère que les bébés connaissent une période de réorganisation spécifiquement culturelle sur leur SMM. Cette réorganisation interviendrait entre l'âge de 6 et 12 mois. En effet, les bébés de 10-12 mois ne sont plus capables de discriminer les changements intervenant dans les gammes non natives, alors qu'ils étaient capables de le faire jusqu'à 6 mois. L'attention du bébé se fixe progressivement sur les schémas tonals spécifiques à son SMM. Ces schémas vont jouer le rôle de filtres en « modifiant » la perception des schémas musicaux non natifs en fonction de leur propre structure. Comme il existe un crible phonologique en faveur de la LM, il existe donc un « crible musical » en faveur du SMM (voir le récapitulatif sur la spécialisation progressive sur la LM et le SMM dans le Volume II, annexe 1.4.).

Alors que l'on peut remarquer les effets de l'acculturation linguistique car elle débouche sur l'acquisition de la LM, il n'en va pas de même pour l'acculturation musicale, dont les auditeurs n'ont souvent pas conscience. Pourtant, sans le savoir, ils ont intégré tous les principes qui leur permettent de structurer la perception de leur SMM. On pourrait donc dire que tous les auditeurs ont une « connaissance » musicale car ils ont une connaissance intuitive des principes qui gouvernent leur SMM, comme ils ont une connaissance intuitive de la grammaire de leur LM.

3.2. Production

Chez l'être humain, les vocalisations se développent dans deux domaines différents et hautement structurés, les productions musicales et les productions verbales. Cependant, pendant les premiers mois de la vie, il paraît très difficile de différencier ce qui relève de la parole de ce qui relève de la musique et plus spécialement du chant.

Selon Papousek (1995 : 123-124) :

« Il n'est pas interdit de penser qu'il est impossible de discriminer les précurseurs du chant spontané des précurseurs du parler précoce. La communication préverbale pourrait bien être une route ontogénétique commune, le long de laquelle deux capacités extrêmement structurées et exclusivement humaines se développent : (a) le langage conduisant à la communication verbale et à la pensée, et (b) le chant autorisant les activités créatrices de la musique vocale. Ces deux capacités sont aussi intimement liées aux fonctions affectives du signal vocal et de la communication ».

Langue et musique sont d'autant plus liées qu'elles sont produites avec le même « générateur » de sons, c'est-à-dire l'appareil vocal :

« à l'évidence, la dualité de la fonction du tractus vocal humain témoigne de la liaison entre langage et musique : organe de la parole, il est aussi instrument de musique. C'est pourquoi les travaux touchant à la petite enfance accordent de plus en plus d'attention à la précocité de son fonctionnement » (Papousek, 1995 : 49).

Pendant ses premiers mois de vie, le bébé va progressivement explorer toutes les possibilités que lui offre cet instrument. A cette période de liberté vocale et d'intense improvisation, succèdera une période où l'enfant devra plier sa virtuosité aux contraintes imposées par son environnement. Il pourra alors s'engager dans la production différenciée de sa LM et de son SMM.

3.2.1. Maîtrise progressive de l'appareil phonatoire

Pendant la première année, l'émergence du langage de l'enfant est aujourd'hui très bien décrite et nous nous référerons le plus souvent aux travaux de Konopczynski (1990, 1991, 1994). Au cours du premier mois, le bébé va apprendre à contrôler ses mouvements respiratoires et le fonctionnement de son larynx. Cette première « prise en main » de son appareil phonatoire passe par des vocalisations réflexes, des cris et des sons végétatifs (**phase de phonation** : contrôle de la phonation et de la respiration). Peu à peu, il va entrer dans la **phase de résonance** qui va s'étendre jusqu'à son quatrième mois. Au cours de cette période, il va produire des syllabes primitives, composées de sons quasi-vocaliques et de contoïdes articulés à l'arrière de la gorge (il se trouve la plupart du temps en position couchée). Selon Papousek (1995), les véritables précurseurs communs à la musique et à la langue se trouvent dans ces sons quasi résonants.

En ce qui concerne la mélodie, le bébé commence à produire des contours descendants dès la fin de la première semaine, ce qui pourrait résulter de la baisse de la pression de l'air sous-glottique. L'apparition des premiers contours ascendants vers la sixième semaine marque le début du contrôle des cordes vocales. Ces contours vont commencer à se complexifier dès la dixième semaine. Vers 3 mois commence une **phase exploratoire et d'expansion vocale**. L'enfant va jouer avec son conduit vocal comme un musicien le ferait avec son instrument, en testant toutes les combinaisons articulatoires possibles et les sons résultant de ces combinaisons, ainsi que les différents modes de phonation. Il se sert de l'ensemble de ses cavités phonatoires et fait pleinement résonner sa voix, son développement physiologique lui permettant d'exercer un contrôle respiratoire, phonatoire et articulatoire croissant. Pendant cette phase, il apprend à maîtriser les différents paramètres, ce qui lui procure un plaisir évident (Papousek et Papousek, 1981). Oller (1980) a relevé une augmentation de la durée des vocoïdes entre 3 et 5 mois, allant de 700 ms à 1400 ms. Vers 3-4 mois, tous les schémas mélodiques sont représentés dans la production de l'enfant, bien que 80 % d'entre eux soient descendants. La tessiture s'élargit, le bébé passant quelquefois brusquement du craquement vocal dans le registre des graves (90 Hz) aux sons sur-suraigus (2000 Hz) et du hurlement au murmure et inversement (Konopczynski, 1991, 1994). A partir du cinquième mois, sa fréquence fondamentale moyenne (Fom) devient très variable alors qu'elle était jusque là relativement stable. C'est pendant cette période que la richesse mélodique est la plus remarquable. Lorsqu'il est en interaction avec un adulte, le bébé essaie d'imiter les sons produits par son « interlocuteur ». Ainsi, dès le troisième mois, le bébé est capable d'imiter une mélodie ou les sons émis par l'adulte si ces sons appartiennent à son répertoire vocal. Les parents encouragent d'ailleurs ce type d'imitation vocale et la récompensent affectivement lorsqu'elle est réussie. Les tours de parole s'installent progressivement, donnant naissance à de petits dialogues vocaux entre l'adulte et l'enfant qui s'imitent mutuellement.

3.2.2. Le babillage canonique : l'entrée différenciée dans le langage et la musique

Vers 6 mois, apparaissent les premières combinaisons de sons contoïdes et vocoïdes avec fermeture du tractus vocal. Ces « proto-syllabes » se caractérisent par une articulation relativement lâche et des transitions lentes entre les sons. Oller (1980) parle de « **marginal babbling** », terme auquel nous préférons celui de babillage rudimentaire (Vinter, 1994). C'est ce type de babillage qui va permettre l'apparition du babillage canonique.

Parallèlement, la position du bébé se verticalise progressivement, entraînant un changement de configuration du tractus vocal vers 7-8 mois et une augmentation de volume de ces cavités supra-glottiques.

Entre 6-8 mois, le **babillage canonique** se met en place de manière relativement brusque. Il se caractérise par une alternance rythmique de mouvements de fermeture et d'ouverture de la bouche, qui permet la production de syllabes bien formées, de type CV. Il est très important de noter que le babillage canonique n'apparaît qu'en situation d'interaction avec un adulte et, lorsque le bébé est seul, dans sa chambre par exemple, il continue ses jeux vocaux, jeux qui se prolongeront bien après l'apparition des premiers mots. Alors que le jeu solitaire possède un caractère fortement aléatoire (arythmie, syllabes rudimentaires, vocoïdes très longs, instabilité de hauteur maximale), le babillage canonique se caractérise à l'inverse par une grande structuration. Selon Konopczynski (1990, 1991), le babillage canonique ouvre la voie du langage propre. Il constitue en effet une étape primordiale pour que l'enfant puisse accéder ensuite au langage référentiel (correspondant aux premiers mots et la mise en place de la syntaxe) ; c'est un véritable « protolangage » selon l'auteur. Papousek explique que son émergence va déclencher un changement de comportement chez les parents :

« dès leur émergence, les syllabes en canon sont considérées comme des protomots potentiels : les parents leur attribuent en effet des significations et en font une utilisation déclarative pour désigner des personnes, des objets et des événements du micro-environnement du bébé » (1995 : 52).

En effet, les premières syllabes du babillage canonique sont très souvent composées d'occlusives labiales [β,μ] et de voyelles centrales [α,A)] (Oller, 1980). Lorsque le bébé produit des combinaisons de syllabes du type [baba], [papa] ou [mama], elles sont immédiatement interprétées par les parents comme « papa » ou « maman » et ce d'autant plus spontanément que les parents attendaient ces « premiers mots » avec impatience. Le bébé va progressivement comprendre que ces séquences suscitent des réactions de joie de la part de ses parents et il est probable qu'il les reproduise pour leur faire plaisir. L'entrée dans la langue se fait donc dans l'interaction. Peu à peu, il va lui aussi leur attribuer un sens et c'est véritablement au moment où elles endosseront une signification pour le bébé qu'elles passeront du statut de « proto-mots » au statut de « mots » à part entière. Le lien entre la forme acoustique et le sens devient arbitraire, et c'est justement la nature de ce lien qui différencie le langage de la musique (voir chapitre I, 3.3.).

Il est possible que le bébé différencie déjà depuis longtemps la musique de la langue, en tout cas dans la parole des adultes (les situations d'émission et les attitudes sont très différentes), mais en revanche, pour l'adulte, il est peut être impossible de différencier clairement ce qui relève de la langue de ce qui relève de la musique dans les productions du bébé durant les premiers mois. D'ailleurs, avant le babillage canonique, l'adulte n'interprète pas les productions du bébé en termes linguistiques. Comme l'affirme Papousek (1995 : 53),

« au stade préverbal du bébé, il est extrêmement difficile de savoir si certaines interventions dues à l'intuition parentale encouragent plus particulièrement l'acquisition du langage, le jeu ou la musicalité. En revanche, lorsque la communication verbale s'installe, la spécificité de ces divers types de soutien s'affirme sous l'influence de facteurs culturels ».

Le babillage étant le véritable point d'entrée dans la communication verbale, il est probable que c'est à partir de son émergence que les parents vont pouvoir identifier facilement les productions linguistiques des productions musicales et en conséquence, apporter des stimulations différenciées à leur bébé. Cette différenciation va à son tour favoriser la séparation de la langue et de la musique dans les productions du bébé. Il est donc possible que l'entrée dans le SMM se fasse clairement à partir du babillage canonique.

3.2.3. Le langage émergent

Le protolangage contient déjà les caractéristiques rythmiques et intonatives de la LM, c'est-à-dire sa musique spécifique, ce qui le rend reconnaissable par des auditeurs de la même communauté linguistique (de Boysson-Bardies et al., 1984). Au niveau temporel, le bébé produit des syllabes ayant approximativement toutes la même durée. Selon Konopczynski (1990), cette durée (250 ms) est proche de la durée moyenne des syllabes adultes (200 ms) ; l'organisation de type isochronique est en place vers 8 mois et on la retrouve quelle que soit la langue de l'environnement. Mais, peu à peu elle va évoluer pour se rapprocher progressivement de la rythmique spécifique de la LM. Ainsi, pour un bébé français, l'organisation isochronique ne répond pas encore à la configuration rythmique du français, où la dernière syllabe du groupe rythmique est allongée (accentuation du français, cf. chapitre V, 2.1.1.1.). Cependant, très rapidement, les syllabes non finales vont se raccourcir, tandis que la durée des syllabes finales devient de plus en plus instable (trop courte ou trop longue, preuve d'un apprentissage par essai/erreur). L'allongement final qui caractérise le français se met progressivement en place pour se stabiliser entre 13 et 16 mois.

La durée des syllabes finales atteindra finalement le double de la durée des syllabes non finales, comme dans les productions adultes. Dans les langues où la place de l'accent est mobile comme l'anglais⁶, l'espagnol, le hongrois et le portugais, la rythmique est beaucoup plus difficile à acquérir.

Au niveau des paramètres mélodiques, le bébé contrôle sa voix en situation d'interaction (Fo moyen à 340 Hz), par rapport à la situation de jeu vocal. Les mélodies, jusque là riches et variées, se réduisent en nombre et leurs formes se stabilisent. Dès 9 mois, le bébé sait reproduire la configuration mélodique spécifique des énoncés interrogatifs, énonciatifs, exclamatifs, ainsi que celle des appels et des ordres. Ces productions sont correctement interprétées par les parents et même par des auditeurs extérieurs au cercle familial qui leur attribuent une fonction (phatique) ou une modalité précise (question, énonciation). Et c'est uniquement l'intonation qui permet cette différenciation (Konopczynski et Tessier, 1994), puisqu'il n'y a pas de couche verbale dans les énoncés du bébé. Ainsi, à la fin du dixième mois,

« un système premier est en place ; l'enfant dispose de quelques schèmes prosodiques, oppositifs, mais vides ; ces schèmes ou macro-segments forment un moule, un cadre prosodique qui va progressivement se remplir » (Konopczynski et Tessier, 1994 : 172).

Ces fameux « moules mélodiques » vont jouer un grand rôle dans la suite de l'acquisition, puisqu'ils vont recevoir le niveau segmental. Dans les stades suivants, l'enfant n'aura plus qu'à les remplir avec les premiers mots, au moment du babillage mixte (9-13 mois) et de la période du mot unique (12-18 mois) :

« au fur et à mesure qu'il s'approprie de nouvelles unités lexicales, il les introduit dans les moules prosodiques préexistants, sans – apparemment – rencontrer la moindre difficulté. Il connaît d'abord la musique de la phrase, il y met ensuite les paroles ; les éventuels trous ne le perturbent en aucune façon, ils sont comblés sans hésitation par du proto-langage » (Konopczynski et Tessier, 1994 : 174).

Plus tard, ces moules mélodiques formeront avec la rythmique spécifique de la LM (en place, rappelons le, dès le seizième mois chez le bébé français), la « structure d'accueil »⁷ de structures syntaxiques plus complexes (combinaison de 2 mots vers 18-20 mois).

⁶ Pour l'acquisition de la rythmique anglaise par le bébé, voir chapitre V, 2.1.3.4.

⁷ Terme emprunté à Konopczynski (1999 : 68).

3.2.4. Le chant émergent

Comparativement aux travaux sur l'acquisition du langage, il existe peu de travaux sur l'émergence des productions musicales et notamment sur l'imitation des mélodies et l'apparition des premiers chants spontanés. C'est d'autant plus dommage que les premières productions musicales et langagières sont indifférenciées durant les premiers mois et qu'il serait intéressant de mettre à jour le moment précis de leur divergence.

De ce point de vue, l'approche de Papousek et Papousek (1981) est particulièrement intéressante, car elle associe une approche linguistique à une approche musicologique. Ils ont établi un journal très détaillé des productions de leur fille Tanya de son premier à son seizième mois, en tentant d'identifier les précurseurs de la langue et les précurseurs de la musique. Ils ont ainsi relevé des capacités d'imitation musicale plus précoces que celles qui avaient été notées dans la littérature (7 mois pour Fridman, 1974⁸ et 14 mois pour Wing, 1941⁸) : dès 3 mois et demi, Tanya réussit à imiter des notes chantées ou jouées au piano. A 4 mois, elle essaie d'imiter des séquences plus longues et réussit à reproduire un pattern ressemblant au modèle adulte (même longueur, certains fragments mélodiques et rythmiques identiques). Ces capacités semblent toutefois communes à la langue, puisque Tanya est capable d'imiter les contours mélodiques de fragments de LAE dès 2 mois. En ce qui concerne ses productions spontanées, elles se caractérisent pendant les six premiers mois par des contours ondulants et continus. Les premières mélodies comportant des hauteurs discrètes apparaissent à 7 mois. La figure n°52 représente la transcription musicale d'une de ces mélodies. On peut y voir une suite de 8 tons de hauteur fixe, nettement délimités les uns des autres par des silences plus ou moins longs.

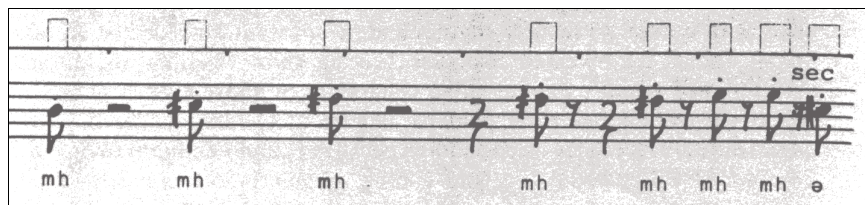


Figure n°52 : mélodie improvisée par Tanya à 7 mois (emprunté à Papousek et Papousek, 1981 : 190).

Selon Dowling (1994 : 172),

« les chansons sont distinctes de la parole en ce qu'elles évoluent à des niveaux discrets de hauteur, chaque voyelle reposant sur une hauteur plus ou moins stable ».

⁸ Cité par Papousek et Papousek (1981 : 193).

Nous prendrons le soin de préciser que cette définition s'applique à la musique tonale, où les hauteurs sont fixées et clairement délimitées les unes des autres (voir chapitre I, 1.2.2.). Les chansons enfantines sont particulièrement fidèles à cette configuration. Par rapport aux contours continus, la réalisation d'une telle séquence « staccato » nécessite une certaine maîtrise du tractus vocal et cette maîtrise coïncide avec l'apparition du babillage canonique (Papousek et Papousek, 1981 : 188).

C'est donc au moment du babillage canonique que les productions musicales deviennent clairement divergentes des productions linguistiques. Après l'apparition des premières mélodies, Papousek et Papousek réussissent clairement à distinguer les deux types de productions chez Tanya :

« in the further development of musical elements, two parallel lines become obvious, the speech-like intonation contours accompanying the linguistic differentiation, and the musical melodic contours in imitated, or invented songs » (1981 : 190).

Lewis (1936)⁹ considère le babillage comme une forme rudimentaire d'art et le début de l'usage esthétique du langage. Il est vrai que la rythmique du babillage canonique est très musicale en raison de sa régularité (écoulement régulier des syllabes à cause de leur isochronie). La mère pourrait la reprendre et lui associer une mélodie chantée, que l'enfant essaierait à son tour d'imiter. Lorsque le bébé est en situation d'interaction, il cherche en effet à

« s'harmoniser à la hauteur, à se mouvoir en synchronie avec la chanson ou à compléter la dernière note, à répéter certaines parties de l'air, voire à anticiper sur la phrase suivantes. De leur côté, les parents encouragent le chant commun en ralentissant le tempo ou en introduisant des silences afin de faciliter la participation du petit enfant par l'imitation ou la formation d'un duo joyeux » (Papousek, 1995 : 128).

L'entrée dans la musique, comme pour la langue, se fait donc dans l'interaction et *« le chant des parents est, semble-t-il, l'élément déterminant du sens musical du bébé »* (Papousek, 1995 : 129).

⁹ Cité par Papousek et Papousek (1981 : 212).

Pour en revenir à Tanya, à 10 mois, elle réussit à imiter des séquences tonales et de courtes mélodies et à 11 mois ½, elle est capable de répéter les premières phrases de chants complexes (mélodies et rythmes). A 15 mois ½, Tanya fredonne spontanément les bribes de ses chansons favorites. Les premières mélodies évoluent donc vers des chansons plus structurées, dans lesquelles l'enfant incorpore des fragments de chansons qui lui sont familières. Ainsi, selon Hargreaves (1995 : 188),

« les babillages vocaux de la phase sensori-motrice se meuvent en chansons articulées et reconnaissables. Celles-ci correspondent aux grandes lignes ou aux « cadres des chansons ». Ces vocalisations spontanées de l'enfant sont essentiellement caractérisées par une incorporation progressive de certains aspects des chansons de son environnement, laquelle donne naissance à des mixtures : Moog (1976) qualifie ce type de chanson de « pot-pourri » ».

En ce qui concerne ses productions spontanées, Tanya commence dès 16 mois à improviser de brefs airs, respectant les patrons mélodiques de base de la musique tonale. La figure n°53 nous donne un exemple de ce type d'improvisation. La tonalité de cet extrait est en la majeur (tonique : la ; médiate : do# ; dominante : mi) et la mélodie s'achève sur la tonique.

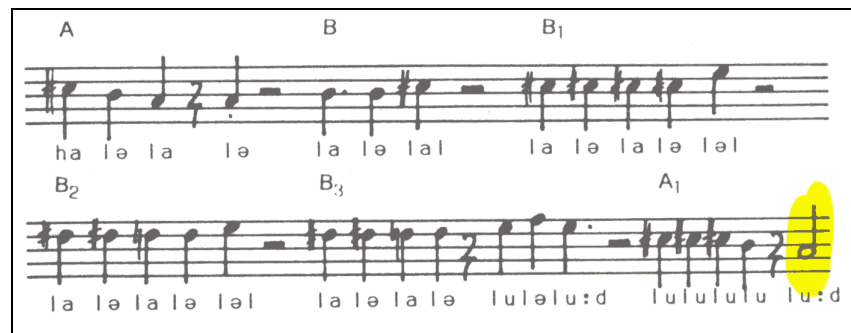


Figure n°53 : improvisation autour de la triade de la majeur (emprunté à Papousek et Papousek, 1981 : 215).

De manière plus précoce, Tanya avait déjà montré des préférences pour les notes de la triade majeure dans ses productions spontanées. Ainsi, Papousek et Papousek (1981) ont relevé dans un monologue ininterrompu produit à 11 mois, que 50 % des 130 tons étaient associés à la gamme de fa majeur. A 14 mois, cette proportion s'élève à 70 % des 138 tons relevés, mais cette fois, pour la gamme de ré majeur. Cependant, il ne faut pas perdre de vue que ce bébé a grandi dans un milieu musicien et qu'il a reçu beaucoup de stimulations de son environnement. Il serait donc particulièrement intéressant de réaliser des études qualitatives sur un plus grand nombre d'enfants, provenant de systèmes musicaux différents (musique tonale / gamelans balinaï par exemple).

Ces observations s'accordent avec les données recueillies pour la perception. En effet, nous avons vu que le bébé connaissait une période de réorganisation perceptive sur son SMM entre 10 et 12 mois. A partir de cet âge, les petits occidentaux échouent à détecter des modifications intervenant dans une autre gamme que la gamme majeure. Ce processus de centration sur les structures musicales maternelles semble très efficace, puisque simultanément dans le cas de Tanya, on peut relever une préférence dans les productions spontanées pour les notes de la triade majeure. En revanche, le poids de la LM semble influencer les capacités perceptives de manière plus précoce que pour la musique, en raison peut-être d'une exposition plus massive à la LM qu'au SMM. Cependant, à l'instar de la musique, les productions du bébé semblent immédiatement refléter cette spécialisation. En effet, la date d'apparition du babillage canonique correspond approximativement à la période où l'attention du bébé se resserre sur les voyelles de sa LM.

Alors que, dans les premières productions du bébé, il était très difficile de différencier la langue de la musique, ses productions sont clairement influencées par sa LM et son SMM à la fin de la première année. A partir de cette différenciation, qui trouve sans doute son origine dans le babillage canonique et qui est renforcée dans les interactions avec l'entourage, les productions musicales et linguistiques vont diverger de manière radicale, comme nous allons le constater dans le chapitre suivant.

Bibliographie chapitre II

- Armitage, S.E., Baldwin, B.A. et Vince, M.A. (1980). The foetal sound environment of sheep. In *Science*, 208, 1173-1174.
- Balzano, G. J. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. In *Computer Music Journal*, 4, 66-84.
- Bartlett, J. C. et Dowling, W. J. (1980). The recognition of transposed melodies: a key-distance effect. In *Journal of Experimental Psychology: Human Perception and Performance*, 6, 501-515.
- Benedict, H. (1979). Early lexical development: comprehension and production. In *Journal of Child Language*, 6, 183-200.
- Bergeson, Tonya R. et Trehub, Sandra E. (1999). Mothers' singing to infant and preschool children. In *Infant Behavior and Development*, 22, 1, 51-64.
- Bernstein Ratner, Nan (1986). Durational cues which mark clause boundaries in mother-child speech. In *Journal of Phonetics*, 14/2, 303-309.
- Bertoncini, Josiane (1993). Initial capacities for speech processing: infants' attention to prosodic cues for segmentation. In De Boysson-Bardies, De Schonen, Jusczyk, MacNeilage and Morton, 249-257.
- Bertoncini, Josiane et de Boysson-Bardies, Bénédicte (2000). La perception et la production de la parole avant deux ans. Dans Kail et Fayol, 95-136.
- Best, Catherine T., McRoberts, Gerald W. and Sithole, Nomathemb N. (1988). The phonological basis of perceptual loss of nonnative contrasts: maintenance of discrimination among Zulu clicks by english-speaking adults and infants. In *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.
- Best, Catherine T. (1993). Learning to perceive the sound pattern of English. In Rovee-Collier and Lipsitt.
- Best, Catherine T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In Goodman and Nusbaum, 167-224.
- Bigand, Emmanuel (1994). Contributions de la musique aux recherches sur la cognition auditive humaine. Dans Mc Adams et Bigand, 249-298.
- Bolinger, Dwight L. (1978). Intonation across languages. In Greeberg, Ferguson, and Moravcik, 2, 471-524.
- Bornstein, Marc H., Kessen, W. and Weiskopf, S. (1976). Color vision and hue categorisation in young human infants. In *Journal of Experimental Psychology: Human Perception and Performance*, 2, 115-129.
- Boysson-Bardies, Bénédicte de (1993). Ontogeny of language-specific syllabic productions. In Boysson-Bardies, Schonen de, Jusczyk, Mac Neilage and Morton, 353-363.
- Boysson-Bardies, Bénédicte de, Sagart, Laurent and Durand, Catherine (1984). Discernible differences in the babbling of infant according to target language. In *Journal of Child Language*, 11, 1-15.
- Boysson-Bardies, Bénédicte de, Hallé, Pierre A., Sagart, Laurent and Durand, Catherine (1989). A cross-linguistic investigation of vowel formants in babbling. In *Journal of Child Language*, 16, 1-17.
- Boysson-Bardies, Bénédicte de, Schonen, Scania, Jusczyk, Peter, MacNeilage and Morton (1993). *Developmental Neurocognition: speech and face processing in the first year of life*. Boston: Kluwer Academic Publishers.
- Bregman, Albert S. (1978). Auditory streaming is cumulative. In *Journal of Experimental Psychology: Human Perception and Performance*, 4, 380-387.
- Bregman, Albert S. (1981). Asking the "what for" question in auditory perception. In Kubovy and Pomerantz, 99-118.
- Bregman, Albert S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge : MIT Press.
- Bregman, Albert S. (1994). L'analyse des scènes auditives. Dans McAdams et Bigand, 11-39.
- Broen, P. A. (1972). The verbal environment of the language-learning child. In *American Speech and Hearing Association Monographs*, 17 (décembre).

- Burns, E. M. and Ward, W. D. (1982). Intervals, scales and tunings. In *Deutsch*, 241-269.
- Busnel, Marie-Claire (1994). Is there prenatal culture ? In Gardner, Chiarelli, Gardner and Plooij, 285-314.
- Busnel, Marie-Claire, Daffos, Fernand, Dolto-Tolitch, Lecanuet, Jean-Pierre et Negri, Romana (1997). *Que savent les Fœtus ?* Ramonville Saint-Agne : Erès, 103 p.
- Busnel, Marie-Claire (1997). Audition fœtale et réactivité prénatale à la voix maternelle “ adressée ”. In Busnel et al, 35-49.
- Caterette, E. C. and Friedman, M. P. (1976). *Handbook of Perception*, VII. New-York: Academic Press.
- Chapman, J. S. (1975). *The Relation between Auditory Stimulation of Short Gestation Infants and Their Gross Motor Limb Activity*. PHD, New-York University.
- Christophe, Anne, Dupoux Emmanuel, Bertoncini, Josianne and Mehler, Anne (1994). Do infants perceive word boundaries ? An empirical study of the bootstrapping of lexical acquisition. In *Journal of the Acoustical Society of America*, 95, 1570-1580.
- Cohen, Anthony J., Thorpe, Leigh A. and Trehub, Sandra, S. E. (1987). Infants’ perception of musical relations in short transposed tone sequences. In *Canadian Journal of Psychology*, 41, 33-47.
- Cohen, Antonie and Nooteboom, Sieb G. (1975). *Structure and Process in Speech Perception*. Berlin: Springer, 351 p.
- Cooper, W. E. and Paccia-Cooper (1980). *Syntax and Speech*. Cambridge: Harvard University Press.
- Cooper, R. P., & Aslin, Richard N. (1990). Preference for infant-directed speech in the first month after birth. In *Child Development*, 66, 1584-1595.
- Cruttenden, Alan (1986). *Intonation*. Cambridge: Cambridge University Press, 214 p.
- Cuddy, L. L., Cohen, Anthony J. and Miller, Joanne (1979). Melody recognition: The experimental application of musical rules. In *Canadian Journal of Psychology*, 33, 148-157.
- Cutting, James E. and Rosner, B. S. (1974). Categories and boundaries in speech and music. In *Perception and Psychoacoustics*, 16, 564-570.
- DeCasper, A. J. and Fifer, W. P. (1980). Of human bonding : Newborns prefer their mothers’ voices. In *Science*, 208, 1174-1176.
- DeCasper, A. J. and Prescott, P. A. (1984). Human newborns’ perception of male voices: preference, discrimination, and reinforcing value. In *Developmental Psychobiology*, 17, 481-491.
- DeCasper, A. J. and Spence, M. J. (1986). Prematernal speech influence newborns’ perception of speech sounds. In *Infant Behavior and Development*, 9, 133-150.
- DeCasper, A.J., Lecanuet, Jean-Pierre, Busnel, Marie-Claire, Granier-Deferre, C. et Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. In *Infant Behavior and Development*, 17, 159-164.
- Delcloque, Philippe (1999). *Speech Technology Applications in CALL*. Dundee: University of Albertay Dundee, 87 p.
- Deliège, Irène et Sloboda, John, A. (1995). *Naissance et Développement du Sens Musical*. Paris : PUF, 307 p.
- Demany, Laurent, McKenzie, B. and Vurpillot, Eliane (1977). Rhythm perception in early infancy. In *Journal of Acoustic Society of America*, 76, 57-66.
- Demany, Laurent (1982). Auditory stream segregation in infancy. In *Infant Behavior and Development*, 5, 261-276.
- Deutsch, D. (1969). Music recognition. In *Psychological Review*, 76, 300-307.
- Deutsch, D. (1982). *The Psychology of Music*. New-York: Academic Press.
- Dowling, Jay W. (1978). Scale and contour: two components of a theory for melodies. In *Psychological Review*, 85, 341-354.
- Dowling, Jay W. (1994). La structuration mélodique. Dans *Zenatti*, 145-176.
- Dowling, Jay W. and Harwood, D. L. (1986). *Music Cognition*. Orlando: Academic Press.
- Eimas, Peter D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. In *Perception and Psychophysics*, 16, 513-521.

- Eimas, Peter D. (1975). Auditory and phonetics coding of the cues for speech: discrimination of the r-l distinction by young infant. In *Perception and Psychophysics*, 18, 341-347.
- Fassbender, Christoph (1993). *Auditory grouping and segregation processes in infancy*. Norderstedt: Kaste Verlag.
- Fassbender, Christoph (1995). La sensibilité auditive du nourrisson aux paramètres acoustiques du langage et de la musique. In Delière et Sloboda, 63-99.
- Feagans, C. Garvey, R. Golinkoff, M. T. Greenberg, C. Harding, & J. Bohannon (1984). *The Origins and Growth of Communication*. Norwood: New Jersey, Ablex Publishing Corporation.
- Feijoo, Jean (1975). Ut conscientia Noscatue. Dans *Cahiers de Soprologie*, 13, 14-20.
- Feijoo, Jean (1983). Le Fœtus, Pierre et le loup. In Herbinet, Etienne et Busnel, Marie-Claire, 193-206.
- Ferguson, Charles A. (1977). Baby talk as a simplified register. In Snow and Ferguson, 219-235.
- Fernald, Anne (1976). The mother's speech to the newborn. Paper presented at the Max-Planck Institute for Psychiatry, Munich (non published).
- Fernald, Anne (1984). The perceptual and affective salience of mothers' speech to infants. In Feagans, Garvey, Golinkoff, Greenberg, Harding & Bohannon, 5-29.
- Fernald, Anne (1985). Four-month-old infants prefer to listen to motherese. In *Infant Behavior and Development*, 8, 181-195.
- Fernald, Anne (1989). Intonation and communication intent in mother's speech to infants: is the melody the message ? In *Child Development*, 60, 1497-1510.
- Fernald, Anne and Kuhl, Patricia (1981). Fundamental frequency as an acoustic determinant of infant preference for " Motherese ". Boston: Paper presented at the Biennial Meeting of the Society for Research in Child Development.
- Fernald, Anne and Simon, Thomas (1984). Expanded intonation contours in mothers' speech to newborns. In *Developmental Psychology*, 20/1, 104-113.
- Fernald, Anne and Kuhl, Patricia (1987). Acoustic determinants of infant preference for motherese speech. In *Infant Behavior and Development*, 10, 279-293.
- Fernald, Anne and Mazzie, C. (1991). Prosody and focus in speech to infants mothers' speech to new-borns. In *Developmental Psychology*, 27, 209-221.
- Fisher, Cynthia L. and Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: cross-linguistic evidence. In *Child Development*, 67, 3192-3218.
- Flege, James E. (1989). Using visual information to train foreign-language vowel production. In *Language Learning*, 38, 365-407.
- Francès, Robert (1958). *La Perception de la Musique*. Paris : Vrin, 408 p.
- Fridman, R. (1974). Los comienzos de la conducta musical. In *Biblioteca de Psicología Evolutiva Serie 2* (Vol. 13). Buenos Aires: Paidós.
- Gardner, R.A., Chiarelli, B.T., Gardner, B.T. and Plooi, F.X. (1994). *The Ethological Roots of Culture*. Kluwer Acad. Publish. Dordrecht Netherlands.
- Gathercole, Susan E. and Baddeley, Alan D. (1993). *Working memory and language*. Hove: Lawrence Erlbaum Associates.
- Gentil-Neymon, V. (1993). *Influence de la Voix de la Mère auprès des Prématurés en Service de Soins Intensifs*. Thèse de Doctorat de Médecine, Université Paris VII.
- Giacomi, A. et De Heredia, C. (1986). Réussites et échecs dans les communications linguistiques entre locuteurs francophones et locuteurs immigrés. In *Language*, 84, décembre.
- Gleitman, Lila R. and Wanner, E. (1982). *Language Acquisition: the State of the Art*. Cambridge. Cambridge: Cambridge University Press.
- Gleitman, Lila R., Gleitman, Harry, Landau, Barbara and Wanner, E. (1988). Where the learning begins: initial representations for language learning. In Newmeyer.
- Goodman, Judith C. and Nusbaum, Howard C. (1994). *The Development of Speech Perception: the Transition from Speech Sounds to Spoken Words*. Cambridge: MIT press.

- Greenberg, J.H., Ferguson, Charles A. and Moravcik, E. A (1978). *Universals of Human Language, Phonology*, 2. Standford: Stanford University Press.
- Grieser and Kuhl, Patricia K. (1989). Categorization of speech by infants: support for speech-sound prototypes. In *Developmental Psychology*, 25, 577-588.
- Halford, Brigitte K. and Pilch, Herbert (1994). *Intonation*. Tübingen: Gunter Narr Verlag.
- Hallé, Pierre A. and De Boysson-Bardies, Bénédicte de (1994). Emergence of an early receptive lexicon: infants' recognition of words. In *Infant Behavior and Development*, 17, 119-129.
- Hargreaves, David (1995). Développement du sens artistique et musical. Dans Deliège et Sloboda, 167-197.
- Hepper, P. G. (1991). An examination of fetal learning before and after birth. In *Irish Journal of Psychology*, 12, 95-107.
- Hepper, P. G., Scott, D. and Shahidullah, S. (1993). Newborn and fetal response to maternal voice. In *Journal of Reproductive and Infant Psychology*, 11, 147-153.
- Hepper and Shaidullah, (1994). Developmental of fetal hearing. In *Arch Dis Childhood*, 71, F81-F87.
- Herbinet, Etienne et Busnel, Marie-Claire (1983). *L'Aube des Sens*. Paris : Stock, 414 p.
- Hirsh-Pasek, Kathryn, Kemler, Nelson, Deborah, Jusczyk, Peter W., Wright, Cassidy, K., Druss, B. and Kennedy, L. (1987). Clauses are perceptual units for young infants. In *Cognition*, 26, 269-286.
- Hulse, S. H., Humpal, J. and Cynx, J. (1984). Discrimination and generalization of rhythmic and arhythmic sound patterns by European starlings. In *Music Perception*, 1, 442-464.
- Journiac, Claire (2002). *La Place de l'Autre dans la Relation Soignante : Contribution à l'Etude des Processus Relationnels à Travers la Voix*. Université de Bourgogne, 360p.
- Jusczyk, Peter W. (1989). Perception of cues to causal units in native and non-native languages. Paper presented at the biennial meeting of the *Society for Research in Child Development*, Kansas City, Avril.
- Jusczyk, Peter W. (1997). *The Discovery of Spoken Language*. Cambridge : MIT, 314 p.
- Jusczyk, Peter W. and Thompson, E. J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. In *Perception and Psychophysics*, 23, 105-109.
- Jusczyk, Peter W. and Krumhansl, Carol (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. In *Journal of Experimental Psychology: Human Perception and Performance*, 19/3, 627-640.
- Jusczyk, Peter W., Hirsh-Pasek, Kathryn, Kemler Nelson, D. G., Kennedy, L., Woodward, A. and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. In *Cognitive Psychology*, 24, 252-293.
- Jusczyk, Peter W., Friederici, A. D., Wessels, J., Svenkerud, V. Y. and Jusczyk, A. M. (1993a). Infants' sensitivity to the sound patterns of native language words. In *Journal of Memory and Language*, 32, 402-420.
- Kagan, Jerome and Lewis, Michael (1965). Studies of attention in the human infant. In *Merill-Palmer Quarterly*, 11, 95-122.
- Kail, Michèle et Fayol, Michel (2000). *L'Acquisition du Langage : le Langage en Emergence de la Naissance à Trois ans*. Vol. I. Paris : PUF, 304 p.
- Katz, G. S., Cohn, J. F. and Moore, F. A. (1996). A combination of vocal fo dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. In *Child Development*, 67, 205-217.
- Kavanagh and Cutting, James E. (1975). *The Role of Speech in Language*. Cambridge: MIT press.
- Kemler Nelson, D. G., Hirsh-Pasek, Kathryn, Jusczyk, Peter W. and Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. In *Journal of Child Language*, 16, 53-68.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. In *Psychological Bulletin*, 112, 24-38.
- Klatt, Dennis K. and Cooper, William E. (1975). Perception if segment duration in sentence contexts. In Cohen and Nooteboom, 69-86.
- Kleiber, Michel (1990). *La Sémantique du Prototype : Catégories et Sens Lexical*. Paris : P.U.F., 199 p.

- Kluender, K. R., Diehl, R. L. and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. In *Science*, 237, 1195-1197.
- Kolinski, M. (1967). Recent trends in ethnomusicology. In *Ethnomusicology*, 11, 1-24.
- Konopczynski, Gabrielle (1990). *Le Langage Emergent : Caractéristiques Rythmiques*. Hamburg : Buske Verlag, 362 p.
- Konopczynski, Gabrielle (1991). *Le Langage Emergent II : Aspects Vocaux et Mélodiques*. Hambourg : Buske Verlag, 425 p.
- Konopczynski, Gabrielle (1999). L'acquisition du système prosodique de la langue maternelle et ses implications pour l'apprentissage d'une L2. Dans Delcloque Philippe, 62-70.
- Konopczynski, Gabrielle et Tessier, Sophie (1994). Structuration intonative du langage émergent. Dans Halford et Pilch, 157-192.
- Krumhansl, Carol L. and Jusczyk, Peter, W. (1990). Infant's perception of phrase structure in music. In *Psychological Science*, 1, 70-73.
- Kubovy and Pomerantz (1981). *Perceptual Organization*. Hillsdale: Erlbaum Assoc.
- Kuhl, Patricia K. (1992). Infants' perception and representation of speech: Development of a new theory. In Ohala, Nearey, Derwing, Hodge and Wiebe, 449-456.
- Kuhl, Patricia K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In Boysson-Bardies de, Schonen, Jusczyk, MacNeilage and Morton, 259-274.
- Kuhl, Patricia K. (1994). Speech perception. In Minifie, 118-142.
- Kuhl, Patricia K. and Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. In *Science*, 190, 69-72.
- Kuhl, Patricia K. and Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. In *Perception and Psychophysics*, 31, 279-292.
- Kuhl, Patricia K. and Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for voicing feature in macaques. In *Perception and Psychophysics*, 32, 542-550.
- Kuhl, Patricia K., Williams, K. A., Lacerda, F., Stevens, K. N. and Lindblom, Björn (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. In *Science*, 255, 606-608.
- Kuhl, Patricia K., Andruski, J.E., Chistovitch, I.A., Chistovitch, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U. and Lacerda, F. (1997). Cross-language analysis of phonetic in language addressed to infants. In *Science*, 277, 684-686.
- Lecanuet, Jean-Pierre, Granier-Deferre, C., Jacquet, A. Y., Capponi, I. et Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. In *Early Development and Parenting*, 2, 217-228.
- Lecanuet, Jean-Pierre (1995). L'expérience auditive prénatale. In Deliège et Sloboda, 7-38.
- Lecanuet, Jean-Pierre (1997). Dans tous les sens... bref état des compétences sensorielles fœtales. In Busnel et al., 17-34.
- Lecanuet, Jean-Pierre, Jacquet, A-Y., Hains S.M.J. et Kisilevsky, Barbara S. (2000). Effects of music on behaviour in term fetuses. *International Conference on Infant Studies* ; Brighton , Grande Bretagne, Juillet 3-6.
- Lecanuet, Jean-Pierre, Jacquet, Anne-Yvonne, Kisilevsky, Barbara, S. and Hains, Sylvia, M. J. (2000). Effects of music on behaviour in term fetuses. Poster, ICIS 2000 (International Conference on Infant Studies), Brighton.
- LeCours, A. R. (1975). Myelogenesis correlates of the development of speech and language. In Lenneberg and Lenneberg.
- Lenneberg, E. H. and Lenneberg, E. (1975). *Foundations of Language Development of Speech and Language*. New-York: Academic Press.
- Levarie, S. and Rudolph, N. (1978). Can newborn infants distinguish between tone and noise ? In *Perceptual and Motor Skills*, 47, 1123-1126.
- Lewis, Michael (1936). *Infant Speech: A Study of the Beginnings of Language*. New-York: Harcourt Brace.
- Liberman, Anatoly M., Cooper, F. S., Shankweiler, D. P. and Studdert-Kennedy, Michael G. (1967). Perception of the speech code. In *Psychological Review*, 74, 431-461.

- Lindblom, Björn, McNeillage and Studdert-Kennedy, Michael G. (1984).
- Lipsitt, Lewis P. and Rovee-Collier, Carolyn K. (1981). *Advances in Infancy Research*. Vol I : XXII. Norwood: Ablex, 267 p.
- Luce, Paul A. and Charles-Luce, Jan (1983). Contextual effects on the consonant/vowel ratio in speech production. Paper presented at the *Meeting of the Acoustical Society of America*, Cincinnati, May.
- Lynch, Michael P. (1993). Prototypical representations of music structure in infancy: theoretical exploration and a pilot study. In *Psychomusicology*, 12, 31-40.
- Lynch, Michael P., Eilers, Rebecca E., Oller, Kimbrough D. and Urbano, Richard C. (1990). Innateness experience and music perception. In *American Psychological Society*, I/4, 272-276.
- Lynch, Michael P. and Eilers, Rebecca, E. (1991). Children's perception of native and nonnative musical scales. In *Music Perception*, 9/1, 121-132.
- Lynch, Michael P., Eilers, Rebecca E., Oller, Kimbrough D. and Urbano, Richard C. (1991). Influences of acculturation and musical sophistication on perception of musical interval patterns. In *Journal of Experimental Psychology: Human Perception and Performance*, 17/4, 967-975.
- Lynch, Michael P. and Eilers, Rebecca, E. (1992). A study of perceptual development for musical tunings. In *Perception and Psychophysics*, 52/6, 599-608.
- McAdams, Stephen et Bigand, Emmanuel (1994). *Penser les Sons : Psychologie Cognitive de l'Audition*. Paris : P.U.F., 402 p.
- Malsheen, Bathsheb (1980). Two hypotheses for phonetic clarification in the speech of mothers-to-children. In Yeni-Komshian, Kavanagh and Ferguson, II, 173-184.
- Martin, Pierre (1996). *Éléments de Phonétique avec Application au Français*. Saint-Nicolas (Québec) : Presses Universitaires de l'Université de Laval, 253 p.
- Mehler, Jacques, Bertoncini, Josiane, Barrière, M., Jassik-Gershenfeld, Dora (1978). Infant recognition of Mother's voice. In *Perception*, 7, 491-497.
- Menyuk, Paula (1977). *Language and Maturation*. Cambridge: MIT Press.
- Menyuk, Paula (1994). Word acquisition: the important role of prosody. In *The Study of Sounds*, 23, 218-222.
- Miller, Joanne L. and Liberman, Anatoly M. (1979). Some effects of later occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25: 457-465.
- Minifie, Fred H. (1994). *Introduction to Communication Sciences and Disorders*. San Diego: Singular Publishing Group.
- Moog, H. (1976). *The Musical Experience of the Pre-school Child*. London: Scott.
- Moon, C., Cooper, R., and Fifer, W. P. (1993). Two-day-olds prefer their native language. In *Infant Behavior and Development*, 16, 495-500.
- Morgan, James L., Meier, R. P. and Newport, Elissa L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language ? In *Cognitive Psychology*, 19, 498-550.
- Morse, Philip A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. In *Journal of Experimental Child Psychology*, 13, 477-492.
- Mosser, Christine (1990). *Effets physiologiques des Stimulations Sonores chez les Prématurés*. Thèse de Doctorat : Mondrini, Paris XII.
- Navon, D. (1977). Forest before trees: the precedence of global feature in visual perception. In *Cognitive Psychology*, 9, 353-383.
- Nazzi, Thierry, Kemler Nelson, D., Jusczyk, Peter and Jusczyk, A.M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. In *Infancy*, 1/1, 123-147.
- Neisser, U. (1967). *Cognitive Psychology*. New-York: Appleton Century Crofts.
- Newmeyer, Frederick J. (1988). *Linguistics: The Cambridge Survey*. Cambridge: Cambridge University Press, III, 349 p.

- Nowik-Stern, Alisa, Marsha G. Clarkson, Mary K. Morris, Roger Bakeman (1998). The effect of premature infants' speech preferences on mother-preterm interaction. Poster, *11th Biennial International Conference on Infant Studies*, Atlanta, Georgia, April 2-5.
- Ohala, J. J., Nearey, T. M., Derwing, B. L., Hodge, M. M. and Wiebe, G. E. *Proceedings of the International Conference on Spoken Language Processing*. Edmonton, Alberta: University of Alberta.
- Oller, Kimbrough D. (1980). The emergence of the sounds of speech in infancy. In Yeni-Komshian, Kavanagh and Ferguson, I, 93-112.
- Panneton, R. K. (1985). *Prenatal auditory experience with melodies: Effects on postnatal auditory preferences in human newborns*. PhD, University of North Carolina at Greensboro.
- Papousek, Mechthild and Papousek, Hanus (1981). Musical elements in the infants vocalizations: Their significance for communication, cognition and creativity. In Lipsitt and Rovee-Collier, 163-224.
- Papousek, Mechthild, Bornstein, Marc H., Nuzzo, C., Papousek, Hanus and Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. In *Infant Behavior and Development*, 13, 539-545.
- Papousek, Mechthild, Papousek, Hanus, & Symmes, D. (1991). The meaning of melodies in motherese in tone and stress languages. In *Infant Behavior and Development*, 14, 415-440.
- Papousek, Hanus (1995). Musicalité et petite enfance. Origines biologiques et culturelles de la précocité. In Deliège et Sloboda, 41-62.
- Papousek, Mechthild (1995). Le comportement parental intuitif, source cachée de la stimulation musicale dans la petite enfance. In Deliège et Sloboda, 101-130.
- Peters, Ann M (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Pisoni, David B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. In *Journal of the Acoustical Society of America*, 61, 1352-1361.
- Piston (1969). *Harmony*. New-York: Norton.
- Polka, M. and Werker, Janet F. (1994). Developmental changes in perception of nonnative vowel contrasts. In *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.
- Proverbio, Alice Mado, Minniti, Alessia and Zani, Alberto (1998). Electrophysiological evidence of perceptual precedence of global vs. Local visual information. In *Cognitive Brain Research*, 6, 321-334.
- Pugin, Maryline (1983). *Les Stimulations Auditives et leur Importance dans la Genèse de la Relation Mère-Enfant*. Université de Besançon: DEA en Neurosciences.
- Pye, Clifton (1983). Mayan telegraphese: intonational determinants of inflectional development in Quiché Mayan. In *Language*, 59/3, 583-604.
- Querleu, Olivier, Renard, Xavier, Versyp, F., Paris-Delrue, L. and Crépin, Gilles (1988). Fetal hearing. In *European Journal of Obstetrics and Reproductive Biology*, 29, 191-212.
- Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. In *Psychology of Music*, 18, 60-72.
- Read, C. and Schreiber, P. (1982). Why short subjects are hard to find ? In Gleitman and Wanner.
- Rovee-Collier, Carolyn K. and Lipsitt, Lewis P. (1993). *Advances in Infancy Research*. Ablex Publishing.
- Rovee-Collier, Carolyn K., Lipsitt, Lewis P. and Hayne, (1998). *Advances in Infancy Research*. Stanford: Ablex.
- Sansavini, A., Bertoncini, Josiane and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. In *Developmental Psychology*, 33, 3-11.
- Schenker, H. (1906). *Harmony*. (ed. Jones, trans. Borgese from 1906 German ed.). Cambridge: MIT.
- Schellenberg, E. G. & Trehub, Sandra E. (1996). Natural musical intervals. In *Psychological Science*, 7 (5), 272-277.
- Schön D., Magne, C., Schrooten, M. et Besson, Mireille (2002). The music of speech: electrophysiological approach. In Bel and Marlien, *Actes de Speech Prosody*, 635-638.

- Scott, D. (1982). Duration as a cue to the perception of a phrase boundary. In *Journal of Acoustical Society*, 71, 996-1007.
- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge : MIT Press.
- Shahidullah, S. and Hepper, P. G. (1992). Hearing in the Fetus: Prenatal Detection of Deafness. In *International Journal of Prenatal and Perinatal Studies*, 4 (3/4), 235-240.
- Smith, Laura, Hains, Sylvia, M.J., Lecanuet Jean-Pierre et Kisilevsky, Barbara S. (2000). Maturation of behaviours elicited by music in 32-37 weeks GA fetuses. Poster, *ICIS 2000* (International Conference on Infant Studies), Brighton.
- Snow, Catherine E. (1977). The development of conversation between mothers and babies. In *Journal of Child Language*, 4, 1-22.
- Snow, Catherine E. and Ferguson, Charles A. (1977) *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press, 219-235.
- Spencer, M. J. and DeCasper, A. D. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. In *Infant Behavior and Development*, 10, 133-142.
- Spring, D. R. and Dale, P. S. (1977). Discrimination of linguistic stress in early infancy. In *Journal of Speech and Hearing Research*, 20, 224-232.
- Stern, Daniel N., Spiker, S. and MacKain, K. (1982). Intonation as signals in maternal speech to pre-linguistic infants. In *Developmental Psychology*, 18, 727-735.
- Stern, Daniel N., Spieker, Susan, Barnett, R. K. and Mac Kain, Kristine (1983). The prosody of maternal speech: infant age and context related changes. In *Journal of Child Language*, 10, 1-15.
- Studdert-Kennedy, Michael G. (1975). From continuous signal to discrete message: Syllable to phoneme. In Kavanagh and Cutting, 113-125.
- Thorpe, Leigh A. (1985). *Auditory-temporal organization: Developmental perspectives*. PhD, Université de Toronto.
- Thorpe, Leigh A., Trehub, Sandra E., Morrongiello, Barbara A. and Bull, Dale (1988). Perceptual grouping by infants and preschool children. In *Developmental Psychology*, 24, 484-491.
- Thorpe, Leigh A. and Trehub, Sandra E. (1989). Duration illusion and auditory grouping in infancy. In *Developmental Psychology*, 25, 122-127.
- Trainor, Laurel (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. In *Infant Behavior and Development*, 19, 83-92.
- Trainor, Laurel and Trehub, Sandra (1992). A comparison of infants' and adults' sensitivity to Western musical structure. In *Journal of Experimental Psychology: Human Perception and Performance*, ??.
- Trainor, Laurel and Rock, Adrienne, M.L. (1997). Distinctive messages in Infant-directed lullabies and playsongs. Poster, SRCD, Washington DC, April.
- Trainor, Laurel, Clark, E.D., Huntley, A. and Adams, B.A. (1997). The acoustic basis of preferences for infant-directory singing. In *Infant Behavior and Development*, 20, 383-396.
- Trainor, Laurel and Zacharias, C.A. (1999). Infants prefer higher-pitched singing. In *Infant Behaviour and Development*.
- Trehub, Sandra E. (1973). Infant's sensitivity to vowel and tonal contrasts. In *Developmental Psychology*, 9, 91-96.
- Trehub, Sandra E. (1989). Infants' perception of musical sequences: Implications for language acquisition. In *Speech-Language Pathology and Audiology*, 13, 3-11.
- Trehub, Sandra E., Bull, Dale and Thorpe, Leigh A. (1984). Infants' perception of melodies: The role of melodic contour. In *Child Development*, 55, 821-830.
- Trehub, Sandra E. and Thorpe, Leigh A. (1989). Infants' perception of rhythm. Categorization of auditory sequences by temporal structure. In *Canadian Journal of Psychology*, 43, 217-229.
- Trehub, Sandra, Thorpe, Leigh A. and Trainor, L. J. (1990). Infants' perception of good and bad melodies. In *Psychomusicology*, 9, 5-19.

- Trehub, Sandra E. and Henderson, J. (1994). Caregivers' songs and their effect on infant listeners. In Deliège (Ed.). *Proceedings of the 3rd International Conference for Music Perception and Cognition*, Liège, Belgium, ESCOM, 47-48.
- Trehub, Sandra E. and Trainor, Laurel, J. (1994). Les stratégies d'écoute chez le bébé : origines du développement de la musique et de la parole. In McAdams et Bigand, 299-347.
- Trehub, Sandra, Trainor, Laurel, J. and Unyk, A. M. (1993a). Music and speech processing in the first year of life. In *Advances in Child Behavior and Development*, 24, 1-35.
- Trehub, Sandra, Unyk, A. M. and Trainor, Laurel J. (1993b). Material singing in cross cultural perspective. In *Infant Behavior and Development*, 16, 285-295.
- Trehub, Sandra E., Unyk, A. M., Kamanetsky, S.B., Hill, D.S., Trainor, Laurel J., Henderson, J.L. and Saraza, M. (1997). Mothers' and fathers' singing to infants. In *Developmental Psychology*, 33, 500-507.
- Trehub, Sandra E. and Trainor, Laurel, J. (1998). Singing to infants: lullabies and play songs. In Rovee-Collier, Lipsitt and Hayne, 43-77.
- Unyk, A. M., Trehub, Sandra, Trainor, Laurel, J. and Schellenberg, E. G. (1992). Lullabies and simplicity: a cross-cultural perspective. In *Psychology of Music*, ??.
- Vinter, Shirley (1992). *Mise en Place des Eléments Prosodiques dans le Langage Emergent de l'Enfant Sourde : Rôle des Stimulations Acoustiques et des Interactions Sociales*. Thèse de Doctorat, Besançon.
- Vinter, Shirley (1994). *L'Emergence du Langage de l'Enfant Déficient Auditif : des Premiers Sons aux Premiers Mots*. Paris : Masson, 147 p.
- Werker, Janet F. and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross language speech perception. In *Journal of the Acoustic Society of America*, 75, 1866-1878.
- Wertheimer, M. (1923c) Untersuchung zur lehre von der gestalt II. In *Psychologische Forschung*, 4, 301-350.
- Whitaker, Harry A. (1976). Neurobiology of Language. In Caterandte and Friedman.
- Wing, H. D. (1941). *Musical Ability and Appreciation*. PhD Thesis, London University.
- Yeni-Komshian, G., Kavanagh, J. F. and Ferguson, Charles A. (1980). *Child Phonology: Production*, tome I. New-York: Academic Press.
- Yeni-Komshian, G., Kavanagh, J. F. and Ferguson, Charles A. (1980). *Child Phonology: Perception*, tome II. New-York: Academic Press, 255 p.
- Zenatti, Arlette (1969). *Le développement génétique de la perception musicale*. Monographies Françaises de Psychologie. Paris : CNRS (n°17).
- Zenatti, Arlette (1994). *Psychologie de la Musique*. Paris : PUF, 391 p.

Chapitre III

*Développement linguistique
et musical de l'enfant :
différenciation
des deux systèmes*

Table du troisième chapitre

1) Développement linguistique de l'enfant

- 1.1. Les premiers mots
- 1.2. La prosodie ou l'émergence de la première syntaxe
- 1.3. Une perte de la primauté de la prosodie au profit du niveau segmental
- 1.4. « Fin » de l'acquisition de la LM

2) Développement musical de l'enfant

- 2.1. Le rythme
- 2.2. La mélodie
- 2.3. L'harmonie
- 2.4. Les productions vocales
- 2.5. « Fin » de l'acquisition du SMM

3) Bilan de l'étude du développement linguistique et musical différencié de l'enfant

- 3.1. Tendances communes dans le développement de l'enfant
- 3.2. Baisse de la sensibilité au profit du système à apprendre
- 3.3. Mettre à profit la richesse perceptive du jeune enfant

4) Conclusions et hypothèses

- 4.1. La différenciation langue/musique au cours du développement ne condamne pas leurs similitudes
- 4.2. L'oreille musicienne
- 4.3. L'apprentissage d'une nouvelle langue, une situation proche de l'apprentissage de la LM
- 4.4. Hypothèse d'un transfert de compétence de l'écoute musicale à l'écoute « linguistique »

Bibliographie du chapitre III

Les philosophes du XVIII^e siècle voyaient une origine commune dans le langage et dans la musique et se représentaient une période, dans des temps très reculés, où ils étaient indifférenciés au sein de la communication humaine ; puis peu à peu, ils se séparèrent, notamment, selon Rousseau et Condillac, avec l'apparition de l'écriture musicale (cf. chapitre I, 1.4.). D'un point de vue ontogénique, nous avons relevé à la fin du chapitre précédent un processus similaire de différenciation entre le langage et la musique au cours du développement du bébé. Cette différenciation survient au sein même des interactions entre l'enfant et son entourage. Les stimulations apportées par les parents à leur bébé sont clairement distinctes pour le langage et la musique, car le LAE et le CAE, malgré les propriétés acoustiques et fonctionnelles qu'ils partagent, sont facilement identifiables respectivement en tant que « productions linguistiques » et « productions musicales ». Au moment de l'apparition des premières productions différenciées chez leur bébé, les parents vont immédiatement identifier celles qui relèvent de la musique et celles qui relèvent de la langue et les encourager dans l'une ou l'autre voie. C'est donc dans l'interaction entre le bébé et son entourage que se cristallisent les influences du milieu et plus largement de la culture. Une fois la séparation consommée, c'est-à-dire au moment du babillage, les deux capacités vont s'engager dans des chemins différents, qui vont diverger définitivement au moment de la scolarisation.

1. Développement linguistique de l'enfant

La prosodie continue à jouer son rôle de structure d'accueil des premiers mots et de la syntaxe émergente. Puis, à mesure que la syntaxe est « assumée » par le niveau segmental, la prosodie semble passer au second plan, d'autant que les efforts portent sur la maîtrise de l'écrit à partir de 6 ans, au moment de l'entrée au CP. Or, au moment où l'enfant semble maîtriser sa LM, il ne maîtrise pas encore toutes les relations entre syntaxe et prosodie, certaines étant très subtiles. Le poids de la LM ne cesse de se renforcer et on relève notamment des stratégies de réorganisation qui modifient les stratégies de perception de l'enfant entre le CP et le CM2.

1.1. Les premiers mots

Les éléments musicaux conservent leur prédominance et jouent un rôle essentiel dans la mise en place des premiers mots. En effet, le bébé s'emploie à remplir les moules mélodiques qu'il a mis en place dans la période précédente.

L'enfant produit de vrais énoncés, structurés au niveau intonatif et rythmique, dans lesquels apparaissent vers douze mois, les premières formes lexicales reconnaissables par l'adulte. Les premiers mots sont majoritairement constitués de sons et de combinaisons de sons appartenant au répertoire utilisé au même moment dans le babillage (Vihman et al., 1985). De Boysson-Bardies et Vihman (1991) ont relevé une continuité entre le babillage et les premiers mots dans quatre langues différentes, l'anglais, le français, le suédois et le japonais. Parallèlement, le jeu vocal coexiste à ce début de langage articulé, et il se maintiendra jusque vers le vingt-quatrième mois, voire plus tard. Cependant, malgré ce maintien tardif, il n'évoluera pas et ne connaîtra jamais de structuration, comparable à celle que l'on trouve dans le proto-langage.

Après l'apparition des premiers mots, le développement lexical du bébé se poursuit : à 13 mois, il est capable de comprendre une cinquantaine de mots, à 16 mois, une centaine. Cependant, les chiffres varient considérablement selon les auteurs. Ainsi, Bates et al. (1995) relèvent que 50 % des bébés comprennent 110 mots à 13 mois et 191 à 16 mois, tandis que Kern (2000) en trouve respectivement 86 et 152. Mais on sait que les données des parents américains sont souvent surévaluées. Cependant, tous les auteurs constatent un décalage de 4 à 5 mois entre le nombre de mots que le bébé peut comprendre et le nombre de mots qu'il peut produire (Benedict, 1979 ; Menyuk, 1994). En effet, à 14 mois, il est capable de produire une dizaine de mots et à 18 mois, une cinquantaine selon Menyuk (1994). Les chiffres de Bates et al. Concernant également une population américaine, sont respectivement de 25 et 100 aux mêmes âges, mais ceux de Kern (2000) paraissent étonnamment faibles : 5 mots à 14 pour 50% des enfants et 26 à 18 mois. A partir de 18 mois, l'enfant connaît une véritable explosion lexicale et acquiert plusieurs mots par jour, ce qui le rapproche très vite du modèle adulte. Selon les cultures, le contenu des mots est relativement universel dans le langage des enfants. Ainsi, en grande majorité, ils désignent des noms qui représentent des objets ou des personnes familières, et en proportion plus réduite, des mots d'action. Quant aux déterminants et aux adjectifs, leur nombre est très restreint dans les productions du petit enfant. Cependant, cette évolution connaît une première période de ralentissement vers 19 mois, suivie quelques mois plus tard d'une autre période pendant laquelle l'enfant paraît régresser. Or quelque temps après ce « relâchement » apparent dans la progression de l'acquisition, l'enfant va commencer à combiner des mots et entrer dans la syntaxe de sa LM.

1.2. La prosodie ou l'émergence de la première syntaxe

Au moment où l'enfant commence à combiner deux mots, c'est-à-dire vers 18-20 mois, c'est la prosodie qui lui permet de créer des liens entre ces deux mots. La syntaxe est créée par la prosodie. Ainsi, selon Konopczynski et Tessier (1994 : 174) :

« dès que l'enfant veut combiner deux ou trois mots, il se sert de la prosodie comme marqueur syntaxique. Ce n'est pas tant l'ordre des mots comme l'ont pensé Bloom (1973) et ses disciples qui structure un énoncé enfantin, mais bien l'emplacement des pauses et des points de culmination de la voix ainsi que le contour mélodique dont est affecté chaque élément de l'énoncé ».

Parallèlement à l'émergence de cette syntaxe primitive, les énoncés du bébé deviennent de plus en plus longs. Il utilise d'une manière de plus en plus marquée les contours mélodiques pour opposer deux énoncés au contenu segmental identique (valeur oppositive) et il est compris sans aucune ambiguïté par son entourage. Les premières modalités ont été mises en place de manière très précoce, dès le neuvième mois (question, énonciation, appel, phatiques). Aux alentours de 20 mois s'ajoutent les modalités impérative et exclamative. Plus l'enfant s'approche du langage articulé et plus il est redondant (contour montant associé à un mot interrogatif par exemple). L'ordre des mots apparaît autour du vingt-quatrième mois et là encore, la prosodie va jouer un rôle de structuration et d'intégration fondamental : dans un premier temps, les deux mots gardent leur individualité puisqu'ils sont séparés l'un de l'autre par une pause. Cette pause disparaît progressivement, mais l'allongement final affectant chacun des mots est conservé pendant une certaine période (parataxe, concaténation d'éléments dont le lien logique est fait uniquement par l'intonation). L'énoncé devient une authentique unité syntaxique lorsque l'allongement final affectant le premier mot disparaît pour se maintenir uniquement sur le deuxième mot (par exemple « *papado'do* » dans Konopczynski, 2001). Ce type de structuration correspond à une phrase simple. Il apparaît vers le vingt-quatrième mois et marque l'entrée dans la syntaxe (ordre des mots et apparition des marques grammaticales). Par ailleurs, jusqu'à 24 mois, les énoncés énonciatifs sont formés d'un seul contour de finalité. A mesure, que les énoncés s'allongent, la continuation majeure se met en place. Cette acquisition est fondamentale, puisqu'elle permet de découper les énoncés en continuation et en finalité (découpage intono-syntaxique).

Jusqu'à 36 mois, le remplissage des moules continue et les trous correspondant aux mots grammaticaux ainsi qu'au lexique encore ignoré sont comblés par du proto-langage, ce qui aboutit à des « énoncés mixtes » dont la structure mélodique est déjà en place. Ces mots grammaticaux sont difficiles à percevoir car ils sont généralement très brefs (qui réduit à [κ] par exemple) et très souvent « avalés » par les adultes ; de ce fait, ils sont acquis plus tardivement. Mais, avant leur apparition dans le langage de l'enfant, leur place est déjà « réservée » et « calibrée ». A 36 mois, l'enfant maîtrise l'ensemble des intonèmes du français ; il est capable de produire des phrases complexes avec plusieurs compléments et plusieurs montées intonatives (Konopczynski, 2001).

1.3. Une perte de la primauté de la prosodie au profit du niveau segmental

Après avoir joué un rôle déterminant dans l'apparition des premiers mots et dans l'apparition de la syntaxe, il semble que le rôle de la prosodie devienne moins primordial. Selon Bertoncini et de Boysson-Bardies (2000 : 105-106) :

« il est probable que la prosodie perde sa préséance au fur et à mesure que l'analyse « segmentale » de la parole se développe et s'automatise et que son traitement évolue au cours du développement, en fonction des réorganisations qui se produisent au niveau attentionnel ».

Après 3 ans, le langage évolue très rapidement et c'est donc le niveau segmental qui prend peu à peu le relais de la prosodie pour exprimer les fonctions syntaxiques. Nous ne détaillerons pas son évolution car elle n'intéresse pas directement notre travail, mais de manière très résumée, l'enfant apprend à utiliser le « je » vers 3 ans et entre progressivement dans la phase des « questions » du type « *C'est qui ? C'est quoi ? Kè c'est,...* » (pour combler la curiosité de l'enfant qui ne connaît pas de limite ; il veut tout savoir !). Cependant, l'enfant n'a pas attendu ce stade pour être capable de poser des questions à son entourage par l'intermédiaire de la modalité intonative correspondante (présente dès 9 mois dans le protolangage du bébé, voir chapitre II, 3.2.3.). Il intègre progressivement dans son langage les marques de négation et les marques de possession. Après 3 ans, il met en place les marques de possession, de négation, les adverbes et les prépositions, la conjugaison des verbes, le genre et le nombre... (Aimard, 1996). A 5 ans, la vitesse d'acquisition ralentit et le langage de l'enfant se rapproche de plus en plus de celui de l'adulte.

On considère généralement que l'enfant maîtrise la syntaxe de sa LM vers l'âge de 6 ans, car il paraît faire preuve d'une bonne compréhension et ses énoncés sont syntaxiquement corrects. Pour cette raison, l'évolution du langage après 6 ans est peu décrite, notamment en ce qui concerne l'acquisition de la prosodie. Cependant des travaux sur les rapports entre prosodie et syntaxe dans les productions de cet âge seraient importants à mener, comme le signalent Kail (1994) et Konopczynski (2002).

En effet, la prosodie, qui était première dans le langage de l'enfant, n'est toujours pas maîtrisée complètement. Les travaux de Konopczynski (1981a) montrent en effet que, si les enfants âgés de 6 à 11 ans se montrent sensibles aux paramètres prosodiques, ils ne sont pas encore capables de comprendre toutes les fonctions linguistiques qu'elle véhicule. Konopczynski (1981a) a réalisé une enquête auprès d'une centaine d'enfants âgés de 6 à 11 ans (du CP au CM2). Les enfants devaient expliquer comment ils comprenaient une série de phrases, dont la prosodie véhiculait des fonctions variées. Les résultats sont surprenants puisqu'ils révèlent que non seulement certaines structures syntaxiques ne sont pas encore dominées à l'âge de 6-7 ans, mais ils montrent en outre qu'une seule modalité est réellement maîtrisée au CP : la modalité assertive. Ainsi, les structures où la continuité prosodique est rompue (par exemple « *Regarde, Catherine !* ») ne commencent à être comprises qu'au CM1 et ne le sont encore pas totalement au CM2. La prosodie utilisée pour le discours direct et pour le discours rapporté n'est toujours pas clairement différenciée au CM2, alors que l'enfant utilise largement le discours rapporté avant 5 ans. Les exclamatives du type « *Qu'est-ce que tu manges !* » sont elles aussi intégrées difficilement, mais reconnues à 100 % au CM2. Ces travaux suggèrent que la **maîtrise de la fonction démarcative et de la fonction contrastive de l'intonation n'est pas achevée avant l'âge de 11 ans.**

Une analyse perceptive et acoustique des phrases répétées par les enfants (Konopczynski, 1981b) révèle en outre que celles dont la continuité prosodique est rompue ne sont reproduites qu'avec beaucoup de difficultés, comme si l'enfant avait besoin d'avoir compris l'énoncé pour pouvoir le répéter correctement. Par ailleurs, l'auteur relève que **plus l'expérience linguistique de l'enfant est importante, plus elle va modifier sa perception des paramètres physiques**, qui passent au second plan.

« Tout se passe comme si, chez les enfants les plus jeunes, il y avait séparation entre la perception rythmique et mélodique d'une part, que l'on pourrait considérer comme « pure » et d'autre part la perception linguistique qui n'agit pas encore sur la première, dans certains cas au moins » (Konopczynski, 1981b : 68).

Ils ignoreraient une pause pourtant très audible pour rapprocher la structure d'une structure connue (perception rythmique moins bonne que perception mélodique). Par exemple la phrase « *Il demande qui parle, à Marie* » serait rapprochée de « *Il demande qui parle à Marie* ». Ce processus de réinterprétation suggère la présence d'un filtre puissant qui adapte les données externes à la structure des représentations internes de la LM stockées par l'enfant.

1.4. « Fin » de l'acquisition de la LM

Aux yeux des adultes, la LM est apparemment en place chez l'enfant de 6 ans ; son évolution, très rapide dans les tranches d'âge précédentes, paraît s'être stabilisée pour atteindre un état relativement proche du système cible. Au niveau scolaire, cet âge correspond au CP, c'est-à-dire en France, au moment où l'enfant commence son apprentissage de l'écriture et de la lecture. Celui-ci doit apprendre à isoler et mémoriser les unités de sa langue et pour y parvenir, il doit mobiliser de grandes capacités analytiques. Ses connaissances métalinguistiques se développent, à travers un apprentissage conscient et non plus intuitif des règles de grammaire de la LM. Son système sémantique est mieux organisé et sa compréhension s'accroît très rapidement entre le CE2 et le CM1, c'est-à-dire entre l'âge de 8 et 10 ans (Konopczynski, 1981a), les structures syntaxiques et prosodiques combinées n'étant cependant pas toutes acquises avant l'âge de 11 ans. Parallèlement, nous avons vu que l'enfant connaissait une période de réorganisation perceptive entre 6 et 11 ans qui renforce encore davantage le poids de la LM et modifie sa perception des paramètres physiques de la prosodie. En effet, l'enfant réinterprète les données prosodiques pour qu'elles entrent en conformité avec les représentations internes de sa LM dans un processus finalement proche des effets de « crible » et de filtrage dont nous avons parlé dans le chapitre II (3.1.2.). On peut en conclure qu'à 11 ans, la LM est bien mise en place et que parallèlement à l'« achèvement » de cette acculturation, son poids n'a jamais été aussi fort sur la perception de l'enfant. Toutefois, les processus d'acquisition de la LM se poursuivent bien au-delà de l'âge de 11 ans et on ne saurait d'ailleurs dire s'ils se terminent un jour, la réponse à cette question étant plutôt d'ordre philosophique que linguistique.

2. Développement musical de l'enfant

Pour mettre à jour le développement musical de l'enfant, on peut étudier un certain nombre de paramètres différents tels que ses préférences pour un pattern musical particulier et ses aptitudes à discriminer deux patterns différents. On peut également analyser sa façon de reproduire un rythme ou une mélodie, ainsi que sa capacité à reproduire la pulsation d'un extrait musical. On peut s'intéresser à sa façon de symboliser un événement musical (par le dessin notamment) et enfin, on peut décrire l'évolution de ses productions vocales spontanées. Outre l'établissement d'une chronologie relatant l'apparition des différentes capacités musicales pendant l'enfance, une telle description permet de mettre à jour l'acculturation progressive de l'enfant à son SMM. Selon Zenatti (1981), l'acculturation musicale de l'enfant va se développer autour de trois grands axes, spécifiques à la musique occidentale : une sensibilité accrue pour les rythmes métriques, l'assimilation du langage tonal au niveau de la mélodie et la préférence grandissante pour la consonance au niveau de l'harmonie. Les données de Zenatti (1981), dont certains résultats sont exposées dans cette partie, résultent d'une expérimentation menée avec 8000 enfants âgés de 4 ans à 10 ans (voir chapitre VI et Volume II, annexe n°6.1. pour la description de certains de ces tests).

2.1. Rythme

Graduellement, le petit enfant réussit à restituer la pulsation d'un extrait musical en « marquant la mesure » par un mouvement du bras, du pied ou du corps entier. Selon Fraisse (1974), **dès 3 ans**, les enfants sont capables de produire **une frappe par battement**, mais ces frappes ne sont pas synchrones avec le rythme du métronome. En revanche, la plupart des enfants réussissent à se **synchroniser avec le rythme de la séquence vers l'âge de 7 ans** (Fraisse, 1974) **et ce, à différents tempi** (Drake et Penel, 1999). Cependant, chaque individu possède un « tempo personnel », c'est-à-dire une cadence déterminée qu'il va choisir spontanément pour produire des frappes rythmiques. Ainsi, si le tempo à reproduire correspond au « tempo personnel » de l'enfant, celui-ci devrait manifester des capacités plus précoces à se synchroniser avec la pulsation. Drake et Penel (1999) relèvent en effet, que si le tempo de la séquence coïncide avec son tempo personnel, le petit enfant est capable de se synchroniser avec une courte séquence dès l'âge de 1 à 2 ans et avec une séquence plus longue entre 2 et 3 ans. Ce tempo spontané est rapide chez l'enfant (Zenatti, 1994 : 324), mais il ralentit avec l'âge.

En effet, nous avons vu dans le chapitre I (1.2.1.2., « musique et métricité ») que plus l'auditeur est expérimenté, plus sa pulsation spontanée est lente car elle reflète un niveau hiérarchique plus élevé (Drake et al., 2000). **Le ralentissement du tempo personnel pendant le développement de l'enfant pourrait refléter son expérience grandissante avec son SMM** ainsi que des capacités de coordination sensorielles et motrices en évolution avec l'âge.

Un autre type de tâche révèle cette maîtrise progressive de la « métrique » : si on demande à un enfant de reproduire un rythme en le dessinant (méthodologie mise au point par Bamberger¹), il pourra réaliser deux types de dessins, soit une représentation « figurale » où les groupement entre les sons apparaîtront, soit une représentation « métrique », dans laquelle figureront tous les battements de la séquence. Ainsi, d'après Hargreaves (1995 : 188), les représentations figurales sont majoritaires chez les enfants de 2 à 5 ans. En revanche, la **capacité à représenter la structure métrique d'une séquence musicale s'améliore après 5 ans**. Ces données sont très intéressantes car cette différence de représentation pourrait refléter un traitement auditif différent, un traitement global pour les enfants jusqu'à 5 ans et un traitement analytique intervenant par la suite (la capacité à décomposer un pattern en ses éléments constitutifs relevant de la capacité analytique). En vertu du fait que les enfants de 5 ans sont également capables de discriminer deux cellules rythmiques, Zenatti (1981) pose l'hypothèse que la perception pourrait être analytique à cet âge pour des situations musicales simples. Il est à noter parallèlement qu'à 4 ans, l'enfant marque une préférence pour les agencements temporels simples, dont les éléments sont régulièrement espacés les uns des autres et **à 4 ans ½-5 ans, une préférence pour les structures organisées de façon métrique** (Zenatti, 1981).

2.2. Mélodie

Le traitement global qui domine la perception du bébé (cf. chapitre II, 2.1.2.2.) paraît se prolonger pendant un certain temps dans la petite enfance. Cependant, **il semble que des capacités de type analytique se mettent en place entre l'âge 4 et 6 ans**. En effet, une étude de Trehub et al. (1985²) montre que des enfants de cette tranche d'âge sont capables de rejeter des chansons dont les intervalles ont été modifiés. Ils manifestent donc une stratégie de traitement par intervalles qui relève de capacités analytiques (chapitre II, 1.2.3.).

¹ Cité par Hargreaves (1995 : 186) et Drake et Penel (1999).

² Cité par Dowling (1994 : 173).

Dans une tâche plus complexe, les enfants de 5-6 ans se rabattent sur un traitement par contour (Andrews et Dowling, 1991) ; ils sont donc capables de combiner les deux stratégies en fonction de la situation, comme les adultes (chapitre II, 1.2.3.).

L'apparition de ces capacités de traitement analytique correspond-elle à une **fixation par l'enfant des intervalles de son SMM**, qui, parce qu'ils ont été mémorisés, faciliteraient le traitement (contextes facilitants, voir chapitre II, 1.2.3.) ? En effet, dès 6 mois, la triade majeure permet aux bébés de dépasser leur stratégie habituelle de traitement par contour (voir chapitre II, 3.1.3.2.). Les enfants rejettent très tôt (vers 4 ans 6 mois, Zenatti, 1981) les intervalles complexes tels que la septième et préfèrent les intervalles plus petits. Vers 7 ans, les enfants marquent une préférence pour la quinte (Imberty, 1969) et à 8-9 ans, ils accordent plus d'importance aux notes de la triade majeure (Krumhansl et Keil, 1982).

Cette préférence pour certains intervalles correspond à une assimilation progressive de la structure des gammes tonales. Dans l'étude de Trehub et al. (1985), les enfants avaient tendance à rejeter les chansons dont la tonalité est éloignée de l'originale ; outre une capacité analytique, ils se montrent donc également sensibles aux changements de tonalité dès 4 ans. Dès 5-6 ans, les enfants préfèrent les musiques tonales aux musiques non tonales (Zenatti, 1981). **A 6 ans, la plupart des enfants ont intégré une partie de la syntaxe du système tonal** : ils préfèrent les extraits musicaux respectant les règles tonales et ils réussissent à distinguer un extrait tonal d'un extrait non tonal dans les épreuves d'identification mélodique (Zenatti, 1981). Parallèlement, l'apprentissage des mélodies non tonales devient plus difficile à 6-7 ans (Zenatti, 1981). Le poids du système tonal ne cessera plus de se renforcer et les préférences des enfants pour les musiques tonales deviennent très marquées à partir de 9-10 ans (Zenatti, 1981).

2.3. Harmonie

L'enfant devrait mettre plus de temps à la maîtriser la dimension harmonique que les dimensions rythmiques et mélodiques, d'autant plus qu'elle est spécifique à la musique tonale (chapitre I, 2.2.2.).

Selon Trehub et Trainor (1994 : 340) :

« l'harmonie ou la structure musicale verticale, c'est-à-dire la combinaison de notes jouées simultanément, est une caractéristique unique de la musique occidentale (Piston, 1969). Ainsi, on s'attend plutôt à ce que la sensibilité à la structure harmonique apparaisse relativement lentement, tout comme les distinctions phonémiques rares apparaissent beaucoup plus tard que celles présentes dans de nombreuses langues (Locke, 1983) ».

Il est probable que la prédominance de la courbe mélodique dans la perception du petit enfant le gêne pour appréhender la dimension harmonique. Il pourrait ne pas la percevoir correctement avant la mise en place de capacités plus analytiques.

Dans une tâche d'identification harmonique, l'enfant de 4 ans n'est pas capable de différencier un accord consonant d'un accord dissonant alors qu'au même âge, il est capable de différencier deux motifs différant uniquement par le rythme ou par la mélodie (Zenatti, 1981). Cependant, cette sensibilité évolue progressivement, puisqu'à 5 ans, il est capable de discriminer facilement des accords consonants et à 6 ans, il réussit l'épreuve d'identification à laquelle il avait échoué à 4 ans (Zenatti, 1981). **La dimension consonance/dissonance est donc difficile à maîtriser avant l'âge de 6 ans.** Cependant, même après cet âge, la prégnance de la partie supérieure gêne la discrimination harmonique, car si la note supérieure est identique dans les deux accords à discriminer, la tâche est beaucoup plus difficile pour les enfants de 6/7 ans (Zenatti, 1981). Cette prédominance de la ligne mélodique est confirmée par le fait qu'à l'âge de 7 ans, ce n'est pas l'harmonie qui donne à l'enfant le sentiment qu'une phrase musicale est achevée, mais la partie mélodique. Ainsi, si une cadence parfaite est produite sur un mouvement mélodique ascendant, elle sera perçue comme inachevée (Zenatti, 1981). **Pour être sensible à la dimension harmonique, l'enfant doit être capable de s'affranchir de la prédominance de la dimension mélodique.**

La sensibilité à l'harmonie se renforce réellement à partir de 8 ans. A cet âge, lorsque des phrases se terminent par une cadence parfaite ou une demi-cadence, les enfants perçoivent la phrase musicale comme achevée (Imberty, 1969, 1995). D'après Imberty (1995 : 246), *« les fonctions tonales sont reconnues, mais pas différenciées sur le plan fonctionnel »*. En effet, l'enfant ne perçoit pas encore les relations entre les accords qui composent la cadence (par exemple, le rapport V-I dans une cadence parfaite, c'est-à-dire dominante-tonique).

C'est seulement vers 10 ans qu'il est capable de « comprendre » les cadences et les demi-cadences. **A cet âge, la sensibilité de l'enfant à l'harmonie s'est suffisamment développée pour lui permettre de commencer à structurer des œuvres musicales harmoniques et polyphoniques** (Imberty, 1969). Il devient sensible à la qualité d'harmonisation, en préférant par exemple une version du « *Bon Roi Dagobert* » composée à partir d'accords parfaits (Imberty, 1969).

2.4. Productions vocales

Dans le chapitre précédent (chapitre II, 3.2.4.), nous avons vu que Papousek et Papousek (1981) avaient relevé chez leur bébé Tanya une tendance à produire des tons appartenant à la triade majeure dans ses chants spontanés et ce, dès 11 mois et de manière plus marquée à 14 mois. En outre, à 16 mois, ses premières improvisations révélaient une sensibilité à la musique tonale. Cette « sensibilité tonale » semble néanmoins beaucoup plus précoce que ce qui est affirmé dans la littérature, mais il existe sûrement une grande variabilité dans l'apparition des premières productions musicales, en raison d'une exposition « musicale » très variée selon les milieux. Ainsi, selon Dowling (1994), entre 1 et 2 ans, certes, l'enfant contrôle l'évolution de la courbe, mais les intervalles ne sont pas encore fixés (fluctuation de la hauteur et absence de centre tonal constant).

A partir du recueil des chansons d'environ 69 enfants (étude transversale) et des chansons de 9 enfants suivis longitudinalement, Davidson (1994³) relève qu'à partir de 3 ans, les enfants prennent appui sur les paroles et produisent des hauteurs distinctes, mais les intervalles ne sont pas stables et il n'y a pas encore de cohérence tonale. Vers 5 ans, ils sont capables de reproduire à la fois les contours et les intervalles.

Dans le cadre de sa « *grammaire musicale évolutive* » (1982, 1989), Imberty explique que, entre 3 et 6 ans, dans les productions improvisées,

« la séquence musicale est construite autour d'un intervalle aux bornes fixes [nous soulignons que ces bornes ont des hauteurs précises], à l'intérieur duquel l'enfant effectue toutes sortes de mouvements vocaux dont les sons réels n'ont aucune hauteur précise ».

³ Cité par Hargreaves (1995 : 188-189).

Par ailleurs, il souligne que la séquence est « *orientée à droite* », c'est-à-dire que

« tout le poids de la séquence est reporté sur sa finale, rendant tout lien avec une éventuelle séquence ultérieure impossible » (1995 : 240).

A partir de 6 ans, l'« *intervalle-pivot* », c'est-à-dire l'intervalle entre la hauteur initiale et la hauteur finale de la séquence, serait souvent, d'après les observations de l'auteur, un intervalle de quarte ou de quinte. Jusqu'à 6 ans, les fragments n'entretiennent pas de rapports les uns par rapport aux autres, ils sont simplement juxtaposés. Pour cette raison, Imberty (1995 : 243) parle de « *schèmes d'ordre simple* ». Mais, à partir de 6 ans et jusqu'à 8-9 ans, l'enfant va les articuler :

« les schèmes d'ordre constituent l'ensemble des intuitions que le sujet a des successions temporelles sans qu'il ait conscience des éléments constitutifs de ces successions » (1995 : 243).

Durant cette période, il est capable de reproduire une chanson, mais s'il est interrompu, il doit reprendre au début, « *la séquence n'ayant d'existence que globalement* » (1995 : 243-244). Plus tard, l'enfant développe des « *schèmes de relation d'ordre* », qui organisent les segments les uns par rapport aux autres selon les règles d'une syntaxe, d'une grammaire de la musique tonale.

2.5. « Fin » de l'acquisition du SMM

Il semble qu'à partir de 12 ans, le développement musical de l'enfant ralentisse d'une manière relativement brusque et que

« tout se passe comme si l'évolution de la perception et de l'intelligence musicales était bloquée » (Imberty, 1969 : 217).

Les effets de l'acculturation musicale semblent moins « visibles » et le SMM est bien dominé. Ce savoir intériorisé permet à l'enfant de structurer sa perception de la musique et une fois que les trois dimensions musicales (rythme / mélodie / harmonie) sont convenablement traitées, il peut se spécialiser entre 8 et 15 ans sur les règles et les conventions artistiques de sa culture musicale (Hargreaves, 1995 : 192). Comme pour la langue quoiqu'un peu plus tard peut-être, cette période correspond au passage d'une acquisition implicite par imprégnation avec l'environnement à un apprentissage « explicite » plus ou moins intensif prodigué par l'institution scolaire (cours de musique à l'école / écoles de musique / conservatoire...).

Par ailleurs, il semble que cette spécialisation provoque chez l'enfant une baisse de sa sensibilité aux autres cultures musicales et aux différents styles musicaux. Ainsi, selon Gardner (1973), la sensibilité pour les divers styles de musique classique est plus marquée chez un enfant de 11 ans que chez des adolescents de 14 à 18 ans. Castell (1982) relate le même phénomène pour les différents styles de musique populaire, à la différence près que cette sensibilité semble plus précoce (8-9 ans) et qu'elle s'atténue plus tôt, vers 11 ans. Par ailleurs, LeBlanc (1991) relève d'après un bilan de la littérature que si, jusqu'à 8 ans, l'oreille de l'enfant est ouverte et qu'il apprécie une grande variété de styles musicaux, son oreille se ferme progressivement à partir de l'adolescence pour se polariser sur la seule musique « pop » s'il n'a été exposé qu'à ce type de musique pendant l'enfance.

3. Bilan de l'étude du développement linguistique et musical différencié de l'enfant

3.1. Tendances communes dans le développement

L'étude successive du développement rythmique, mélodique et harmonique de l'enfant se révèle riche d'enseignements sur le développement musical de l'enfant jusqu'à 8-10 ans. Elle révèle notamment que la sensibilité à l'organisation métrique (stabilisée vers l'âge de 7 ans) et tonale (stabilisée vers 6 ans) de la musique occidentale semble se développer plus rapidement que la sensibilité à l'organisation harmonique (renforcée vers 8 ans et stabilisée vers 10 ans). La perception de l'harmonie nécessite plus d'expérience de la part de l'enfant. Pour pouvoir la percevoir, il doit d'abord se soustraire de la domination de la mélodie qui occulte les parties inférieures. Cette domination serait semble-t-il caractéristique d'un traitement de type global, alors que la perception de l'harmonie nécessiterait un traitement de type analytique : en effet, pour pouvoir discriminer deux accords ayant la même note supérieure, il faut pouvoir en identifier les composantes et les différencier.

Or, entre 4 et 6 ans, l'enfant passe justement d'une stratégie « globale » à une stratégie « locale » : à partir de 5 ans, il est capable de représenter le rythme de façon métrique (ce qui nécessite de pouvoir isoler les éléments les uns des autres au niveau perceptif avant de pouvoir les représenter par le dessin) et au même âge, il est capable de traiter des mélodies par intervalles.

Ce changement de stratégie provient-il d'une réorganisation cognitive ou est-il provoqué par le poids grandissant de l'expérience avec le SMM ? En effet, nous avons vu dans le chapitre II (1.2.3.) que la mémorisation des contextes les plus fréquents dans le SMM (les fameux contextes « facilitants ») oriente vers une stratégie plus analytique. Mais nous avons vu également que la maturation du cortex auditif n'est pas achevée avant l'âge de 5 ans, favorisant un traitement de type global. Il est difficile de trancher et peut-être même n'est-il pas souhaitable de le faire, cette réorganisation perceptive résultant d'une interaction étroite entre une réorganisation cognitive et l'assimilation progressive du milieu musical.

Quoi qu'il en soit, jusqu'à 5 ans, l'enfant éprouve des difficultés à intégrer les détails en un ensemble structuré et à traiter plusieurs niveaux en même temps. Pour lui, l'ensemble prime sur les éléments. Ce mode de fonctionnement « global » est général et s'applique également au traitement de la langue. Mais, progressivement il va réussir à isoler telle ou telle composante sonore au sein de cet ensemble et à en abstraire les propriétés pour construire une représentation et la mémoriser. Vers l'âge de 6-7 ans, les capacités analytiques sont déjà bien en place et elles permettront à l'enfant d'apprendre à lire et à écrire sa LM lors de son entrée au cours préparatoire (CP).

Le changement de traitement permet à l'enfant de se spécialiser sur sa LM et son SMM, mais cette centration provoque également une baisse de sa sensibilité pour les autres langues et les autres systèmes musicaux.

3.2. Baisse de la sensibilité au profit d'une centration sur le système à apprendre

Pour apprendre sa LM et son SMM, l'enfant va devoir centrer son attention exclusivement sur eux et se conformer à leur organisation pour en apprendre les règles. Drake et Penel (1999 : ?) se posent la question suivante, « *Do we need to conform before we can produced once again creatively ?* ». Il est en effet très probable que l'enfant doive plier ses larges capacités pour s'adapter à son environnement. Cette adaptation est bien sûr nécessaire, car sans elle, il ne pourrait apprendre ni sa LM, ni son SMM. Une fois ces règles intériorisées, l'enfant peut à nouveau jouer avec les sons de la langue comme il joue à nouveau avec les sons de la musique. Mais retrouvera-t-il pour autant la créativité qui était la sienne pendant ses premières années de vie ?

La question reste en suspens, mais si les règles, une fois maîtrisées, peuvent être transcendées (nous en avons donné un exemple avec la remise en cause des lois tonales par les compositeurs contemporains, chapitre I, 2.2.2.), ces transgressions semblent ensuite dépendre des individus (cf. le musicien de formation classique qui n'arrive pas à improviser en jazz).

Cette spécialisation, si elle est nécessaire à l'apprentissage, provoque chez l'enfant la perte d'une grande partie de sa sensibilité aux systèmes langagiers et musicaux non natifs. Nous avons vu qu'au cours du développement, il se produit plusieurs réorganisations perceptives, les premières intervenant dès le milieu de la première année (Chapitre II, 3.1.2.). L'acculturation consiste donc, pour la langue comme pour la musique, en un cycle de réorganisations qui oriente l'enfant sur un **traitement de plus en plus profond, mais parallèlement à cet approfondissement, une perte de sa sensibilité intervient pour les systèmes sur lesquels il ne se spécialise pas**. Nous pourrions donc énoncer le théorème suivant sur les processus d'acculturation : « **L'imprègnation progressive de l'enfant avec son système natif provoque une baisse de sa sensibilité aux systèmes non natifs** ».

3.3. Mettre à profit la richesse perceptive du jeune enfant

En vertu de ce théorème, il paraît logique de penser que, si on veut sensibiliser un enfant à un nouveau système musical ou à une nouvelle langue, il faudra le faire avant que la LM et le SMM soient complètement établis. Sinon, il est probable que les représentations fixées par l'environnement fonctionnent comme de puissants filtres qui vont entraver l'apprentissage de cette nouvelle langue ou de ce nouveau système musical. Il conviendrait donc de diversifier l'écoute musicale et linguistique de l'enfant de façon la plus précoce possible.

4. Conclusions et hypothèses

4.1. La différenciation langue/musique au cours du développement ne condamne pas leurs similitudes

Au cours du chapitre I, nous avons fait apparaître les similitudes très fortes de forme entre la langue et la musique, avec comme principal point de convergence les éléments musicaux dans la langue, c'est-à-dire la prosodie. Le chapitre II a révélé un traitement de la langue et de la musique très proche au début de la vie, le bébé manifestant une sensibilité très similaire à la musique et à la langue et ses parents s'adaptant à cette sensibilité par des adaptations vocales très musicales.

Les premières productions du bébé vont également s'illustrer par une grande musicalité et il est dans un premier temps très difficile d'y différencier ce qui relève de la musique et ce qui relève de la parole.

Cependant, le développement musical et le développement linguistique connaissent ensuite des chemins très divergents, qui conduisent l'enfant à l'apprentissage progressif de sa LM et de son SMM. Cette séparation se creuse encore plus nettement au moment de la scolarisation et notamment au moment où l'enfant entre au CP. Par le passé, l'oral a été délaissé car l'accent était mis sur l'écrit pour permettre l'apprentissage de la lecture et de l'écriture (ne serait-ce que parce que le programme très chargé du CP ne laissait pas assez de temps pour d'autres préoccupations). La musique de la langue maternelle est mise de côté, alors que la prosodie n'est véritablement en place qu'à 11 ans (Konopczynski, 1981a). Cette non prise en compte de l'oral à l'école a contribué à classer la langue et la musique dans deux sphères nettement délimitées. Mais, **ce n'est pas parce qu'elles sont culturellement différenciées, qu'il n'y a plus de similitudes possibles et il nous semble légitime de se demander si la pratique musicale n'influence pas la pratique de la langue**, notamment au niveau de ses caractéristiques prosodiques.

L'école devient de plus en plus sensible à la musique (multiplication des classes d'éveil musical, augmentation du taux horaire de la pratique musicale hebdomadaire à l'école...) et à l'oral. Ce n'est pas une totale nouveauté, puisque dès 1972, le Plan Rouchette insistait sur la « priorité à l'expression orale » (en L.M.). Il fut complété par de nombreuses circulaires (Fayol...) et plus récemment, cette priorité se manifeste également à travers l'apprentissage précoce des langues. Il serait temps donc de se poser la question des rapports qu'entretiennent la langue et la musique et comment les exploiter leurs similitudes au niveau scolaire. Ainsi, dans l'introduction de son livre au nom évocateur « *Développer les Capacités d'Ecoute à l'Ecole, Ecoute Musicale, Ecoute des Langues* », Ribière-Raverlat (1997 : 4) souligne que :

« si l'on tient compte du fait que l'organe sensoriel de l'ouïe permet de capter toute matière sonore qui parvient à nos oreilles, les bruits de l'environnement, la musique, mais aussi la parole [...], on prend véritablement conscience que l'éducation de l'oreille ne se limite pas au seul domaine musical mais qu'elle a une dimension transversale qui concerne le domaine de la langue ».

4.2. L'oreille musicienne

4.2.1. Qu'est-ce que l'oreille musicienne ?

Si l'éducation de l'oreille a une « dimension transversale » sur le traitement de la langue, en quoi consiste le fait d'avoir une oreille éduquée musicalement et qu'est-ce qui caractérise une oreille musicienne d'une oreille non musicienne ?

Il existe deux types d'oreilles musicales. La plus mythique des deux, c'est bien sûr **l'oreille absolue**, bien que ce type de traitement reste exceptionnel chez l'homme. L'oreille absolue se mettrait en place à la suite d'une exposition musicale précoce et prolongée pendant l'enfance (Cohen et Baird, 1990). Ce type de traitement permet au musicien d'identifier et de nommer automatiquement les sons, sans avoir besoin de se référer à un son « référence », le la^3 par exemple (cette référence ayant été mémorisée au cours de l'enfance). Un musicien qui possède l'oreille absolue n'a donc théoriquement pas besoin d'accordeur. Le violoniste Julian Rachlin décrit cette capacité dans les termes suivants :

« Je l'avais à trois ans bien avant d'apprendre à lire et à écrire, et bien avant d'apprendre le solfège. La musique et avec elle toutes les notes des instruments avaient chacune pour moi un visage aussi évident que ceux de mes parents et de tous les éléments de mon entourage. Mais, à l'époque, j'étais incapable de les nommer dans leur totalité, parce que je savais à peine parler » (Chouard, 2001 : 71).

Cependant, **l'oreille relative** constitue le type de perception le plus courant et le plus naturel. Dans ce type de perception, les hauteurs ne sont pas reconnues en fonction de leur hauteur absolue, mais elles sont déterminées les unes par rapport aux autres en fonction d'une hauteur référence, qui n'a pas été intériorisée, mais qui est donnée au musicien (par le diapason par exemple). La perception relative est commune aux musiciens et aux non musiciens, puisque tous les auditeurs perçoivent la hauteur de façon relative, y compris dans la langue, la différence se faisant uniquement dans le degré d'entraînement de l'oreille musicienne. En effet, grâce à une gymnastique auditive intervenant au début de l'apprentissage musical, le musicien apprend à identifier les notes le plus rapidement possible et à les transcrire. Cette gymnastique est difficile et nécessite beaucoup de concentration, comparée à la reconnaissance automatique des hauteurs qui caractérise l'oreille absolue.

Qu'elle soit relative ou absolue, l'oreille musicienne se caractérise par des capacités supérieures à l'oreille non musicienne pour le traitement des sons. Ces capacités se développent au cours de la formation musicale, par la pratique des dictées musicales (transcription d'une mélodie, d'un rythme ou les deux à la fois, dictée d'accords), du complément de texte (écoute d'une œuvre musicale et retranscription des passages manquants, la partie de hautbois dans une partie d'orchestre par exemple), d'analyse d'œuvres (reconnaissance des thèmes et identification de la structure de l'œuvre...). L'oreille musicienne apprend donc à isoler les différents éléments qui composent un « ensemble » sonore et à les identifier, ce qui permet ensuite leur retranscription sur portée musicale. Après un tel entraînement, un musicien ne perçoit plus la musique comme un non musicien. Ainsi, dans le domaine rythmique, Drake et al. (2000) ont étudié, chez des enfants musiciens et non musiciens, la capacité à restituer la pulsation d'un extrait musical (voir également ce chapitre 2.1., chapitre I, 1.2.1.). Les résultats montrent que, plus l'expérience musicale de l'enfant est grande, plus ses battements sont lents. Ce ralentissement témoigne que grâce à la pratique musicale, l'enfant accède à des niveaux hiérarchiques supérieurs.

Si un musicien entend différemment, il est donc légitime de se demander comment son oreille entraînée va percevoir les éléments musicaux de la langue et s'il se produit un transfert de compétence du traitement de la musique à celui de la prosodie. Il existe peu de travaux expérimentaux dans ce domaine spécifique, mais plusieurs études se sont intéressées aux transferts de la pratique musicale sur les performances scolaires et notamment sur la pratique linguistique en général (lecture, aisance langagière...) et sur la pratique d'autres disciplines comme les mathématiques.

Gardiner et al. (1996) ont réalisé une expérimentation avec une population de 96 enfants, âgés de 5 à 7 ans (4 classes contrôles). Les classes expérimentales ont suivi un entraînement musical régulier à partir de la méthode Kodaly (fondée sur le chant et le chant populaire). Après 7 mois de cet entraînement, il est apparu d'après les commentaires des professeurs et les résultats au jardin d'enfant, que les **performances des enfants « musiciens » étaient plus élevées en mathématiques** que celles des enfants « non musiciens », alors qu'aucune différence n'était relevée entre les groupes pour la lecture. L'année suivante, l'étude a continué dans la même école et avec les mêmes enfants (4 classes expérimentales et 5 classes contrôles) et elle est venue confirmer les premiers résultats. Les auteurs en ont conclu que l'apprentissage musical facilitait l'apprentissage des mathématiques.

Comme dans l'étude de Gardiner et al., Mingat et Suchaut (1996) ont relevé des **influences de la pratique musicale sur les acquis en mathématiques, mais également sur la lecture**. Leur étude porte sur une population de 900 enfants, répartis au sein de 46 classes. Les classes expérimentales ont bénéficié d'un enseignement musical spécifique de 2 heures hebdomadaires ou de 4 heures hebdomadaires, fondé sur une méthode originale⁴. Les résultats font apparaître des performances scolaires supérieures dans les classes expérimentales. Ces résultats valent pour les acquis de grande section de maternelle, mais également sur les scores en lecture et en mathématiques en fin de C.P. Par ailleurs, les enfants de nationalité étrangère ayant les acquis les plus faibles en début de grande section et les enfants n'ayant pas bénéficié d'une scolarisation précoce profitent plus que les autres enfants de l'apprentissage musical pour combler leur retard. Dans ce cas, la musique constitue un **facteur d'intégration à l'école pour les enfants en difficulté**. Ces résultats rejoignent le constat établi par Gordon (1975) qui montre que des enfants issus de milieux défavorisés et rencontrant des difficultés scolaires en CM1 et CM2 peuvent égaler le niveau des enfants issus de milieux hétérogènes au bout de cinq ans si on leur donne la possibilité d'apprendre un instrument.

Enfin, Weber et al. (1993) ont travaillé avec 70 classes d'enfants âgés de 7 à 15 ans. 35 de ces classes bénéficiaient de 5 séances de 45 minutes de musique par semaine, contre 1 ou 2 séances dans les classes contrôle. Par ailleurs, le temps accordé aux mathématiques ou à l'étude de la langue a été réduit dans ces classes expérimentales. Après 3 années, les auteurs ont relevé que les enfants des classes expérimentales avaient des **capacités linguistiques supérieures** (notamment pour restituer, à l'écrit et à l'oral, une histoire qui leur avait été racontée). Quant aux mathématiques, ils ne se montraient pas plus mauvais que les classes qui avaient suivi un enseignement plus important. En outre, les professeurs ont rapporté que ces enfants apprenaient à lire plus rapidement et que les relations étaient plus coopératives et plus amicales en classe.

Ces travaux montrent donc que la pratique musicale semble exercer un certain nombre d'effets positifs sur les performances scolaires, et en particulier, pour ce qui nous intéresse sur la pratique de la LM. Mais ils ne nous donnent aucune donnée sur l'écoute musicienne en elle-même. En effet, cet accroissement des performances peut trouver son origine dans d'autres capacités mises en jeu dans la formation musicale (telle que la transcription musicale pour l'écrit et la lecture de notes pour la lecture).

⁴ Bachelard, Annie, Coulon, Daniel et Loisy, Jean-Paul (1996). *Musique au Quotidien au Cycle 2 : Des Parcours d'Apprentissage de la Maternelle au CE1*. CRDP de Bourgogne, huit cahiers, un CD.

Par ailleurs, nous manierons ces résultats avec de très grandes précautions à cause de l'influence de l'expérimentateur et de la situation expérimentale sur la performance des sujets étudiés. La première observation de ce type a été réalisée sur les ouvrières de la Western Electric Company à Hawthorne, près de Chicago, dans les années 30. A leur grande surprise, les chercheurs constatèrent une hausse de leur productivité lorsque l'éclairage baissait, parce qu'elles étaient observées. Cet effet fut baptisé « effet Hawthorne ». Rosenthal et Jacobson (1968) ont démontré le même effet au sein du milieu scolaire : si le professeur a prophétisé la réussite de certains de ses élèves, ceux-ci auront tendance à mieux réussir que les autres : c'est l'« *effet Pygmalion* » à l'école (Rosenthal et Jacobson : 1968) (pour une description détaillée de l'effet Hawthorne et de Pygmalion, se reporter au Volume II, annexe n°7.1.).

Le montage expérimental utilisé dans les travaux que nous avons exposés plus haut correspond exactement à une situation où l'effet Hawthorne se manifeste (classe bénéficiant d'un apprentissage musical, classe bénéficiant d'un apprentissage musical moindre, classe ne bénéficiant pas d'un apprentissage musical), mais il semble que ces travaux n'en aient pas tenu compte. Les classes « musique » se caractérisent par une activité particulière qui les met en valeur par rapport aux classes « non musique ». A cause de l'effet Hawthorne, il est impossible de déterminer si c'est cette caractérisation, cette « mise en valeur » qui est responsable des bons résultats des enfants ou la variable indépendante étudiée, ici la musique. Il est donc impossible de déterminer de manière assurée que, par exemple dans l'étude de Gardiner et al. (1996), la musique favorise l'apprentissage des mathématiques.

4.3. L'apprentissage d'une nouvelle langue, une situation proche de l'apprentissage de la LM

Il semble difficile de faire apparaître un transfert de compétence des capacités de traitement de l'oreille musicienne au traitement de la prosodie de la LM chez des enfants en âge scolaire, car la prosodie semble moins prédominante qu'en début d'acquisition (1.3.) ; elle a perdu une partie de sa primauté au profit du segmental. De plus, l'importance du niveau segmental est renforcée par l'apprentissage de la lecture et de l'écriture. Pour ces raisons, il nous semble plus intéressant de déterminer comment une oreille musicienne perçoit la prosodie d'une langue étrangère. En effet, lorsqu'un auditeur est confronté à une langue inconnue, son attention se porte en priorité sur le niveau prosodique, comme le fait le bébé avec sa LM.

L'analogie avec l'acquisition de la LM est certes un peu facile d'autant que les processus mis en jeu dans l'apprentissage d'une langue étrangère sont très différents de ceux mis en jeu dans l'acquisition de la LM. Mais elle pourrait s'avérer pertinente au moins pour la primauté du niveau prosodique.

Il est en effet probable que l'oreille musicienne soit particulièrement sensible à la nouvelle organisation musicale de la langue étrangère, qu'elle en « décode » les caractéristiques plus rapidement et qu'elle s'oriente vers un traitement de type analytique. Cependant, il convient d'étudier ce transfert de compétence avec des enfants jeunes, afin d'éviter que le poids de la LM ne soit trop renforcé (cf. 3.3.) et ne provoque des effets de filtre trop importants dans la perception d'une langue étrangère. En effet, s'il est probable qu'une oreille musicienne se montre mieux entraînée à percevoir les éléments musicaux, ces capacités ne lui permettront pas forcément d'échapper au phénomène de crible. Les travaux de Lynch et al. (Lynch et Eilers, 1992 ; Lynch et al., 1991) exposés dans le chapitre II (3.1.3.3.) prouvent même le contraire dans le domaine de la musique, puisque si les musiciens se montrent plus performants à percevoir les gammes natives, ils ne se montrent en revanche pas plus performants dans la perception des gammes non natives.

4.4. Hypothèse d'un transfert de compétence de l'écoute musicale à l'écoute « linguistique »

A l'issu de ces réflexions, nous posons comme hypothèse qu'il existe un transfert de compétence de la pratique musicale à la pratique d'une langue étrangère, du fait que l'entraînement qui caractérise une oreille musicienne se reporte au traitement des éléments musicaux de la nouvelle langue. Par ailleurs, il serait intéressant de déterminer si les capacités *analytiques* qui caractérisent une oreille musicienne se reportent au traitement des phonèmes de la langue étrangère, dont la perception nécessite également un traitement de type analytique. Nous établirons donc une analogie entre l'analyse de la forme musicale et celle de la forme verbale à deux niveaux : un niveau « global » correspondant au traitement de la prosodie et un niveau « local » correspondant au traitement des phonèmes.

Afin d'étudier le transfert de compétence de la pratique musicale sur l'apprentissage d'une langue étrangère, il faudrait profiter d'une période où ce transfert serait le plus facile, c'est-à-dire au moment où l'oreille de l'enfant est ouverte et pas encore trop centrée sur sa LM et son SMM.

Il nous faut donc déterminer la période la plus appropriée pour l'apprentissage d'une langue en fonction des données mises à jour dans le bilan sur le développement linguistique et musical de l'enfant, mais également en fonction des travaux menés sur les effets de l'apprentissage des langues aux différents âges. Il nous faudra ensuite déterminer si cette période est mise à profit par l'école.

Bibliographie chapitre III

- Aiello, Rita and Sloboda, John A. (1994). *Musical Perceptions*. New-York: Oxford University Press, 304 p.
- Aimard, Paule (1996). *Les Débuts du Langage chez l'Enfant*. Dunod : Paris, 226 p.
- Andrews, M. W. and Dowling, Jay W. (1991). The developments of perception of interleaved melodies and control of auditory attention. In *Music Perception*, 8, 349-368.
- Bates, Elizabeth, Dale, Philips and Thal, Donna (1995). Individual differences and their implications for theories of language development. In Fletcher and Mac Whinney, 96-151.
- Benedict, H. (1979). Early lexical development: comprehension and production. In *Journal of Child Language*, 6, 183-200.
- Bertoncini, Josiane et de Boysson-Bardies, Bénédicte (2000). La perception et la production de la parole avant deux ans. Dans Kail et Fayol, 95-136.
- Boysson-Bardies, Bénédicte de and Vihman, Marilyn M. (1991). Adaptation to language: evidence from babbling and first words in four language. In *Journal of Child Language*, 67, 297-319.
- Castell, K. C. (1982). Children's sensitivity to stylistic differences in "classical" and "popular" music. In *Psychology of Music*, Special Issue, 22-25.
- Chouard, Claude-Henri (2001). *L'Oreille Musicienne*. Paris : Gallimard, 347 p.
- Cohen, A. and Baird, R. (1990). Acquisition of absolute pitch: The question of critical periods. In *Psychomusicology*, 9, 31-37.
- Davidson, L. (1994). Songsinging by young and old: a developmental approach to music. In Aiello and Sloboda, 99-130.
- Deliège, Irène et Sloboda, John A. (1995). Naissance et Développement du Sens Musical. Paris : PUF, 307 p.
- Dowling, Jay W. (1994). La structuration mélodique. Dans Zenatti, 145-176.
- Drake, Carolyn and Penel, Amandine (1999). Learning to play music: rhythm and timing. In *Parole*, 9/10, 49-62.
- Drake, Carolyn, Jones, Riess, Mari and Baruch, Clarisse (2000b). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. In *Cognition*, 77, 251-288.
- Fletcher, Paul and Mac Whinney, Brian (1995). *The Handbook of Child Language*. Cambridge: Blackwell.
- Fraisse, Paul (1974). *La Psychologie du Rythme*. Paris : PUF, 244 p.
- Gardiner, Martin F, Fox, Alan, Knowles, Faith and Jeffrey, Donna (1996). Learning improved by arts training. In *Nature*, 381/6580, 284.
- Gardner, H. (1973). Children's sensitivity to musical styles. In *Merrill-Palmer Quarterly*, 19, 67-7.
- Girard, Denis (1995). *Enseigner les Langues : Méthodes et Pratiques*. Paris : Bordas, 175 p.
- Gordon, Edwin E. (1975). Fourth-year and fifth-year final results of a longitudinal study of the musical achievement of culturally-disadvantaged students. In *Experimental Research in the Psychology of Music*, 10, 24-52.
- Hargreaves, David (1995). Développement du sens artistique et musical. Dans Deliège et Sloboda, 167-197.
- Halford, Brigitte K. and Pilch, Herbert (1994). *Intonation*. Tübingen: Gunter Narr Verlag.
- Imberty, Michel (1969). *L'Acquisition des Structures Tonales chez l'Enfant*. Klincksieck : Paris, 226 p.
- Imberty, Michel (1995). Développement linguistique et musical. Dans Deliège et Sloboda, 223-249.
- Kail, Michèle (1994). Le Développement du langage et les sciences cognitives. Dans *Psychologie Française*, 39/1, 29-47.
- Kail, Michèle et Fayol, Michel (2000). *L'Acquisition du Langage : le Langage en Emergence de la Naissance à Trois ans*. Vol. I. Paris : PUF, 304 p.
- Kern, Sophie et Lange, Jacques (2000). Des premiers gestes aux premiers mots : le développement communicatif chez l'enfant de 8 à 30 mois. *Troisième Journées Scientifiques de l'Ecole d'Orthophonie*, Lyon, CD-Rom, 18 p.

- Konopczynski, Gabrielle (1977). Jugement linguistique, jugement psycho-physique dans la perception de contours intonatifs : interrelations. Dans *Travaux de L'Institut de Phonétique de Strasbourg*, 9, 205-220.
- Konopczynski, Gabrielle (1981a). Compréhension des structures intonatives et de leur fonction linguistique par des enfants de 6 à 11 ans : un exemple d'expérimentation. Dans *Working Papers of Speech and Voice Society*, Toronto, 2, 50-90.
- Konopczynski, Gabrielle (1981b). Remarques sur la perception des structures rythmiques et mélodiques par les enfants de 6 à 11 ans. Toronto, *Symposium Prosodie*, 61-75.
- Konopczynski, Gabrielle (2001). Maturation syntaxique chez le jeune enfant (9-36 mois). Continuité ou rupture ? Neuchâtel : Tranel, *Actes du Colloque « Le Changement Linguistique »*, 34/35, 251-272.
- Konopczynski, Gabrielle (2002). *Linguistique de l'Acquisition : Etat des Lieux*. Cours de D.E.A. de Sciences du Langage, mention F.L.E. à distance, Université de Franche-Comté, 15 pages.
- Konopczynski, Gabrielle et Tessier, Sophie (1994). Structuration intonative du langage émergent. Dans Halford et Pilch, 157-192.
- Krashen, Stephen D. and Seliger, H. W. (1975). Conscious language learning.
- Krumhansl, Carol L. and Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. In *Memory and Cognition*, 10, 243-251.
- LeBlanc, A. (1991). Effect of maturation/aging on music listening preference: a review of the literature. Communication au *9th National Symposium on Research in Musical Behavior*, Cannon Beach, Oregon, USA.
- LeCours, A.R. (1975). Myelogenetic correlates of the development of speech and language. In Lenneberg and Lenneberg.
- Lenneberg, Eric Heinz and Lenneberg, Elizabeth (1975). *Foundations of Language Development of Speech and Language* (Vol. 1). New-York: Academic Press, 350 p.
- Lipsitt, Lewis P. and Rovee-Collier, Carolyn K. (1981). *Advances in Infancy Research*. Vol I : XXII. Norwood: Ablex, 267 p.
- McAdams, Stephen et Bigand, Emmanuel (1994). *Penser les Sons : Psychologie Cognitive de l'Audition*. Paris : P.U.F., 402 p.
- Menyuk, Paula (1994). Word acquisition: the important role of prosody. In *The Study of Sounds*, 23, 218-222.
- Mingat, A. et Suchaut, B. (1996). Incidences des activités musicales en grande section de maternelle sur les apprentissages précoces au cours préparatoire. Dans *Les Sciences de l'Education*, 29, 3, 49-76.
- Newport, Elissa L. (1990). Maturational constraints on language learning. In *Cognitive Science*, 14, 11-28.
- Papousek, Mechthild and Papousek, Hanus (1981). Musical elements in the infants vocalizations: Their significance for communication, cognition and creativity. In Lipsitt and Rovee-Collier, 163-224.
- Rivière-Raverlat, Jacquotte (1997). *Développer les Capacités d'Ecoute à l'Ecole. Ecoute Musicale, Ecoute des Langues*. Paris : PUF, 207 p.
- Schön D., Magne, C., Schrooten, M. and Besson, Mireille (2002). The music of speech: electrophysiological approach. In Bel and Marlien, *Actes de Speech Prosody*, 635-638.
- Trehub, Sandra E., Bull, Dale and Thorpe, Leigh A. (1984). Infants' perception of melodies: The role of melodic contour. In *Child Development*, 55, 821-830.
- Trehub, Sandra E., Morongiello, B. A. and Thorpe, Leigh A. (1985). Childrens' perception of familiar melodies: The role of intervals, contour, and key. In *Psychomusicology*, 5, 39-48.
- Trehub, Sandra E. et Trainor, Laurel, J. (1994). Les stratégies d'écoute chez le bébé : origines du développement de la musique et de la parole. In McAdams et Bigand, 299-347.
- Vihman, Marylin M., Macken, M. A., Miller, R., Simmons, H. & Miller, J. (1985). From babbling to speech: A reassessment of the continuity issue.
- Weber, E. W., Spychiger, M. und Patry, J. L. (1993). Musik macht Schule. Biografie und Ergebnisse eines Schulversuchs mit erweiten Musikunterricht. Essen.
- Zenatti, Arlette (1981). *L'Enfant et son Environnement Musical. Etude Expérimentale des Mécanismes Psychologiques d'Assimilation Musicale*. Issy Les Moulineaux : E.A.P., 275 p.
- Zenatti, Arlette (1994). *Psychologie de la Musique*. Paris : PUF, 391 p.

Chapitre IV

*L'apprentissage précoce
des langues étrangères*

Table du quatrième chapitre

1. Age et apprentissage

- 1.1. Existence d'une période critique pour l'apprentissage d'une LE
 - 1.1.1. La découverte d'une période sensible pour l'acquisition de la LM...
 - 1.1.2. ...étendue à l'apprentissage des langues secondes
- 1.2. A quel âge commencer l'apprentissage d'une LE ?
 - 1.2.1. Pendant l'âge heureux ?
 - 1.2.2. Pendant l'âge critique ?
 - 1.2.3. Pendant le seuil fatidique ?
 - 1.2.4. Conclusion

2. Etat des lieux de l'apprentissage précoce

- 2.1. Qu'est-ce qu'un apprentissage « précoce » ?
- 2.2. Bref historique
 - 2.2.1. Péril mental et apprentissage des langues
 - 2.2.2. Des débuts aux résultats mitigés
 - 2.2.3. Le renouvellement de l'intérêt pour l'apprentissage précoce dans un monde en pleine mutation
- 2.3. L'apprentissage précoce en France
 - 2.3.1. 1989 : « Enseignement d'Initiation aux Langues Vivantes » (EILE)
 - 2.3.2. 1995 : « Initiation aux Langues Vivantes » (ILV)
 - 2.3.2.1. Présentation de la réforme : les textes officiels
 - 2.3.2.2. La mise en pratique
 - 2.3.3. 2000 : Plan « langue vivante à l'école primaire »

Bibliographie du chapitre IV

L'âge joue un rôle déterminant dans les processus d'acquisition et si, pour une raison ou pour une autre, l'enfant était privé de stimulation linguistique pendant une période « critique » qui s'étend en gros jusqu'à l'âge de dix ans, la mise en place de la LM serait gravement perturbée. Cette période est également déterminante dans les processus d'apprentissage d'une langue étrangère, notamment en ce qui concerne sa prononciation. En effet, il apparaît qu'après dix ans, l'apprentissage du système intono-accentuel d'une langue seconde devient beaucoup plus difficile en raison de la perte partielle de la plasticité cérébrale. Par ailleurs, comme le soulignent Bertoncini et De Boysson-Bardies (2000 : 106) :

« compte tenu de l'antériorité de la prosodie sur les autres dimensions du langage et de la façon dont les processus se spécialisent au cours de l'acquisition, une des difficultés pour l'apprentissage tardif d'une langue seconde pourrait provenir de la capacité décroissante du système à utiliser la prosodie en priorité, avant de porter attention au niveau segmental »
(Bertoncini et De Boysson-Bardies, 2000 : 106).

Le niveau segmental pose également des difficultés à l'enfant, puisque ses réalisations phonétiques se détériorent avec le temps et son articulation devient « *pesante* » (Petit, 1985). Il est donc fondamental de commencer l'apprentissage d'une langue seconde avant ce « *seuil fatidique* » (Hagège, 1996). Mais une fois cette limite posée, à quel moment précis l'introduction d'une langue étrangère sera-t-il le plus adapté en fonction du développement de l'enfant ? Au Cours Moyen, au Cours Élémentaire ou à la maternelle ? Pour le savoir, nous ferons un bilan des différentes périodes développementales que traverse l'enfant jusqu'à ce qu'il atteigne ses dix ans et nous tenterons de déterminer un âge « idéal » pour l'apprentissage d'une langue étrangère. Mais est-ce que cet âge est en adéquation avec l'âge d'introduction d'une nouvelle langue à l'école ? Après avoir évoqué rapidement l'historique de l'apprentissage précoce, nous détaillerons les dernières réformes qui sont apparues en France depuis la fin des années quatre-vingt pour aboutir à la situation actuelle. Ces données nous aideront à déterminer l'âge de notre population pour l'étude du transfert de compétence de la musique à la langue.

1. Age et apprentissage

1.1. L'existence d'une période critique pour l'apprentissage d'une LE

1.1.1. La découverte d'une période sensible pour l'acquisition de la LM...

Pendant la période de l'acquisition, c'est moins la durée d'exposition à une langue qui est important, mais bel et bien l'âge auquel commence cette exposition. Il existe en effet une période favorable où les capacités de l'enfant sont maximales pour l'acquisition de la LM. Cette acquisition se fait, comme nous l'avons vu dans les deux précédents chapitres, de manière naturelle, par imprégnation progressive. Mais si l'enfant n'est pas exposé à sa LM pendant cette période, l'acquisition du langage sera fortement entravée. Le retard sera irréversible et ne pourra pas être comblé, même si l'enfant est exposé massivement à une langue (cas des enfants sauvages et de certains sourds profonds). Pour cette raison, on parle de « **période critique** ». D'après Bijeljac-Babic (2000 : 183) :

« dans l'acquisition du langage en général, on définit la « période critique » pour les capacités langagières comme la période pendant laquelle le système est doté de la plus grande sensibilité à développer des échanges spécifiques avec l'environnement » (Bijeljac-Babic, 2000 : 183).

L'existence d'une période critique a été notamment mise en lumière par les travaux de Penfield et Roberts (1959). Ceux-ci ont montré que des enfants de moins de dix ans étaient capables de récupérer intégralement leurs fonctions langagières après un traumatisme cérébral, contrairement aux enfants plus âgés et aux adultes. Grâce à une malléabilité exceptionnelle provenant d'une surabondance de connexions neuronales (la fameuse « frange d'adaptabilité » de Changeux, 1983), le cerveau du jeune enfant peut se réorganiser en cas de lésion. Après un accident touchant l'hémisphère gauche, l'enfant peut ainsi récupérer l'intégralité de ses capacités de langage car une nouvelle zone cérébrale prend le relais de la zone endommagée. Ce processus de réorganisation intervient également chez un adulte, mais, contrairement à l'enfant, l'adulte recouvre plus difficilement un langage normal, et surtout avec un délai plus long, car s'il est démontré maintenant qu'il y a adaptabilité et plasticité (avec neurogénèse) même durant la vie adulte, celle-ci reste moindre que dans un mécanisme jeune (cf. Ecole d'Été de Neurosciences organisée par le Collège de France en août 2002 sur le thème « Plasticité neuronale et adaptation fonctionnelle »).

1.1.2. ...étendue à l'apprentissage des langues secondes

S'il existe une période critique propice à l'acquisition de la LM, il est probable que cette période soit également particulièrement adaptée à l'apprentissage d'une langue étrangère, comme le suggère Penfield, cité par Petit (1985 : 108) :

« Souvenez-vous que pour l'apprentissage des langues, le cerveau humain devient progressivement raide et rigide après l'âge de neuf ans ».

Lenneberg (1967), s'inspirant des travaux de Penfield et Roberts, pense qu'il existe effectivement une période privilégiée pour l'apprentissage d'une L2, l'« **âge critique** ». Passé cet âge, c'est-à-dire à la puberté, l'apprentissage d'une L2 devient laborieux et un déclin linguistique s'amorce : les réalisations se détériorent en raison d'une certaine perte de plasticité des contrôles moteurs impliqués dans la production des phonèmes (Moreau et Richelle, 1981). Johnson et Newport (1989, 1991) situent ce déclin de manière plus précoce, entre l'âge de 6 et 8 ans.

1.2. A quel âge commencer l'apprentissage d'une LE ?

Pendant la période critique, l'enfant apprend une L2 de manière « intuitive », c'est-à-dire qu'il va progressivement intérioriser les règles de fonctionnement de cette langue. Cet apprentissage se fait au contact de la langue et de manière implicite ; une fois les règles intériorisées, l'enfant est capable de se corriger, sur la base d'une intuition grammaticale. Cette forme d'« apprentissage intériorisé » et l'intuition grammaticale qui en résulte offrent des points de similitude avec l'acquisition de la LM. Les processus d'apprentissage étant non conscients, on parle d'« **acquisition naturelle** » (Krashen, 1983). C'est donc une période intéressante pour étudier le transfert de compétence de la musique à la langue. Cependant, cela ne veut pas dire qu'il suffit de placer un enfant au contact d'une langue pour qu'il la parle comme un vrai bilingue (erreur qui a été commise dans les années 60, voir historique de l'apprentissage précoce). Après la période critique, l'enfant ne pourra plus apprendre les règles de la L2 de manière implicite, mais il devra mettre en jeu des mécanismes conscients d'apprentissage. C'est donc pendant cette période d'acquisition naturelle, que les aptitudes de l'enfant pour apprendre une L2 devraient être maximales, le cas idéal étant celui des enfants né dans un milieu bilingue. Mais l'école ne peut pas reproduire le développement naturel d'un enfant placé dans un tel environnement.

L'exposition ne serait pas assez massive et surtout, l'école s'adresse à des enfants plus âgés : il est difficile de commencer cet apprentissage avant l'école maternelle, tous les enfants n'étant pas concernés. Si, comme le souligne Garabédian (1996), le plus tôt serait le mieux afin de mettre à profit les grandes capacités perceptives de l'enfant, quand commencer l'apprentissage d'une L2 exactement ? Peut-on vraiment acquérir une seconde langue dans le strict cadre scolaire sans complémentarité familiale ?

1.2.1. Pendant « l'âge heureux¹ » ?

Pour Guberina (1991), c'est pendant l'« âge heureux » que l'enfant fait preuve des plus grandes facultés pour l'apprentissage d'une LE, c'est-à-dire jusqu'à six ans et notamment, pendant la période qui précède l'école primaire. Nous avons vu dans le chapitre III (3.1.), qu'entre l'âge de 4 et 6 ans, l'enfant passe progressivement d'une stratégie de traitement global à une stratégie de traitement local, qui sera définitivement en place vers 6-7 ans. Mais pendant cette phase de transition, c'est encore le traitement global qui domine, et notamment celui de la prosodie (domination du traitement par contour). L'enfant se montre particulièrement habile à restituer les éléments musicaux de la langue, en particulier ceux d'une nouvelle langue, et ce d'autant plus que ses capacités d'imitation sont maximales entre l'âge de 4 et 8 ans. L'influence de la L.M. se renforce de plus en plus, notamment pendant l'année de CP, mais pas suffisamment pour avoir eu le temps de fermer totalement l'oreille de l'enfant. Avant de se plier aux situations institutionnelles d'apprentissage, l'enfant manifeste encore une grande spontanéité et une grande curiosité qui lui permettent d'aborder une LE de manière « naturelle » et ludique. Cependant, si les facultés d'imitation de l'enfant sont exceptionnelles, il oublie tout aussi rapidement ce qu'il a appris rapidement. Les travaux de Pallier et al. (à paraître) montrent que des adultes ayant été adoptés entre les âges de 3 à 10 ans et pour la plupart d'origine coréenne, ont totalement oublié leur langue maternelle. Une étude d'IRMf révèle en effet que pour les sujets coréens, il n'y a pas de différence d'activation des aires corticales pour des phrases de leur LM que pour des phrases d'autres langues étrangères. La LM a donc été totalement remplacée par la langue seconde et les tests comportementaux ne montrent pas de différence entre ces sujets et les sujets contrôles français.

¹ Terme emprunté à Guberina (1991).

En ce qui concerne l'apprentissage d'une langue seconde, jusqu'à 6 ans, l'enfant possède une mémoire immédiate restreinte et Girard (1995) rappelle que

« les enfants, malgré leurs grandes capacités d'apprentissage ont une grande faculté d'oubli qui se manifeste dès que le processus est interrompu » (cf. problèmes de continuité dans l'apprentissage des langues).

Si la période qui s'étend jusqu'à l'âge de 6 ans se montre particulièrement « heureuse » pour l'introduction d'une LE, les auteurs sont partagés sur l'âge exact de sa mise en œuvre. Certains pensent qu'il faut commencer cet apprentissage au CP. Pour Hagège (1996 : ?) :

« c'est alors que la disponibilité totale du jeune écolier, aux oreilles et au regard déclois sur l'univers, doit être exploitée comme un trésor aussi précieux que périssable. C'est alors que les germes de la langue que l'on sème profitent d'une terre avide. Ils y seront accueillis avec d'autant de zèle qu'on n'en retrouvera jamais par la suite ».

D'autres pensent qu'il faut commencer dès la maternelle comme Guberina (1991), Garabédian (1996) et Cohen (1997). Cette dernière lance un véritable plaidoyer en faveur d'un enseignement des langues dès la maternelle. Quant à Garabédian (1996 : 31), elle écrit :

« ce n'est pas par goût de la provocation que je dis « non pas à six »... - qui est encore une période charnière au plan institutionnel, difficile à gérer pour les enseignants, les parents et les enfants principaux destinataires et acteurs – « mais à trois-quatre ans » ».

Il est vrai que l'année du CP est une année délicate à gérer pour les enseignants, car le programme est chargé, notamment à cause de l'introduction de la lecture et de l'écriture, comme nous l'avons déjà signalé à plusieurs reprises.

1.2.2. Pendant l'« âge critique² » (7-9 ans) ?

A partir de 6/7 ans, l'enfant bascule définitivement vers un traitement de type analytique, qui lui permet de se spécialiser sur sa LM. Cette stratégie requiert beaucoup d'attention et de contrôle, ce qui expliquerait les pertes de discrimination relevées à cet âge.

² Terme emprunté à Lenneberg (1967).

Cependant, les influences de la LM sont encore trop récentes pour avoir définitivement altéré les capacités auditives de l'enfant ; les pertes constatées concernent les contrastes robustes NN, c'est-à-dire des contrastes qui sont saillants au niveau psycho-acoustique et facilement réappris à l'âge de 8 ans, soit de manière spontanée, soit après un léger entraînement auditif. Cependant, l'enfant fait preuve d'un développement cognitif plus avancé qui pourrait venir compenser les difficultés qu'il rencontre au niveau perceptif. Ainsi, on constate un bond de la compréhension entre le CE2 et le CM1 (Konopczynski, 1981a) et l'enfant manifeste de meilleures connaissances des caractéristiques linguistiques de sa LM. Il est passé d'un savoir intuitif à un apprentissage explicite des règles de grammaire, ce qui contribue à développer ses capacités métalinguistiques. Quant à son système sémantique, il est mieux organisé. Dans une expérience d'enseignement précoce de l'allemand qui a concerné 3600 élèves du CE2 au CM2, Petit (1985) a relevé que les enfants de CE2 manifestaient d'extraordinaires capacités phoniques, justifiant le bien fondé d'une approche exclusivement orale de la langue étrangère. Cependant, à 8 ans, une certaine « *pesanteur* » articulatoire s'installe (Petit, 1985), directement issue des habitudes articulatoires spécifiques à la LM.

1.2.3. Pendant le « seuil fatidique³ » ?

Lorsque l'enfant atteint le « seuil fatidique » (Hagège, 1996 : 27) à la fin de la période critique, la restitution d'une LE devient difficile au niveau phonétique. Selon Petit (1985 : 360),

« l'âge de 10 ans constitue le jalon à partir duquel s'amorce un processus de décroissance rapide et irréversible des possibilités d'acquisition phonique d'une L2 ».

Les habitudes articulatoires de la LM sont en effet complètement établies et la « *pesanteur articulatoire* », apparue dès 8 ans (Petit, 1985), continue à se renforcer. Par ailleurs, les capacités d'imitation de l'enfant ont décliné. Au niveau de la perception, le poids grandissant de la LM a réduit la richesse perceptive de l'enfant en centrant davantage son attention sur les catégories natives. Les phénomènes de crible prosodique et phonologique n'ont jamais été aussi puissants chez l'enfant de 10 ans et leurs effets se rapprochent de ce qui se passe pour un auditeur adulte.

³ Terme emprunté à Hagège (1996 : 27).

L'enfant rencontre de grandes difficultés à se corriger, car il n'entend plus correctement les sons de la LE. Selon Hagège (1996 : 34-35) :

« au-delà de 10 ou 11 ans, la fossilisation des aptitudes non stimulées n'est guère réversible. En effet, vers cet âge, l'oreille, jusque-là organe normal d'audition, devient nationale. On peut appeler oreille nationale (Dalgalian et al., 1981), celle qui, au lieu de traiter de la même façon la totalité des sons perçus en leur ouvrant un itinéraire jusqu'à l'enregistrement cérébral, fonctionne comme un filtre n'ouvrant passage qu'à ceux que la LM connaît sous une forme identique ou analogue [...] Il est une des causes principales de l'accent étranger ».

1.2.4. Conclusion

L'enfant possède de grandes capacités pour l'apprentissage d'une langue jusqu'à l'âge de 7/8 ans. La première période qui s'étend jusqu'à 6 ans paraît particulièrement appropriée pour l'apprentissage phonétique. Jusqu'à 8 ans, ses capacités phonétiques sont encore bonnes, bien que le poids de l'environnement commence à se faire sentir. L'enfant fait preuve de capacités linguistiques plus développées qu'aux âges précédents, qui pourraient être profitables à l'apprentissage d'une LE. **Il est idéal de commencer l'apprentissage d'une langue pendant l'âge heureux et il est encore temps pendant l'âge critique.** Au-delà, on perd les avantages d'un apprentissage précoce pour se rapprocher de la situation d'apprentissage traditionnel qui intervient trop tardivement car les habitudes maternelles sont trop lourdes. Pendant la période propice, **un apprentissage fondé sur l'oral paraît le plus approprié** à cause des capacités de l'enfant. C'est de toute façon un ordre plus naturel, et plus proche de l'acquisition de la langue maternelle. Malheureusement, pendant longtemps, l'apprentissage de la première langue étrangère a commencé à l'entrée au collège, c'est-à-dire à 11 ans, au moment du seuil fatidique, ce qui est en totale contradiction avec le développement de l'enfant et l'évolution de ses capacités.

2. Etat des lieux de l'apprentissage précoce

A partir de quand a-t-on commencé à s'intéresser à l'apprentissage précoce des langues ? En effet, celui-ci n'a pas toujours été considéré comme bénéfique ; bien au contraire, on pensait qu'il était à l'origine de retards importants dans le développement de l'enfant. Beaucoup d'expériences ont été menées dans les différents pays, avec des résultats très inégaux et des conclusions très diverses.

Par ailleurs, les sociétés monolingues ont longtemps considéré le multilinguisme comme une situation si ce n'est « exceptionnelle », en tout cas minoritaire et spécifique à un certain type de population, la population immigrée. Pourtant, c'est le monolinguisme qui reste une situation minoritaire à l'échelle mondiale. Depuis les années 60 où les expérimentations ont véritablement commencé, il y a eu beaucoup de tâtonnements, d'avancées et de retours en arrière avant d'aboutir à notre conception actuelle de l'apprentissage précoce. De plus, la signification de ce type d'apprentissage devrait évoluer en fonction du développement de la circulation des personnes et de la multiplication des échanges linguistiques qui caractérisent notre société contemporaine.

2.1. Qu'est-ce qu'un apprentissage précoce ?

Le terme « précoce » qualifie un être vivant qui atteint la maturité avant les autres individus. On parle d'« enfants précoces », c'est-à-dire des enfants qui ont développé des capacités intellectuelles avant la majorité des autres enfants de leur âge. Par extension, « précoce » a désigné un développement intellectuel supérieur (enfants surdoués). Mais, ce qui est important dans ce terme, c'est sa signification chronologique, puisqu'il qualifie ce qui se produit avant. Appliqué aux langues, il désigne non pas un apprentissage meilleur, mais un apprentissage qui intervient plus tôt que ce qui se fait habituellement. Dans le domaine scolaire, il correspond à un apprentissage qui se met en place avant celui qui est fixé par les institutions. Selon Hagège (1996 : 79-80)

« il ne s'agit pas ici d'un sens lié à la personnalité biologique ou intellectuelle de l'enfant, mais d'un sens institutionnel [...] La notion de précocité s'attache donc ici à ce qui est pratiqué avant l'âge prévu par l'institution ».

Puisqu'en matière d'apprentissage des langues, le terme « précoce » dépend de l'institution scolaire, que veut dire apprentissage précoce en France ? C'est une problématique très récente. Jusqu'à la rentrée de 1989 et la mise en place de la réforme « Enseignement d'Initiation aux Langues Vivantes » (désormais EILE), l'enseignement d'une langue étrangère commençait officiellement au collège, en sixième. Cependant, à partir de 1989, il a été avancé au cours moyen avec l'EILE, puis au cours élémentaire à partir de 1995 avec l'« Initiation aux Langues Vivantes » (désormais ILV). Cet enseignement n'était toutefois pas généralisé à toutes les écoles du territoire (il se faisait sur la base du volontariat) et restait au niveau d'une simple initiation.

Cependant, depuis 1998, nous sommes entrés dans une autre logique, une logique de généralisation qui devrait aboutir en 2005. A cette date, tous les enfants de 5 ans devraient commencer l'apprentissage d'une langue vivante, c'est-à-dire à partir de la grande section de maternelle. En conséquence, le terme d'apprentissage précoce ne sera plus approprié, puisque nous sommes passés progressivement du temps de l'expérimentation, où il ne concernait qu'un nombre limité d'établissements, au temps de la généralisation où tous les établissements devraient être concernés. Avec le plan « Langues Vivantes à l'Ecole Primaire » (2000), l'enseignement des langues est aujourd'hui intégré au cursus scolaire dès l'école élémentaire et la pratique des langues est considérée comme une discipline à part entière, inscrite dans les programmes.

Certains auteurs ont discuté de la pertinence du terme « précoce » pour désigner ce type d'apprentissage. Maier (1996 : 59) rejette le terme d'« enseignement précoce des langues vivantes » tel qu'il est utilisé en Allemagne, car selon lui, il maintient l'idée que l'enseignement des langues vivantes dans le secondaire serait la norme. Pour ces raisons, il préfère parler d'« *enseignement primaire des langues vivantes* ». En effet, ce terme lui paraît plus approprié pour :

« désigner [...] l'identité particulière d'un enseignement dont les méthodes, les contenus et les objectifs sont adaptés à l'enfant et qui s'inscrit dans la continuité du processus d'apprentissage dans la carrière scolaire ultérieure » (Maier, 1996 : 59).

2.2. Bref historique de l'apprentissage précoce

2.2.1. Péril mental et apprentissage des langues

Jusque dans les années 60, l'apprentissage précoce et l'enseignement bilingue des langues étrangères ont souffert de graves préjugés. Certains auteurs pensaient que l'apprentissage trop précoce d'une langue se faisait forcément au détriment de la LM et qu'il se montrait même dangereux pour le développement de l'enfant, provoquant des retards au niveau linguistique et même des retards mentaux plus ou moins marqués.

Petit (1985 : 109) cite l'un de ces auteurs, qui se livre à une condamnation sans appel du bilinguisme :

« Le bilinguisme est une infériorité intellectuelle ; cette conclusion pessimiste est celle de tous les auteurs qui ont, sur du matériel concret, étudié la question [...] Cette nocivité du bilinguisme est explicable ; car d'une part, l'effort demandé pour l'acquisition de la langue seconde semble diminuer la quantité disponible d'énergie intellectuelle pour l'acquisition d'autre connaissances ; d'autre part et surtout, l'enfant se trouve ballotté entre des systèmes de pensée différents l'un de l'autre ; son esprit ne trouve d'assiette ni dans l'un ni dans l'autre, et il les adúltere tous les deux en les privant de leur originalité et en se privant par là lui-même des ressources accumulées depuis des siècles par ses prédécesseurs dans chaque idiome » (Pichon, 1936 : 101-103).

Ces arguments, qui nous paraissent complètement fantaisistes aujourd'hui, connurent leur apogée dans les années 20-30 et se sont propagés jusque dans les années 60. Or, il a été montré par la suite que l'apprentissage précoce et le bilinguisme n'étaient absolument pas nocifs pour l'enfant et que non seulement ils ne provoquaient aucun retard mental, mais que les petits bilingues possédaient même une malléabilité et une souplesse cognitive supérieures à celles des monolingues.

2.2.2. Des débuts aux résultats mitigés

Ce sont les travaux de Wilder et Penfield (1959) qui contribuèrent à marquer un tournant décisif en révélant les extraordinaires capacités de récupération dont faisaient preuve les enfants de 10 ans (voir paragraphe 1.1.). Une fois ces conclusions étendues à l'apprentissage des langues, l'enthousiasme pour l'enseignement bilingue et l'apprentissage précoce gagnât beaucoup de pays en Europe, comme en témoignent les deux conférences internationales organisées par l'Unesco en 1962 et en 1966. Un certain nombre d'expérimentations d'apprentissage bilingue furent lancées, mais leurs conclusions étaient contradictoires. Alors que les programmes d'immersion canadiens faisaient apparaître de bons résultats, il en était tout autrement pour les programmes lancés aux Etats-Unis, en Scandinavie et en Australie. La dimension sociale n'avait pas été suffisamment prise en compte et les évaluations se montrèrent négatives pour les enfants issus des populations immigrées. **L'environnement social avait été largement sous-estimé**, alors qu'il expliquait en grande partie les mauvais résultats obtenus.

Par ailleurs, les capacités d'apprentissage de l'enfant avaient été surestimées dans le sens où **il ne suffit pas de le mettre au contact d'une langue étrangère pour qu'il devienne bilingue** :

« l'erreur souvent commise consistait à croire qu'il suffisait de commencer l'enseignement d'une deuxième langue avant l'âge critique de dix ans pour obtenir d'excellents résultats, et à oublier que pour être efficace, l'enseignement d'une langue vivante doit bénéficier de conditions favorables » (Girard, 1995).

L'analogie avec l'acquisition de la langue maternelle a ses limites et l'apprentissage des langues secondes à destination d'un public très jeune nécessite une pédagogie spécifique. Porcher et Groux (1998 : 49-51) expliquent qu'en France, l'apprentissage précoce est apparu dès la fin de la seconde guerre mondiale, avec l'introduction de l'anglais à l'école maternelle. Beaucoup d'expériences ont ensuite été menées et cet enseignement n'a pas cessé de se développer jusqu'en 1973. A cette date, la circulaire du 11 mai 1973 a contraint les maîtres titulaires des classes à se charger de l'enseignement des langues vivantes, celui-ci étant auparavant assuré par des instituteurs dont la LM était la langue enseignée. Dans la plupart des cas, les maîtres n'étant pas formés, les écoles durent abandonner l'apprentissage précoce. Selon Favard (1992 : 31), les tentatives étaient éparées et ne furent jamais coordonnées au plan national.

Dans l'Est de la France, région pour laquelle nous disposons de quelques renseignements, la situation était assez typique du flou régnant au niveau national. En Alsace, l'enseignement de l'allemand s'est fait très tôt, sur la base du bénévolat, avec des maîtres qui, pour la plupart, étaient dialectophones (alsacien). En Franche-Comté, et plus spécialement dans l'Académie de Besançon, une forte motivation régnait dans le corps enseignant. Ainsi, selon les écoles, une initiation précoce s'est faite dès la crèche à la crèche municipale de Lons-le-Saulnier (39) où certaines puéricultrices anglophones ne parlant que l'anglais aux bébés (continuation en maternelle assurée) et dès la maternelle à l'école de l'Helvétie de Besançon. Dès le CE1, allemand (Ecole de Fontaine-Ecu) dans une expérimentation pilotée par Petit, professeur d'allemand à la Faculté des Lettres de Besançon, expérimentation dont on trouve un bilan dans Petit (1985).

Quoique les remarques qui suivent soient de type anecdotique, puisqu'elles n'ayant donné lieu ni à publication, ni à évaluation sérieuse, elles ne manquent pas d'intérêt. Elles m'ont été communiquées par Konopczynski, qui faisait parti du groupe de pilotage d'allemand. Cette expérience a été bien réalisée dans l'ensemble, car tous les enseignants volontaires savaient l'allemand au départ et ont bénéficié de stages dits de recyclage pris sur leur temps libre. La méthode employée, œuvre de Petit « *Sing und Spiele mit* » leur a été aussi présentée et explicitée par l'auteur lui-même. Les élèves étaient aussi motivés que les enseignants. Les deux enfants de Konopczynski et nombre d'enfants de ses amis ont pu bénéficier de cet enseignement précoce et elle rapporte qu'ils ont quasi tous acquis une prononciation quasi native tant au niveau segmental que prosodique, que la plupart d'entre eux se sont ensuite dirigés au collège vers des classes dites « d'allemand bilingue » où l'enseignement de la langue était renforcé, puis qu'ils ont choisi des sections de langues (A2 à l'époque) au lycée ; plusieurs ont embrassé des carrières où les langues, l'allemand notamment, leur sont indispensables. Certes, plus de la moitié des élèves concernés étaient issus de milieux socio-culturels favorisés et regroupés dans les classes dites « d'application » de l'École Normale de Besançon où la musique (chant choral) était également privilégiée. En outre, ces enfants bénéficiaient presque tous d'une formation musicale (solfège et instrument) au conservatoire. Cependant, ces classes comportaient aussi des enfants immigrés. Tous les enfants avaient été testés au début de C.P. pour leur niveau de conscience phonologique ; il est apparu que les enfants immigrés, en contact avec une autre langue que le français, avaient une conscience phonologique plus développée que leurs camarades monolingues. Leurs résultats en allemand furent très bons, contrairement aux résultats dans la plupart des autres matières. Cet enseignement n'ayant pas donné lieu à une véritable expérimentation calibrée et évaluée, on ne saurait en tirer de conclusions, plusieurs facteurs se combinant pour expliquer le bon niveau des jeunes apprenants. Cependant, il était intéressant de la signaler.

Un autre écueil des expériences d'enseignement précoce des années 70 fut la non prise en compte des acquis lors du passage au collège, où les élèves furent mélangés avec des débutants complets et l'enseignement repris à zéro, d'où la démotivation extrême pour les élèves déjà formés. Cette dernière remarque introduit un autre problème crucial pour la réussite de l'apprentissage précoce : pour qu'il soit vraiment efficace, il faut qu'il y ait un **suivi de cet enseignement entre le primaire et le secondaire.**

En effet, n'oublions pas que, si l'enfant fait preuve de grandes capacités d'assimilation, il est également capable d'oublier très vite. Girard (1995) qui avait réalisé en 1974 un rapport d'évaluation sur l'ensemble des expériences d'apprentissage précoce réalisées sur le territoire français, explique que ce problème de continuité constitue l'une des raisons principales qui ont fait échouer les premières expériences d'apprentissage précoce. Il met également en lumière le **problème de la formation des enseignants**, insuffisante pour leur permettre d'assurer ce type d'apprentissage, que ce soit au niveau pédagogique ou au niveau de leurs capacités linguistiques. Si on ajoute le **manque de matériel pédagogique** et un **taux horaire insuffisant**, leur tâche devient encore plus difficile :

« c'est ainsi que vers 1975, la grande vague d'intérêt était largement retombée dans la plupart des pays, pour plusieurs insuffisances constatées : manque d'enseignants correctement formés, de matériaux pédagogiques adéquats, horaires insuffisants et, par-dessus tout, absence de continuité et de suivi entre le primaire et le secondaire » (Girard, 1995 : ??).

2.2.3. Le renouvellement de l'intérêt pour l'apprentissage précoce dans un monde en pleine mutation

Devant des évaluations aux résultats contradictoires, l'enthousiasme pour l'apprentissage précoce des langues est progressivement retombé jusqu'à ce qu'il se réveille à nouveau au début des années 80 et devienne une préoccupation centrale des gouvernements. Porcher et Groux (1998) voient dans cet intérêt renouvelé de nouvelles préoccupations liées à **l'internationalisation des activités et des savoirs** (qualifications professionnelles qui nécessitent la pratique d'une ou de plusieurs LE ; multiplication des déplacements internationaux ; nous ajouterons la suppression des frontières en Europe), ainsi qu'un **développement très important des médias**, contribuant à réduire la planète à un « village planétaire » (Mac Luhan), où il sera nécessaire pour pouvoir communiquer avec son « voisin » de pratiquer sa langue ou une langue véhiculaire commune :

« l'internationalisation (qui chez nous, se manifeste institutionnellement par la dimension européenne) et les médias constituent à coup sûr les éléments lourds de l'évolution qui rend aujourd'hui indispensable l'enseignement précoce des langues vivantes, au plus jeune âge possible » (Porcher et Groux, 1998 : 30).

Ce n'est certainement pas un hasard si en 1991, pour la première fois, le projet « Langues Vivantes » du Conseil de l'Europe inscrivait l'enseignement et l'apprentissage des langues à l'école primaire parmi ses priorités et que

« vingt pays membres aient envoyé des participants à l'atelier organisé sur ce thème à Edimbourg en 1991 et à celui que la France a organisé à Sèvres en décembre 1993 » (Girard, 1995).

Cependant, en 1992, le Figaro titrait « *Langues étrangères : la cacophonie* », car les pays de la communauté hésitaient entre expérimentation et généralisation de l'apprentissage précoce. En 1996, Hagège (1996 : 79) faisait le constat que l'« *offre scolaire* » sous-exploite les capacités du jeune enfant en France. Est-ce que cette situation a évolué durant ces six dernières années ?

En Europe, s'il reste encore des pays où l'apprentissage obligatoire d'une langue étrangère ne commence pas avant l'âge de 9-10 ans (Allemagne, Danemark, Grèce, Pays-Bas, Portugal et Royaume-Unis⁴), il est des pays où cet apprentissage commence à 7 ou 8 ans (Belgique, Espagne, Finlande, Italie et Suède⁴) et d'autres encore où il commence pendant l'âge heureux, à 6 ans (Autriche, Luxembourg⁴). Pour le Luxembourg, l'apprentissage précoce est devenu une véritable priorité nationale, avec l'instauration d'un trilinguisme (luxembourgeois/allemand/français). L'Autriche, l'Italie et l'Espagne⁴ se démarquent par des projets pilotes où les enfants débute une nouvelle langue dès l'âge de 3 ans.

Comment la France se situe-t-elle par rapport à ces pays ? Depuis 1989, dans le cadre d'une expérimentation contrôlée, l'enseignement commence au CM1, puis il a été étendu à la rentrée de 1995 au CE1, mais ce n'est encore qu'une sensibilisation avant l'apprentissage qui commence en sixième et il ne concerne que des classes volontaires. Cependant, depuis la rentrée de 1998, on assiste à un véritable effort de généralisation à partir du CM2 (10 ans), puis à la rentrée de 1999, à partir du CM1 (9 ans). Mais c'est en 2000 que la France se donne les moyens d'une réforme qui va lui permettre de rattraper progressivement le retard accumulé en instaurant d'ici 2005 l'apprentissage généralisé des langues à partir de la grande section de maternelle (5 ans).

⁴ Source : Eurydice (juillet 2000), « la place des langues étrangères dans les systèmes éducatifs européens ».

2.3. L'apprentissage précoce en France

2.3.1. 1989 : « Enseignement d'Initiation aux Langues Vivantes »

En 1989, le ministère de l'Education Nationale lance dans les classes de CM1 et de CM2, une « *expérimentation contrôlée d'enseignement des langues à l'école élémentaire* », l'EILE, dont l'organisation est définie par la circulaire du 16 mars 1989 (complétée par la circulaire du 14 juin 1989). Cet enseignement prend la forme d'une **sensibilisation** à la langue étrangère, c'est-à-dire que la pédagogie utilisée passe par le jeu et donne la prédominance à l'oral (circulaire du 6 septembre 1991). Dans ce type d'enseignement, il n'y a pas d'apprentissage explicite : l'enfant s'approprie la langue par simple exposition à cette langue, de manière implicite et naturelle. Cette sensibilisation le prépare à l'apprentissage qui ne commencera toutefois pas avant le collège. Les séances de langue ont lieu deux à trois fois par semaine, pendant une durée d'une heure. Les instituteurs qui ont choisi de participer à l'expérimentation ont le choix entre trois langues : l'allemand, l'anglais et l'espagnol. Pendant l'année de sa mise place (année scolaire 1989-1990), l'EILE a concerné 7,5 % des élèves de CM1 et 12,5% des élèves de CM2.

En 1990, devant le succès de l'expérimentation initiale, la nature de l'enseignement est modifiée. La circulaire du 26 mars fixe en effet le passage d'un enseignement de sensibilisation à un enseignement d'initiation, qui vise déjà une production orale :

« Cet enseignement doit consister essentiellement en une initiation à la LE destinée à faciliter son apprentissage en classe de sixième ».

Au terme de l'expérimentation, une évaluation a mis en lumière les points positifs de cet enseignement (Favard, 1992) : tout d'abord, la participation des enseignants a été enthousiaste et efficace et les approches pédagogiques, riches et variées. En ce qui concerne les enfants, ils ont manifesté un réel plaisir d'apprendre. Il apparaît cependant que les enseignants manquent de supports pédagogiques et qu'il faudrait qu'ils soient réellement formés à l'enseignement des langues étrangères. Par ailleurs, les bénéfices de l'ILV semblent se situer plus au niveau de l'enrichissement personnel apporté à l'enfant, que dans le développement d'une véritable compétence linguistique.

2.3.2. 1995 : « Initiation aux Langues Vivantes »

2.3.2.1. Présentation de la réforme : les textes officiels

A la rentrée scolaire 1995, dans le cadre du Nouveau Contrat pour l'Ecole, une nouvelle réforme appelée ILV instaure une première sensibilisation aux langues vivantes dès le CE1 (circulaires n°95-103 et n° 95-104 du 3 mai 1995). Cette réforme prend appui sur l'EILE dont elle ne remet pas en cause le dispositif. Les langues concernées sont l'allemand, l'anglais, l'arabe, l'espagnol, l'italien et le portugais. En même temps que l'ILV est mis en place un dispositif d'observation et de suivi, ainsi que des indications pédagogiques précises sur les contenus. Les séances sont quotidiennes et durent chacune quinze minutes. La méthodologie est laissée au choix de l'enseignant, mais des cassettes vidéo ont été créées spécialement et distribuées à tous les enseignants volontaires pour leur apporter un support pédagogique. La série « *CE1 et CE2 sans frontières* » réalisée par le Centre National de Documentation Pédagogique (CNDP) était même censée dans la logique initiale remplacer la compétence linguistique de l'enseignant. Autrement dit, grâce à ces cassettes, un instituteur pouvait théoriquement assurer un enseignement d'initiation sans réelle compétence linguistique et pédagogique dans la langue étrangère enseignée.

En continuité avec l'EILE, l'ILV est particulièrement intéressante parce que cet enseignement accorde la priorité à l'oral du point de vue de la perception, de la compréhension et de la production ; il privilégie en effet d'abord :

« l'éducation de l'oreille et le développement de l'aptitude à la compréhension orale. L'éducation de l'oreille passera par l'écoute attentive et sélective, l'imitation, la mémorisation et la production de structures langagières simples » (circulaire n°95.104 du 3 mai 1995).

Le Ministère a fixé la liste précise des compétences langagières de base et des fonctions globales à aborder, langue par langue (circulaire n°95.104 du 3 mai 1995). Il est intéressant de relever dans cette même circulaire des indications qui concernent l'apprentissage de la prosodie de la langue cible : en matière de compréhension globale, l'enfant devra être entraîné à l'identification :

« de l'intonation correspondant aux différents types d'énoncés (énoncés déclaratifs, interrogatifs, injonctifs, exclamatifs) [...] et à l'identification de groupes de souffles. Il devra percevoir l'intonation et les phonèmes voisins » (circulaire n°19).

La production vient ensuite, l'enfant devant être capable de

« reproduire un énoncé en respectant le rythme et le schéma d'intonation (comptine, chanson, poème...) » (circulaire n°19).

Dans les guides pédagogiques accompagnant les cassettes vidéo *« CE1, CE2 sans frontières »*, les rubriques consacrées à la prosodie et à la phonétique sont par ailleurs très détaillées (Dodane, 2000). Elles apportent notamment des recommandations à l'enseignant pour aborder la prononciation des phonèmes difficiles, ainsi que l'accentuation et proposent des chansons et des jeux qui vont permettre de travailler ces difficultés spécifiques.

La rentrée 1998-1999 constitue une étape essentielle dans l'ILV, puisqu'elle marque le passage d'une phase encore largement consacrée à l'expérimentation, à une véritable volonté de généralisation de cet enseignement à toutes les classes du cours moyen (voir tableau n°4, circulaire n°98-105 du 11 mai 1998). Par ailleurs, la circulaire 99-093 du 17 juin 1999 insiste sur le fait que l'enseignement en langue délivré en CM2 doit se démarquer de l'initiation pour constituer un véritable apprentissage et qu'il doit être assuré par des enseignants compétents :

« l'enseignement ne peut être confié qu'à des personnels linguistes : enseignants du premier degré habilités, enseignants du second degré, assistants étrangers et intervenants extérieurs agréés ».

Le tableau n°4 présente un résumé des différentes étapes de la mise en place de l'ILV et les références des textes officiels associés.

Rentrée concernée	Textes	Classe concernée	Contenu
1995/1996	n°95-103 du et n°95-104 du 3/05/1995	ILV –CE1	Première sensibilisation aux langues vivantes dès le CE1.
1996-1997	N°96-131 du 9/05/1996	ILV –CE2	Mise en place de la deuxième année d'initiation aux langues vivantes au CE2.
1997-1998	N°97-102 du 24/04/1997	ILV – CM1	Mise en place de la troisième année d'initiation aux langues vivantes au CM1.
1998-1999	N°98-105 du 11/05/1998	ILV – CM2	Mise en place de la deuxième année d'initiation aux langues vivantes au CM2 (généralisation et de manière plus limitée au CM1).
1999-2000	N°99-093 du 17/06/1999	ILV	Priorité à l'achèvement de la généralisation (CM2) ; une heure 30 par semaine, développement enseignement en CM1 ; ne remet pas en cause initiation au CE et dans les classes de CM1.

Tableau n°4 : textes officiels qui concernent la mise en place de l'ILV en France.

2.3.2.2. La mise en pratique

Mais qu'en est-il au niveau de la pratique ? Les rapports établis par les experts du Ministère en juin 1997 se montrent très positifs. Il apparaît qu'à l'entrée au collège, les élèves manifestent une avance de 6 mois sur le plan cognitif. L'apprentissage précoce des langues se révèle donc bénéfique pour le développement de l'enfant. Cependant, un certain nombre de problèmes n'ont pas été résolus par l'ILV. Le premier d'entre eux concerne la compétence linguistique des enseignants. Dans un premier temps, les cassettes vidéo réalisées par le CNDP étaient censées remplacer cette compétence lorsqu'elle n'était pas présente. Mais, il est vite apparu qu'elles ne pouvaient constituer qu'un complément et en aucun cas, le seul et unique support du travail de la classe. Par ailleurs, leur contenu pédagogique a souvent été jugé comme décevant par les enseignants (manque de réalisme, manque d'ancrage dans la réalité du pays dans lequel est parlée la langue, dialogues alourdis par l'utilisation d'un vocabulaire programmé par l'inspection générale...). Des efforts ont-ils été entrepris pour améliorer la formation des professeurs ?

Les professeurs des écoles fraîchement nommés ont pour la plupart obtenu une habilitation pour ce type d'enseignement pendant leur formation à l'Institut de Formation Universitaire des Maîtres (désormais IUFM), mais la généralisation prévue à partir de 1999 nécessite un contingent de personnes qualifiées bien supérieur. Ainsi, à la rentrée 1999, l'enseignement en langue était encore assuré dans 42 % des classes par des instituteurs ou des professeurs des écoles volontaires, alors que dans 17,2% des classes, il était assuré par des instituteurs et des professeurs habilités (par la commission académique d'habilitation en langues ou par l'IUFM), dans 16 % des classes par des professeurs de langue du second degré et dans 13,2% des classes par des intervenants extérieurs⁵. La circulaire n°99-093 du 17 juin 1999 a donné la possibilité aux assistants de langue de faire une partie ou la totalité de leur service à l'école primaire, ce qui explique que leur nombre soit en légère progression en 1999 (6,1% des classes)⁵. Ce problème de compétence linguistique est d'autant plus important que l'enseignant de langue constitue une véritable modèle identificatoire pour l'enfant (Bruner, 1983), qui doit jouer

« un rôle de médiation en mettant en place les conditions d'une écoute attentive, en gérant le rythme de la séquence et les moments d'interactivité avec le document audiovisuel, en veillant à la compréhension et à la reformulation correcte par les enfants grâce aux énoncés de référence proposés par le support vidéo » (BOEN du 3/05/1995, n°95-104).

Que se passe-t-il si ce modèle s'exprime avec un fort accent français ? Etant donné les capacités d'imitation maximales dont les enfants font preuve jusqu'à 8 ans, il est probable qu'ils fixent cet accent et que les bénéfices de l'apprentissage précoce s'en trouvent donc fortement réduits. Une expérimentation est en cours à l'Université de Franche-Comté, laboratoires de Phonétique et de Psychologie, pour élucider le problème (crédits du Ministère de la Recherche, dans le cadre du Programme Cognitique 2001 « École et Sciences Cognitives », thème « Apprentissage des Langues »).

Mais le plus gros problème de l'ILV tient directement à son caractère initial d'expérimentation, puisqu'au moment de sa mise en place en 1995, cet enseignement ne concernait pas tous les élèves. Or, il est impossible pour les professeurs de langue de collège de tenir compte du niveau des enfants formés lorsque ceux-ci sont mélangés à des enfants qui n'ont pas suivi d'apprentissage précoce.

⁵ Source : Note d'information 01-18, avril 2001, Ministère de l'Éducation Nationale (figure dans le Volume II, annexe 2).

En 1995, 63,2 % des élèves se trouvaient dans ce cas (Porcher et Groux, 1998 : 69). Si la continuité n'est pas assurée entre le primaire et le collège, les effets sont très négatifs sur les enfants, qui perdent une partie de leur motivation pour la LE et commencent à régresser. Selon Hagège :

« si la continuité n'est pas assurée de la formation à l'école à l'enseignement au collège, les profits tirés de l'apprentissage précoce sont en danger d'être annihilés ou fortement compromis » (Hagège, 1996 : 44).

Cependant, la volonté de généralisation affichée à partir de 1999 a commencé à porter ses fruits et à la rentrée scolaire de 1999, 82 % des élèves de CM2 et 60 % des élèves de CM1 bénéficiaient de l'apprentissage d'une langue vivante (mais seulement 36 % en CE1 et 42,1 % en CE2). Cette généralisation ne va pourtant malheureusement pas dans le sens de la diversité linguistique, puisqu'il contribue à renforcer la prédominance de l'anglais. Ainsi, à la rentrée de 1999, 79,8 % de la totalité des élèves pratiquant une langue étrangère à l'école primaire (public et privé confondus) apprenaient l'anglais, contre 15,2% l'allemand, 2,4 % l'espagnol et 1,3% l'italien⁵.

2.3.3. 2000 : Plan « Langues vivantes à l'école primaire »

Le plan « Langues vivantes à l'école primaire » marque une étape décisive vers un enseignement continu et généralisé des langues vivantes en France. Avec cette réforme, l'enseignement des langues vivantes change de statut et devient une véritable discipline, qui occupe une place dans les nouveaux programmes scolaires et à laquelle a été affecté un horaire déterminé. Le calendrier de la réforme commence par la généralisation de l'enseignement d'une langue en CM2 dès la rentrée 2000 (en fait, nous avons vu que cette généralisation était déjà bien engagée en 1999) et en CM1 à la rentrée suivante. A partir de 2002, cet enseignement sera généralisé au début de la grande section de maternelle pour remonter jusqu'au CE1 d'ici à 2004. La dernière étape du calendrier sera atteinte à la rentrée 2005, lorsque l'enseignement des langues vivantes sera généralisé de la grande section de maternelle jusqu'au CM2 inclus. Par ailleurs, la continuité de l'apprentissage de la première langue vivante devrait être assurée entre le CM2 et la sixième, et en 2005, tous les élèves de sixième commenceraient l'apprentissage d'une deuxième langue vivante. Parallèlement, une évaluation des acquis et notamment de la prononciation des élèves à la sortie de l'école élémentaire a été lancée. Elle sera généralisée à partir de 2004.

Ce projet de généralisation de l'enseignement des langues à partir de la grande section maternelle et le début d'une langue seconde en sixième sont des objectifs très ambitieux à atteindre, mais il semble que le plan « Langues vivantes à l'école primaire » prenne en compte les leçons des précédentes expérimentations. Ainsi, en ce qui concerne la **formation des maîtres**, « *cet enseignement sera confié à des maîtres ou à des intervenants dont la compétence linguistique est attestée* » (?). Dans une première phase, la réforme prévoit une mobilisation des maîtres ayant une compétence linguistique et notamment, les professeurs des collèges (la baisse des effectifs d'élèves au collège pour certaines langues libère en effet du temps d'enseignement). Mais à terme, l'enseignement des langues sera confié en majorité aux enseignants du premier degré et pour y parvenir, l'épreuve de langue deviendra obligatoire à partir de 2002 lors du concours de professeur des écoles et des formations initiales et continues en langues seront créées. Le plan « Langues vivantes » devrait également résoudre, nous l'avons vu, le **problème de la continuité** entre le primaire et le collège. Il conserve la priorité accordée à la langue orale déjà retenue par l'EILE et l'ILV. Par ailleurs, pour éviter la main mise de l'anglais sur l'enseignement primaire, il préconise la **diversification des langues** dès le primaire en apportant la garantie que l'enseignement de la première langue commencé au primaire pourra être continué en sixième et qu'il sera alors possible de **commencer l'apprentissage de l'anglais dès la sixième**. Par ailleurs, le choix disponible sera étendu à des **langues extra-européennes** telles que l'arabe et le chinois (reconnaissance des langues des pays d'immigration) :

« Cette diversification témoigne du double souci de préserver la richesse linguistique et culturelle de l'Europe et du monde et de former les enfants à une véritable citoyenneté européenne ».

Enfin, le plan prévoit une intégration plus large des ressources informatiques (échanges par courrier électronique avec des correspondants étrangers par exemple) et utilisation de cédéroms et logiciels de langues.

Le « Plan Langues vivantes à l'école primaire » s'inspire pour beaucoup du livre d'Hagège (1996), « *L'Enfant aux Deux Langues* », notamment pour certains principes tels que la l'instauration de l'anglais comme seconde langue pour lutter contre sa primauté dans l'enseignement, la diversification des langues avec l'introduction de langues de proximité (langues des pays voisins, langues de l'immigration) et l'introduction des langues régionales dans l'enseignement primaire. Le linguiste a d'ailleurs participé à l'élaboration du plan.

Conclusion

L'apprentissage précoce a connu une évolution en dents de scie et les gouvernements n'ont pas toujours exploité au mieux la richesse perceptive et les ressources phoniques qui caractérisent la période de l'« âge heureux » chez l'enfant. Cependant l'internationalisation grandissante des échanges et la construction européenne l'imposaient de fait. Les différents pays européens ont donc entrepris un certain nombre de réformes et certains se lancent même dans des expérimentations très ambitieuses (introduction de l'apprentissage d'une langue vivante à trois ans par exemple). Jusqu'en 1999, la France faisait un peu figure de mauvais élève. Avec la dernière réforme en date, elle va rejoindre le peloton de tête des pays qui ont mis en place l'apprentissage d'une langue comme matière obligatoire avant l'âge de 7 ans. On ne peut que s'en féliciter surtout lorsque l'approche orale est privilégiée et que les connaissances sur les ressources de l'enfant, ainsi que sa capacité à restituer une bonne prononciation sont prises en compte. Il est encore trop tôt pour juger de la mise en pratique de cette réforme, mais il est possible que son principal défaut soit justement celui de ses qualités. En effet, l'introduction généralisée d'une langue étrangère dès l'âge de 5 ans nécessite la mobilisation d'un contingent de professeurs de langues très important, sans compter l'introduction d'une deuxième langue en sixième et la diversité linguistique couverte. Les objectifs apparaissent d'autant plus difficiles à atteindre que la généralisation de l'enseignement des langues vivantes au CM2 n'est effective que depuis 1999. Par ailleurs, les enseignants se demandent, en cette rentrée 2002-2003 qui voit les crédits de l'Éducation stagner et même amputés, si la réforme pourra être applicable, étant donné qu'elle nécessite des moyens financiers relativement importants.

Bibliographie chapitre IV

- Bijeljac-Babic, Ranka (2000). Acquisition de la phonologie et bilinguisme précoce. Dans Kail et Fayol, 169-192.
- Changeux, Jean-Pierre (1983). *L'Homme Neuronal*. Paris : Fayard, 419 p.
- Cohen, Rachel (1997). Plaidoyer pour les Apprentissages Précoces : Stratégies Educatives pour la Réalisation des Potentialités Humaines. Paris : P.U.F., 317 p.
- Dalgalian, Gilbert, Lieutand, S. and Weiss, F. (1981). *Pour un Nouvel Enseignement des Langues et une Nouvelle Formation des Enseignants*. Paris : Clé International, 143 p.
- Dehaene-Lambertz, Ghislaine, Christophe, Anne et Van Ooijen, Britt (1997). Bases cérébrales de l'acquisition du langage. Dans Kail et Fayol, 61-93.
- Favard, Jean (1992). *Les langues étrangères à l'école primaire : la problématique française*. Dans *L'Apprentissage des Langues Etrangères à l'Ecole Primaire*. Paris : Didier.
- Garabédian, Michèle (1996). Apprendre les langues étrangères le plus tôt possible à l'école primaire : pourquoi ? Pourquoi faire ? Dans *Les Langues à l'Ecole, un Apprentissage ? Actes du Colloque IUFM*, 13 et 14 mars, Dijon, 27-37.
- Girard, Denis (1991). Bilan didactique qualitatif des expériences d'enseignement précoce d'une langue étrangère 1960-80. Dans *Les Langues Vivantes à l'Ecole Elémentaire*, INRP.
- Girard, Denis (1995). *Enseigner les Langues : Méthodes et Pratiques*. Paris : Bordas, 175 p.
- Guberina, Pierre (1991). Rôle de la perception auditive dans l'apprentissage précoce des langues. Dans *Le Français dans le Monde*, août-septembre, 65-70.
- Hagège, Claude (1996). *L'Enfant aux Deux Langues*. Paris : Odile Jacob, 298 p.
- Johnson, E.L. and Newport, S.J. (1989). Critical period effects in second language learning : the influence of maturational state on the acquisition of english as a second language. In *Cognitive Psychology*, 21, 60-99.
- Johnson, E.L. and Newport, S.J. (1991). Critical period effect on universal properties of language : the status of subadjacency in the acquisition of a second language. In *Cognition*, 39, 215-258.
- Kail, Michèle et Fayol, Michel (2000). *L'Acquisition du Langage : le Langage en Emergence de la Naissance à Trois ans*. Vol. I. Paris : PUF, 304 p.
- Konopczynski, Gabrielle (1981a). Compréhension des structures intonatives et de leur fonction linguistique par des enfants de 6 à 11 ans : un exemple d'expérimentation. Dans *Working Papers of Speech and Voice Society*, Toronto, 2, 50-90.
- Krashen, (1983). *The Natural Approach : Language Acquisition in the Classroom*. Englewood Cliffs (New-Jersey): Alemany Press, 191 p.
- Lenneberg, Eric Heinz (1967). *Biological Foundations of Language*. New-York: Wiley & Sons, 489 p.
- Luria, Alexander R. (1974). *L'enfant Retardé Mental*. Paris : Privat, 254 p.
- Maier, Wolfgang (1996). Théorie et pratique de l'enseignement primaire des langues vivantes. Dans *Les Langues à l'Ecole, un Apprentissage ? Actes du Colloque IUFM*, 13 et 14 mars, Dijon, 57-71.
- Moreau, Marie-Louise et Richelle, Marc (1981). *L'Acquisition du Langage*. Bruxelles : Mardaga, 261 p.
- Ninio, A. (1992). The relation of children's world utterances to single word utterance in the input. In *Journal of Child Language*, 118, 87-110.
- Pallier, Christophe, Dehaene, Stanislas, Poline, S., LeBihan, D., Argenti, A-M, Dupoux, Emmanuel et Mehler, Jacques (soumis). Brain imaging of language plasticity in adopted adults: can a second language replace the first ? In *Cerebral Cortex*.
- Penfield, Wilder and Roberts, Lamar (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.
- Penfield, Wilder and Roberts, Lamar (1963). *Langage et Mécanismes cérébraux*. Paris : PUF (trad. Fse texte 1959), 311 p.
- Petit, Jean (1985). *De l'Enseignement des Langues Secondes à l'Apprentissage des Langues Maternelles*. Paris :

Champion, 693 p.

Piaget Jean (1937). *La Construction du Réel chez l'Enfant*. Paris : Delachaux et Niestlé, 399 p.

Pichon, Edouard (1936). *Le Développement Psychique de l'Enfant et de l'Adolescent : Evolution Normale, Pathologie, Traitement : Manuel d'Etude*. Paris : Masson, éd. 1965, 375 p.

Porcher, Louis et Groux, Dominique (1998). *L'Apprentissage Précoce des Langues*. Paris : PUF, 127 p.

Rivière-Raverlat, Jacquotte (1997). *Développer les Capacités d'Ecoute à l'Ecole. Ecoute Musicale, Ecoute des Langues*. Paris : PUF, 207 p.

Selinker, Larry (1969). Language Transfer. In *General Linguistics*, 9, 67-92.

Deuxième partie

Exploration empirique

Chapitre V

Catalogue d'interférences anglais-français

*« Une étude comparative du système de la langue étrangère par rapport au système de la langue maternelle fournit les bases d'une meilleure prise de conscience par l'élève des différences entre les deux systèmes »
(Lanchec, 1976 : 162).*

Table du cinquième chapitre

1) Discussion méthodologique : choix et principes généraux utilisés

- 1.1. L'anglais comme L2
- 1.2. Une théorie de l'interférence : du connu vers l'inconnu
- 1.3. Nécessité d'une étude contrastive
- 1.4. Application de ces principes
 - 1.4.1. Catalogue d'interférences
 - 1.4.2. Batterie de tests de perception

2. Description comparée des systèmes prosodiques de l'anglais et du français

- 2.1. Rythme
 - 2.1.1. Localisation de l'accent
 - 2.1.2. Différence de réalisation acoustique de l'accent
 - 2.1.3. Différence de rythmique
 - 2.1.4. Accent de groupe vs accent de mot
 - 2.1.5. Langue non accentuée vs langue accentuée ?
- 2.2. Intonation
 - 2.2.1. L'accent de phrase
 - 2.2.2. Les tons mélodiques
 - 2.2.3. Description comparée de l'intonation déclarative en anglais et en français
 - 2.2.3.1. Finalité
 - 2.2.3.2. Continuation

3. Description comparée du système vocalique de l'anglais et du système vocalique du français

- 3.1. Caractéristiques générales
 - 3.1.1. Conséquences de la différence d'organisation rythmique sur le niveau segmental
 - 3.1.2. Différences d'organisation de l'espace articulatoire
 - 3.1.3. Conséquences pour un apprenant francophone

- 3.1.3.1. Prédiction d'une typologie des fautes découlant de ces différences
- 3.1.3.2. Paramètres importants nécessaires à la description de l'anglais pour un apprenant francophone
 - 3.1.3.2.1. Durée
 - 3.1.3.2.2. Timbre et stabilité de timbre
- 3.1.3.3. Nécessité d'une description plus poussée
- 3.1.4. Réalisation d'une étude descriptive
 - 3.1.4.1. Locutrices
 - 3.1.4.2. Matériel de parole
 - 3.1.4.3. Conditions d'enregistrement
 - 3.1.4.4. Analyses
- 3.2. Caractéristiques détaillées
 - 3.2.1. Les noyaux syllabiques simples de l'anglais et du français
 - 3.2.1.1. Durée
 - 3.2.1.1.1. Les voyelles brèves
 - 3.2.1.1.2. Les voyelles longues
 - 3.2.1.1.3. Les voyelles du français
 - 3.2.1.1.4. Interférences de durée du français vers l'anglais
 - 3.2.1.2. Timbre
 - 3.2.2. Les noyaux syllabiques complexes de l'anglais
 - 3.2.2.1. Définitions
 - 3.2.2.1.1. La diphtongue
 - 3.2.2.1.2. La triptongue
 - 3.2.2.2. Description des diphtongues et des triptongues de l'anglais en fonction des productions de la locutrice R.P. comparées à celles de la locutrice francophone.
 - 3.2.2.2.1. Les diphtongues
 - 3.2.2.2.2. Les triptongues

Bibliographie du chapitre V

1. Discussion méthodologique : choix et principes généraux utilisés

1.1. L'anglais comme L2

Nous avons choisi de travailler sur l'anglais, pour au moins deux raisons, la première étant d'ordre institutionnel et la seconde, d'ordre linguistique. On constate tout d'abord une prédominance de l'enseignement de l'anglais dans le premier degré. Pendant l'année scolaire 1999-2000, la proportion d'élèves apprenant l'anglais dans le primaire était de 79,8 % contre 15,2 % pour l'allemand, 2,4 % pour l'espagnol et 1,3 % pour l'italien. Qui plus est, la tendance est au recul pour ces trois dernières langues (respectivement 15,8 %, 2,8% et 1,5% en 99-2000¹). Par ailleurs, on trouve cette prédominance dans toutes les académies, sauf celles de Strasbourg et de Nancy. Ce pourcentage ne fait que confirmer la tendance nationale au niveau du secondaire et la tendance européenne (82,7 % des classes de l'Union étaient concernées par l'enseignement de l'anglais en 1991-92). Dans ces conditions, il était donc beaucoup plus simple de trouver des classes en apprentissage précoce de l'anglais. Mais le choix de l'anglais s'explique surtout par ses caractéristiques phonétiques. Si cette langue est relativement proche du français du point de vue du vocabulaire et de l'ordre des mots, il n'en va pas de même en ce qui concerne la prononciation. Les deux langues possèdent des systèmes très différents. Alors que la rythmicité de l'anglais repose sur une tendance accentuelle (Pike, 1946), celle du français repose sur une tendance syllabique. Cette différence de structuration rythmique engendre un régime de tension musculaire très différent, qui conditionne directement la prononciation des voyelles et en particulier la stabilité de leur timbre. Par ailleurs, on pourrait qualifier l'anglais de langue « musicale » car elle se caractérise par de fortes variations de hauteur et d'intensité et elle couvre une large tessiture (plus de deux octaves selon Ginésy, 1995 : 223). Du point de vue descriptif, Faure (1962) a souligné la richesse mélodique de

¹ INSEE (2001). Note d'Information 01-18. Edu050, avril.

l'anglais :

« cette richesse mélodique, à laquelle on a rarement accordé toute l'attention qu'elle mérite, est encore plus frappante lorsqu'on la compare aux ressources correspondantes d'autres langues modernes, comme par exemple, le français ou l'allemand ».

Cette musicalité frappe d'ailleurs l'oreille de l'apprenant francophone :

« nous avons de nombreuses fois entendu les réflexions de nos apprenants français qui, captant une multitude de contrastes sur les syllabes en anglais, ressentent la langue anglaise comme une langue qui « chante » et « accentue » » (Scott, 1987 : 135).

Enfin, le système vocalique de l'anglais est plus riche que celui du français en ce sens qu'il contient plus de contrastes (26 contre 16) et que certains de ces contrastes, les diphtongues et les triptongues n'existent pas en français.

1.2. Une théorie de l'interférence : du connu vers l'inconnu

L'attitude qui consiste à analyser l'inconnu à partir du connu est tout à fait naturelle et nous portons des « préjugés », c'est-à-dire que nous appliquons les règles de fonctionnement de ce que l'on connaît à ce que l'on ne connaît pas. Dans un premier temps, nous avons besoin de ces préjugés, car ils nous permettent d'interpréter le monde à la lumière de ce que nous en savons et par la même, de nous adapter à lui. Le problème fondamental, c'est que nous n'avons pas conscience que la réalité que nous percevons est une construction établie à partir de la structure de notre propre système, autrement dit que cette réalité est filtrée par notre propre système perceptif. Ainsi, lorsqu'on se trouve confronté à une langue étrangère, on la compare intuitivement à sa LM. En effet, l'expérience a forgé un ensemble d'attentes perceptives (Lhote, 1995 : 48) et d'habitudes articulatoires qui ont permis l'acquisition de la LM, mais qui se montrent inadaptées au traitement et à la production d'une langue étrangère, surtout dans le cas où la langue cible est très éloignée de la LM (comme c'est le cas pour l'anglais et le français). Comme le souligne Troubetzkoy dès 1949 :

« l'homme s'approprie le système de sa langue maternelle. Mais s'il entend parler une autre langue, il emploie involontairement pour l'analyse de ce qu'il entend, le « crible phonologique »

de sa langue maternelle qui lui est familier. Et comme ce crible ne convient pas pour la langue étrangère entendue, il se produit de nombreuses erreurs et incompréhensions. Les sons de la langue étrangère reçoivent une interprétation phonologiquement inexacte, puisqu'on les fait passer par le « crible phonologique » de sa propre langue » (Troubetzkoy, 1949 : 54).

Cet effet de filtrage est particulièrement puissant, puisqu'il provoque chez l'apprenant une surdité sélective, une surdité dite « phonologique » aux sons de la langue étrangère.

Précisons que lorsque l'on parle de « surdité », on ne fait pas référence à un déficit d'audition, mais à une réorganisation perceptive en faveur de la LM. Tous les sons ne sont pas égaux devant cet « effet de filtrage » (Ostiguy et al., 1996 : 79) : certains sont assimilés purement et simplement aux sons de la LM en raison de leur similitudes de timbre ;

« on parlera alors de transfert négatif ou d'interférence lorsque l'apprenant va se servir en L2 d'un élément de L1 qu'il croit identique alors que celui-ci est différent » (Guimbretière, 1994 : 17).

D'autres sont analysés comme des exemplaires déviants, d'autres comme des sons différents, d'autres enfin comme des sons non linguistiques. Pour classer ces différences, Best propose un modèle d'assimilation perceptive, le PAM (« *Perceptual Assimilation Model* ») (1988, 1993), qui part du principe que, lorsque les différences entre les phonèmes de la langue étrangère et ceux de la LM sont claires, la discrimination est facilitée. Le modèle propose six cas de figure dans la discrimination d'une paire de phonèmes non natifs (les contrastes UU et UC ont été rajoutés au modèle en 1993) :

1. Les deux phonèmes sont assimilés à deux phonèmes différents de la LM. Dans ce cas, la discrimination est excellente (« *Two-Category, TC* »).
2. Les deux phonèmes sont assimilés à un seul phonème de la LM, mais divergent de l'« idéal natif » (l'un est acceptable et l'autre, considéré comme déviant). Comme ils sont différenciés, leur discrimination est modérée, voire bonne (« *Category-Goodness Difference, CG* »).
3. Les deux phonèmes sont assimilés à un seul phonème de la LM et divergent tous les deux de l'« idéal natif ». L'auditeur ne fait pas de différence entre les deux contrastes et la discrimination est très pauvre (« *Single-Category, SC* »).

4. Les deux phonèmes ne sont pas catégorisables par l'auditeur, car ils tombent en dehors de l'espace phonétique familier. Dans ce cas, la discrimination est modérée, voire très pauvre (« *Both Uncategorizable, UU* »).
5. L'un des deux phonèmes est assimilé à un phonème natif et l'autre tombe dans un espace phonétique non familier, en dehors des catégories natives. Leur discrimination est très bonne (« *Uncategorized vs. Categorized, UC* »).
6. Les deux phonèmes tombent en dehors du domaine du langage et sont considérés comme des sons n'appartenant pas à la parole. Leur discrimination est bonne, voire très bonne (« *Non-Assimilable, NA* »).

Les auditeurs adultes perçoivent donc les phonèmes étrangers par rapport à la structure des phonèmes appartenant à leur LM.

Ces phénomènes de crible sont d'une importance capitale pour la prononciation de la langue cible : en effet, on ne peut pas produire correctement des sons qu'on perçoit mal ou qu'on ne perçoit pas :

« ce qu'on appelle l'« accent étranger » ne dépend pas du fait que l'étranger en question ne peut pas prononcer un certain son, mais plutôt du fait qu'il n'apprécie pas correctement ce son. Et cette fausse appréciation des sons d'une langue étrangère est conditionnée par la différence existant entre la structure phonologique de la langue étrangère et celle de la langue maternelle du sujet parlant » (Troubetzkoy, 1949 : 56).

La notion de crible phonologique se réfère essentiellement au niveau segmental. En ce qui concerne le niveau prosodique, Delattre (1965) parle d'« interférences d'intonation » et De Bot (1986), de « transferts intonatifs ». Nous préférons parler de crible prosodique, de manière à englober les interférences d'ordre rythmique, qui vont également provoquer des erreurs graves,

comme nous le verrons plus loin (voir 2.2.1.). Les interférences prosodiques sont peut être les plus difficiles à combattre. En effet, l'organisation rythmique et intonative de la LM est tellement ancrée, tellement naturelle, car première dans l'ordre d'acquisition, qu'elle en devient « transparente » pour le locuteur : la plupart du temps, il n'a pas conscience de la façon dont il accentue et module ses énoncés, alors qu'il a en partie conscience de la façon dont il prononce les éléments segmentaux. Les locuteurs sont totalement imprégnés par la rythmique de leur LM et s'ils ne sont pas sensibilisés aux différences prosodiques dont nous venons de parler, ils appliqueront le patron rythmique et intonatif de leur LM à la LE, sans en avoir la moindre conscience. La surdit  prosodique est donc tr s puissante et il est tr s difficile de s'en affranchir, d' chapper   des automatismes d'autant plus install s qu'ils charpentent l'ensemble de la langue.

1.3. Nécessité d'une étude contrastive

Pour connaître les interférences qui se produiront entre deux langues, il est nécessaire de connaître le système de la langue cible et celui de la LM. Pour arriver à cette connaissance, on peut faire une analyse fine et détaillée de chacun de ces systèmes. La description prosodique et segmentale est actuellement poussée, comme en témoigne la sortie récente de plusieurs ouvrages et articles consacrés au français (Lacheret-Dujour et Beaugendre, 1999 ; Rossi, 1999 ; Di Cristo, 1999) et à l'anglais (Hirst, 1998 ; Montreuil, 2001). En effet, ces deux langues sont parmi les plus abondamment décrites. Cependant, une analyse comparative permettra de limiter la description, en mettant uniquement en lumière les points d'achoppement entre les deux langues. Ce type d'analyse « contrastive » (on parle également d'analyse différentielle) est né aux Etats-Unis dans les années 50. D'après Lado (1957), il repose sur le principe qu'en situation d'apprentissage, lorsque deux langues sont en contact, il peut y avoir deux sortes de transferts :

- soit un transfert positif, si les caractéristiques de la langue cible ressemblent à celles de la langue source ; dans ce cas, l'apprentissage de la langue cible sera facilité ;
- soit un transfert négatif, si les caractéristiques de la langue cible divergent par rapport à celles de la langue source ; dans ce cas, l'apprentissage de la langue cible sera gêné.

Du point de vue de l'apprenant, cette approche est très intéressante, car elle permet de prédire les erreurs qu'il pourrait produire. Le principal danger reste de relever plus d'erreurs potentielles que l'apprenant n'en produira réellement, notamment au niveau du lexique et de la grammaire. Cependant, l'analyse contrastive s'avère particulièrement adaptée à la prédiction des interférences de type phonétique. Par ailleurs, d'après Léon et Martin (1971 : 37), l'enseignement de l'intonation suppose une analyse préalable, comparative, de la langue de départ et de la langue d'arrivée.

Il existe des études comparatives pour l'anglais et le français, à commencer par celles de Delattre (1965) s'intéressant aux deux niveaux, prosodique et phonémique. D'autres qui ne s'occupent que d'intonation (Delattre, 1961 ; Pritchard, 1985 ; Grover et al., 1987) ou que de segmental. Certains manuels à destination des apprenants francophones sont fondés sur une analyse comparative de l'anglais par rapport au français pour la totalité du système (Faure, 1948 ; Adamczewski et al., 1973 ; Ostiguy et al., 1996).

1.4. Application de ces principes

1.4.1. Catalogue d'interférences

A partir de la littérature et de nos propres observations, nous réaliserons une analyse comparée du français et de l'anglais. Celle-ci nous permettra de prédire les interférences les plus graves qui se produiront entre ces deux langues et d'en faire le relevé systématique. Nous donnerons à cette liste d'interférences le nom de « catalogue d'interférences » (Dodane, 2001) et son élaboration respectera les principes suivants :

- à l'image de ce qui se passe lors de l'acquisition de la LM, nous accorderons la primauté à la prosodie ;
- la prosodie étant constituée du rythme et de l'intonation, nous les aborderons séparément, mais sans jamais perdre de vue leur interdépendance ;
- la première partie du catalogue concernera la prosodie et la seconde partie, l'organisation segmentale, sans jamais oublier que le premier niveau sous-tend le second ; nous ne nous intéresserons qu'aux voyelles car ces sons paraissent assumer un grand rôle pendant l'acquisition (cf. baby talk et berceuses, chapitre II, 2.4.) ; en outre, ce sont elles qui portent la plupart des paramètres prosodiques.

1.4.2. Batterie de tests de perception

Deux batteries de tests seront réalisées à partir des principales interférences relevées dans ce catalogue : une batterie « prosodie » et une batterie « contrastes vocaliques » (Volume II, annexe n°4.2.). Ces batteries serviront d'outil d'évaluation pour déterminer les capacités de discrimination en anglais et pour l'étude des interférences prosodiques et phonémiques entre le français et l'anglais. Nous avons donc réalisé une batterie de 8 tests portant sur :

- les spécificités prosodiques de l'anglais par rapport à celles du français (accent de mot, place du pivot au sein de l'énoncé, rythmicité accentuelle, tons mélodiques et schémas mélodiques).

- les contrastes vocaliques spécifiques de l'anglais par rapport aux français (oppositions vocaliques entre voyelles de timbres très proches en anglais ; oppositions vocaliques entre voyelles de timbres très proches en anglais et en français ; diphtongues et triphthongues de l'anglais).

2. Description comparée du système prosodique de l'anglais et du système prosodique du français

2.1. Rythme

Chaque langue alterne différemment les syllabes faibles et les syllabes fortes et « marque » le rythme différemment.

« Linguistiquement, nous sommes conditionnés par le rythme de notre langue maternelle. Apprendre une langue étrangère, c'est aussi, c'est avant tout, changer de rythme, subir une sorte de « recyclage » rythmique et intonatif » (Meschonnic, 1982 : 420).

Pour décrire la spécificité de cette rythmicité, nous commencerons par décrire la localisation des syllabes accentuées et observer comment se combinent les paramètres acoustiques pour réaliser l'accent. Ensuite, nous déterminerons le type de rythmique engendrée par l'alternance de ces accents.

2.1.1. Localisation de l'accent

L'accent primaire est désigné par de multiples noms, selon les auteurs. Pour souligner la différence entre les deux accents (au niveau de la localisation, de la réalisation acoustique, de la fonction...) et dans un but de clarté, nous parlerons d'accent final pour le français et d'accent tonique pour l'anglais (*« Nous préférons le désigner désormais sous le nom d'accent tonique, en signifiant par là que le ton en est le facteur essentiel »*, Faure, 1962 : 150).

2.1.1.1. Prédicibilité de l'accent final en français

Une opinion répandue parmi les phonéticiens, c'est que le français est une langue qui possède un seul accent et que cet accent affecte toujours la dernière syllabe du groupe rythmique. Dans ce sens, c'est un accent de groupe. Le groupe rythmique est un groupe de syllabes formant une unité sonore coïncidant souvent avec le syntagme grammatical.

Dans un groupe rythmique, les mots perdent leur individualité, car l'accent de mot disparaît pour se reporter à la fin du groupe. Cet accent, appelé également final, logique, objectif, tonique, normal ou encore interne, se réalise dans 100 % des cas sur la dernière syllabe (Delattre, 1965²), c'est-à-dire la pénultième du groupe rythmique, et ceci, quel que soit le nombre de syllabes composant le groupe rythmique. Pour cette raison, on parle d'oxytonie. Le français est donc une langue à accent fixe, tout comme le tchèque, le finnois, le hongrois et l'islandais (mais pour celles-ci, l'accent affecte l'initiale et non la finale). En français, l'accent primaire frappe donc la syllabe finale du groupe rythmique, engendrant ainsi un rythme de type iambique³.

2.1.1.2. Variabilité de l'accent tonique en anglais

En anglais, les mots conservent leur particularité, même s'ils sont mis en contexte au sein d'un énoncé. Tous les mots lexicaux (« *content words* » en anglais), c'est-à-dire les mots à sémantisme plein comme les verbes, les adjectifs, les substantifs, les démonstratifs et les adverbes portent un accent tonique. Ainsi, dans la phrase « *The workmen from Boston were leaving* », ce sont les mots « *workmen* », « *Boston* » et « *leaving* » qui sont frappés par l'accent. L'accentuation affectant les mots n'empêche bien sûr pas leur intégration dans une unité de niveau supérieur et il convient de respecter la distinction faite par Faure (1948 : 101) entre l'accent syllabique (accent de mot) et l'accent de phrase (pivot) :

« Chaque mot du groupe garde en général son accent propre, cet accent syllabique se doublant d'un accent de phrase qui porte sur les mots les plus importants de la phrase, dont l'accent syllabique se trouve ainsi renforcé ».

² Delattre (1965) : statistiques sur 2000 syllabes utilisées dans chaque langue (1000 en textes narratifs et 1000 en passages dramatiques)

³ En métrique ancienne, l'iambe désigne un pied composé de deux syllabes, la première brève et la seconde longue (*Dictionnaire Hachette Encyclopédique* sur CD, 2000).

L'accentuation anglaise est réputée complexe, car il est difficile pour un apprenant de savoir où placer l'accent tonique dans les mots multisyllabiques. La place de l'accent y est tellement mobile que certains phonéticiens ont affirmé qu'elle était non prédictible. Cependant, on sait maintenant qu'elle est gouvernée par des règles générales (décrites entre autres par Faure, 1948 ; Chomsky et Halle, 1968 ; Guierre, 1984 ; Ginésy, 1995 ; Hirst, 1998). Ces règles évitent à l'apprenant de mémoriser le schéma accentuel de chaque mot à partir du dictionnaire.

Par exemple, la loi d'alternance pousse les anglais à alterner les syllabes accentuées et les syllabes atones. Dans les mots de plus de deux syllabes, si la première syllabe est accentuée, la troisième portera un léger accent (appelé accent secondaire). Pour les mots de deux syllabes, trop courts pour être soumis à la loi d'alternance, c'est la loi de valeur qui se montre déterminante : ils seront accentués de préférence sur leur radical (« *Work* » dans « *workmen* », « *lea* » dans « *leaving* »).

Les statistiques données par Delattre (1965) montrent une tendance à l'accentuation de la syllabe initiale. Faure (1948 : 90) explique que les anglais ont

« tendance à accentuer dans un mot la syllabe la plus chargée de sens, c'est la raison pour laquelle ils font, en général, tomber l'accent sur la syllabe radicale, à l'exclusion des éléments accessoires (préfixes, suffixes, etc...). Or, en anglais, cette syllabe du radical se trouve très souvent au début des mots ».

Dans la phrase « *The workmen of Boston were leaving* », c'est la syllabe initiale de chaque mot lexical qui porte l'accent, soit « *work* », « *Bos* » et « *lea* », ce qui engendre un rythme de type trochaïque⁴. La localisation de l'accent tonique en anglais est variable, affectant de préférence la syllabe initiale des mots lexicaux. Les anglais sont donc habitués à un rythme dont

⁴ En métrique ancienne, le trochée ou le chorée désigne un pied composé de deux syllabes, une longue et une brève

l'élément acoustiquement fort est placé sur la syllabe initiale du mot lexical.

2.1.2. Différence de réalisation acoustique de l'accent

L'accent en français se matérialise avant tout grâce au paramètre acoustique de durée, les syllabes accentuées étant en moyenne deux fois plus longues que les syllabes atones. Pour cette raison et puisque cet accent est localisé en finale de groupe rythmique, on parle d'allongement final. La durée de la syllabe constitue le paramètre le plus important dans la perception de l'accentuation. Les écarts d'intensité sont en effet peu importants et, en ce qui concerne la hauteur, on ne constate aucune rupture mélodique, mais plutôt un « glissando », c'est-à-dire une variation de fréquence au cours de la production de la voyelle (Delattre, 1965).

(Dictionnaire Hachette Encyclopédique sur CD, 2000).

En anglais, l'accent se réalise avant tout grâce aux paramètres de hauteur et d'intensité : la syllabe accentuée est beaucoup plus forte, plus audible que les autres ([+intensité]), au détriment des autres syllabes qui deviennent presque inaudibles et qui disparaissent même. Mais surtout, la syllabe accentuée est prononcée sur une note différente des autres, ce qui crée une rupture tonale ([+mélodique]). Et selon Faure (1962 : 125), c'est ce paramètre qui se montre déterminant dans la mise en relief des syllabes accentuées en anglais :

« une syllabe réputée accentuée doit, à peu près toujours, l'essentiel de son relief acoustique au contraste mélodique qui l'oppose aux syllabes inaccentuées ».

Il pense que les différences de hauteur sont « *beaucoup plus efficaces que les variations de durée et surtout d'intensité* » (Faure, 1962 : 72). On parle d'ailleurs de « pitch accent ». Pour lui, l'augmentation d'intensité et l'augmentation de durée sont dépendantes de la variation de hauteur.

« L'élément essentiel de ce rythme, c'est-à-dire celui qui donne aux syllabes « fortes » leur plus haut degré de perceptibilité, est un facteur mélodique, généralement associé à des variations correspondantes d'intensité et de durée, dont le rôle nous paraît être beaucoup moins importants qu'on semble l'avoir pensé jusqu'ici ».

2.1.3. Différence de rythmique

2.1.3.1. La régularité syllabique du français

Nous avons vu que la syllabe finale de groupe rythmique est en moyenne deux fois plus longue que les syllabes non finales. Les syllabes atones ont donc toutes approximativement la même durée (du point de vue perceptif), ce qui donne une impression de régularité syllabique au français, d'isochronie. Pour cette raison, Pike (1946) parle de langue « syllable-timed ». C'est d'ailleurs le nombre de syllabes qui fonde notre versification, et non l'accentuation : seul est

retenu le nombre de syllabes ou pieds comme critère métrique. Ainsi, même avec un débit rapide, le français garde sa netteté articulatoire et individualise chaque syllabe. A cause de la place de l'accent en finale, l'articulation est très tendue, car l'élément important est toujours à venir. De facto, depuis Wenk et Wioland (1982), on parle maintenant plutôt de langue « *trailer timed* », c'est-à-dire à remorque (l'élément important figurant en arrière).

En fait, l'isochronie en français n'est pas si parfaite et il semble qu'il y aurait un allongement progressif des syllabes inaccentuées plus on se rapproche de la syllabe accentuée finale (Pasdeloup, 1990 : 224-225)

« La durée des syllabes inaccentuées tend à progresser jusqu'à la syllabe accentuée dans un mouvement général de ralentissement progressif ».

Wenk et Wioland (1982) se demandent cependant si c'est bien la répétition des syllabes successives qui donne sa spécificité rythmique au français. Le rythme ne serait-il pas créé par le retour de l'accent plutôt que par la régularité des syllabes atones ? S'ils ont raison, le français serait alors plutôt une langue à rythmicité accentuelle, et ils invoquent comme arguments contre l'isochronie du français que les syllabes accentuées n'ont pas la même longueur que les syllabes atones et que l'accent peut remplir une fonction distinctive dans le cadre du groupe rythmique. Mais le fait que l'accent puisse être distinctif en français ne va pas à l'encontre d'un écoulement régulier des syllabes atones. Quant aux syllabes accentuées, on ne peut pas dire que ce soit leur retour régulier qui assure la rythmique, à cause précisément de leur emplacement en finale. Leur apparition est conditionnée par la longueur du groupe rythmique et ne peut en conséquence pas être régulière. Il suffit d'essayer de reproduire l'écoulement des syllabes en frappant un crayon pour se rendre compte que la rythmicité syllabique est bien une réalité. Au lieu de classer les langues dans deux groupes opposés, « syllable timed » ou « stress timed », Dauer (1987) suggère que les différentes langues occupent différentes positions le long d'un axe allant du « syllable-timed » au « stress-timed » : l'emplacement de chaque langue le long de cet axe dépendrait dans chaque cas de la satisfaction d'un certain nombre de critères.

Depuis les travaux de Fonagy (1980) sur l'accent probabilitaire, on sait qu'une description de la rythmique du français contemporain ne peut plus se limiter à l'accent final et à la rythmicité syllabique. En effet, il existe une autre proéminence qui affecte de manière optionnelle la

syllabe initiale et de manière obligatoire la syllabe finale de chaque item lexical (Hirst et Di Cristo, 1984). Contrairement à l'accent final, cet accent ne se manifeste pas par un allongement de la durée, mais par un « pic » mélodique, beaucoup plus discret que celui affectant les syllabes accentuées en anglais.

Cet accent, qu'on appelle secondaire (Di Cristo, 1999 ; Astesano, 2001), permet de délimiter à droite une unité plus petite que le mot prosodique : le pied⁵ métrique, considéré par Di Cristo (1999 : 148) comme l'unité accentuelle minimale. Le système accentuel du français est donc beaucoup plus complexe, comme le montre Di Cristo (1999), qui tente d'établir le cadre accentuel du français contemporain. Il démontre notamment la présence de « *schèmes barytoniques* » et d'« *arcs accentuels* » (Di Cristo, 1999 : 185), dans lesquels seules les syllabes initiales et finales d'un groupe de mots reçoivent un accent (comme dans son exemple, « *les ONdées nuaGEUSES* »).

2.1.3.2. La régularité accentuelle de l'anglais

En revanche, la rythmique de l'anglais se caractérise par le retour régulier, non pas des syllabes atones, mais de l'accent tonique. Pour cette raison, Pike (1946 : 34) parle de langue « *stress-timed* », c'est-à-dire que

« *les unités rythmiques ont tendance à se suivre de façon à ce que le laps de temps entre le début de leurs syllabes proéminentes soit uniforme* ».

L'isochronie engendrée par le retour régulier de l'accent tonique est très similaire à un rythme de percussion, où les temps forts reviendraient à intervalles réguliers. C'est pourquoi, on parle de « *drum beat* », pour souligner la particularité de cette rythmique. L'important est de maintenir un intervalle analogue entre deux accents (1 seconde environ selon Faure, 1948 : 101). Peu importe le nombre de syllabes atones : si elles sont peu nombreuses, elles seront dilatées et si elles sont nombreuses, elles seront au contraire compressées ou même éliminées. Etant donné que chaque mot lexical est affecté d'un accent tonique, il en résulte que plus ces

⁵ Pied (Abercrombie, 1964) : notion empruntée à la poétique ("Chaque syllabe d'un vers dans la Métrique Ancienne") ; « *dans notre acception, toute suite de syllabe terminée par un accent secondaire* » (Lacheret-Dujour et Baugendre, 1999 : 283).

mots sont nombreux au sein d'un énoncé, plus il y aura de syllabes accentuées et plus le rythme sera ralenti.

Cette organisation rythmique interfère directement avec le niveau segmental : à l'intérieur d'une syllabe compressée, la voyelle est réduite ; elle n'a pas le temps de se réaliser pleinement et son timbre se centralise pour s'approcher du [ə]⁶. Par ailleurs, l'inégalité de durée entre les syllabes introduit un rythme beaucoup plus complexe qu'en français.

Pour s'en rendre compte, il suffit d'essayer de le transcrire musicalement, comme l'a fait Faure (1948 : 102, voir figure n°54) pour l'énoncé « *Shut the door* ».

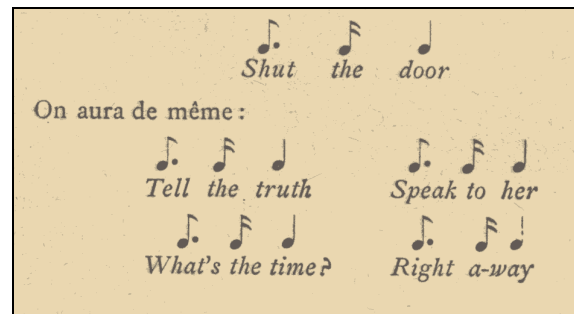


Figure n°54 : Faure (1948 : 102)

Il obtient ainsi la séquence musicale  suivante : avec comme référence la noire symbolisant la durée d'une syllabe accentuée.

« *Shut* » et « *door* » sont deux mots monosyllabiques à sémantisme plein, donc ils portent un accent. En revanche, le déterminant « *the* » n'est pas accentué, car c'est un mot outil. Pour indiquer la réduction de sa durée, Faure l'a transcrit à l'aide d'une double croche, c'est-à-dire par un élément trois fois plus court que la croche pointée initiale correspondant à la durée du mot « *shut* » et quatre fois plus court que la noire finale correspondant à la durée du mot « *door* ».

⁶ Voir paragraphe 3.1.1.1.

2.1.3.3. Régulation initiale versus régulation finale

En anglais, la syllabe accentuée tend à se placer en position initiale et elle est suivie de syllabes non accentuées. La régulation s'effectue donc à partir du début du groupe rythmique (régulation initiale). En français, au contraire, la syllabe accentuée se trouve en position finale et la régulation s'effectue en fonction de la fin du groupe rythmique (régulation finale). Cette différence de régulation engendre selon Wenk et Wioland (1982) un régime de tension musculaire différent dans les deux langues : en anglais, le régime de tension est décroissant car la syllabe accentuée apparaît en début de groupe rythmique ; les syllabes atones qui suivent sont prononcées de manière très relâchée et le timbre des voyelles se trouve réduit. En français au contraire, le régime de tension est croissant car la syllabe accentuée apparaît en fin de groupe rythmique. De ce fait, les syllabes atones, et en particulier le timbre de leurs noyaux, c'est-à-dire des voyelles, ne manifeste pas de réduction.

Par ailleurs, la montée mélodique marquant la syllabe accentuée peut se produire progressivement sans avoir recours à des mouvements brusques comme c'est le cas en anglais. La tension favorise un contrôle plus fin et indépendant des paramètres. En outre, l'augmentation de la durée ne s'accompagne pas d'augmentation de l'intensité.

2.1.3.4. Différences dans la durée d'acquisition de la rythmique anglaise et de la rythmique française

A cause de la mobilité de la place de l'accent en anglais, il est très difficile pour un apprenant d'inférer les règles qui lui permettront de bien placer les syllabes accentuées, surtout si celui-ci parle une langue où la place de l'accent est fixe, comme en français. Un francophone aura tendance à appliquer les règles d'accentuation de sa LM à l'anglais et donc à accentuer la syllabe finale. S'il n'est pas étonnant qu'un apprenant rencontre de telles difficultés dans l'apprentissage d'une langue étrangère, en raison des phénomènes de crible « prosodique » au niveau de la perception et d'automatismes articulatoires au niveau de la production, il semble que, lors de l'acquisition de leur LM respective, le « coût cognitif » soit beaucoup plus élevé pour les bébés anglais que pour les bébés français. Les travaux de Konopczynski (1990) montrent que l'allongement final est en place chez les bébés francophones, et ce de manière stable, dès l'âge de 16 mois. En effet l'acquisition du cadre rythmique est facilitée par le fait que la position de l'accent primaire est entièrement prédictible en français. Curieusement, le babillage des petits anglophones est également marqué par un allongement final pendant une certaine période, bien qu'il soit moins stable qu'en français. Konopczynski (1991 : 333) explique ces ressemblances par le fait que :

« l'enfant acquiert d'abord un rythme de type « universel » avant d'acquérir la structuration accentuo-temporelle propre à sa langue maternelle ».

Il se démarque ensuite graduellement de ce rythme « naturel » pour se rapprocher

progressivement de la rythmicité accentuelle de l'anglais, qui ne sera pas complètement mise en place qu'entre l'âge de 3 ans et 3 ans et demi. Ainsi, un bébé anglais met beaucoup plus de temps à maîtriser le schéma accentuel de sa LM, sans doute à cause d'une plus grande complexité. En effet, l'acquisition du rythme est généralement plus longue dans les langues à accent mobile, car la place de l'accent est moins facilement prédictible que dans les langues à accent fixe ou quasi fixe.

2.1.4. Accent de groupe vs accent de mot

2.1.4.1. Valeur de l'accent en français

En français, les mots perdent leur individualité au profit du groupe rythmique⁷ (nous préférons parler de groupe rythmique, car ce terme est plus simple à comprendre pour un apprenant).

« La langue française n'est pas une langue à accent de mot : cela veut dire que les différentes syllabes d'un mot ne sont pas hiérarchisées par une relation accentuelle » (Lhote, 1995 : 140).

Mackay (1987) définit l'accent de mot comme la proéminence qui affecte une ou plusieurs syllabes d'un mot multisyllabique. En effet, qu'est-ce qui peut indiquer à l'auditeur que le groupe « *La voiture jaune* » est composé de deux (« *La voiturejaune* ») ou de trois mots (« *La voiture jaune* ») ? Llorca (1987 : 54) souligne que la « *la spécificité du mot n'apparaît qu'à l'écrit et dans la conceptualisation* ». Une unité rythmique correspond en français à une unité de sens et n'est pas facilement identifiable pour des auditeurs non natifs, en particulier ceux dont la LM est une langue à accent de mots. En effet, alors qu'en anglais les mots sont nettement différenciés, l'interdépendance établie au niveau sonore entre les unités en français favorise leur interdépendance au niveau sémantique et réduit le poids de chacun de ses constituants (Llorca, 1987 : 53). Llorca (1987 : 52) donne l'exemple suivant : « *Il joue du violon* » prononcé d'une seule traite fonctionne comme unité simple équivalent à « *Il violonne* ». Les deux segments perdent donc quelques caractéristiques propres au profit d'une configuration globale qui les unit. Au contraire, si on dit « *Il joue / du violon* », en marquant une pause entre les deux, on attribue à chaque segment une configuration sonore propre. Pour

⁷ Lacheret-Dujour et Beaugendre (1999 : 45) définissent le groupe rythmique (ou groupe accentuel) comme un groupe « *correspondant à une suite de syllabes dont la dernière est marquée par un accent primaire* ». Di Cristo (1999 : 148) préfère le terme de mot prosodique qu'il définit comme le « *domaine de projection de l'accent final primaire* ».

préserver l'identité des deux segments, le locuteur oppose à l'enchaînement filé, la juxtaposition de deux signifiants distincts (« *Il joue du violon, du piano et de la guitare* »). L'opposition de la forme liée et de la forme comportant une rupture possède un caractère phonologique. La fusion entre les mots réalisée à cause de leur proximité est un phénomène spécifique au français. Certes, Rossi (1985) pense qu'un accent léger intervient dans l'identification et la démarcation du mot en français (en affectant la dernière syllabe de chaque mot lexical).

Cet accent interne assumerait donc bien selon lui une fonction morphologique. Di Cristo (1999 : 192-193) pense également que

« le mot, qui constitue l'unité accentuable minimale en français est doté dans la représentation sous-jacente d'une proéminence initiale et d'une proéminence finale. En d'autres termes, nous posons l'hypothèse que l'organisation métrique des mots accentogènes est assujettie à un principe de base que nous appellerons principe de bipolarisation ».

Toutefois, il insiste sur le fait qu'on ne peut pas parler d'accent lexical : « *le français ne possède pas d'accent lexical et cette remarque ne souffre aucune exception* » (1999 : 192). En effet, l'accent en français n'est pas distinctif, dans le sens où il ne sert pas à opposer des mots, comme c'est le cas en anglais.

En revanche, dans les langues à accent de mots comme l'anglais, les mots conservent leurs schémas accentuels, c'est-à-dire leur individualité, même s'ils sont mis en contexte. Bien sûr, il y a une réorganisation globale des unités dans une unité de niveau supérieur, mais le schéma qui s'applique à chacun de ces mots garde sa spécificité, à savoir son patron accentuel. Comme le souligne Llorca (1987 : 55), ces langues sont « *des langues à mots avant d'être des langues à accent de mots* ». Un anglais peut ainsi segmenter facilement les mots dans sa langue, même ceux qu'il ne connaît pas : il est guidé par la proéminence qui les affecte. Concernant la

production, la place de l'accent est même plus importante que la réalisation des phonèmes, car l'accent permet d'identifier les syllabes accentuées, c'est-à-dire les plus importantes au niveau du sens (par exemple, la syllabe portant le radical). Or, un accent mal placé au sein d'un mot en perturbera gravement la compréhension..

2.1.4.2. Valeur phonologique de l'accent lexical en anglais

Puisque l'accent en français est un accent de groupe, il n'a pas de valeur distinctive au niveau lexical. Il ne sert pas à opposer des mots, comme c'est le cas en anglais pour les deux versions du mot « *abstract* », qui se distinguent uniquement par leur accentuation (**extrait sonore n°23**). Comme adjectif, « *abstract* » se traduit en français par « *abstrait* » et comme substantif, par « *résumé* ». La transcription phonétique de ce mot [ˈæbstrækt] indique que l'accent porte sur la première syllabe et en effet, on remarque une rupture tonale sur la figure n°55 (à gauche), marquée par un ton mélodique ascendant (226-251 Hz), tandis que la syllabe atone est prononcée avec un ton beaucoup plus bas et descendant (188-143 Hz).

On parle de ton mélodique quand la hauteur varie au cours d'une même syllabe.

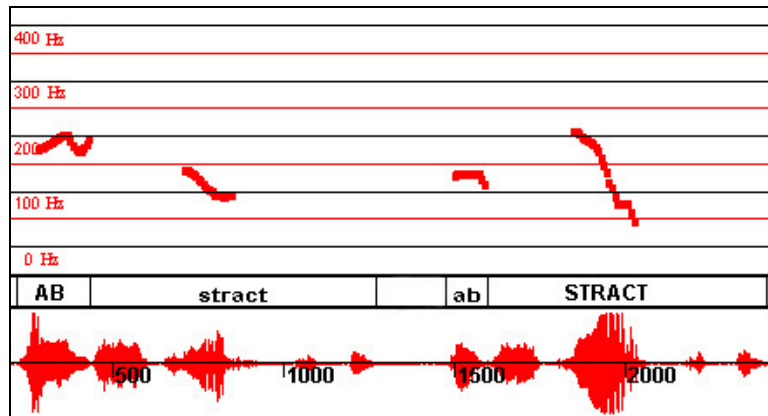


Figure n°55 : ['æbstraekt] et [æb'straekt] prononcés par la locutrice anglophone

Le mot suivant, le verbe « *abstract* » se traduit par « *soustraire, dérober qch à qn, détourner de l'argent* ». Selon sa transcription, l'accent porte sur la dernière syllabe [æb'straekt]. Effectivement, sur le tracé, on observe exactement l'inverse du premier mot : le décrochement mélodique se fait sur la seconde syllabe (258-126 Hz), alors que la première syllabe est « recto-tono » (178-180 Hz). Il apparaît encore une fois, au travers de cet exemple, que le paramètre mélodique est essentiel pour marquer l'accent de mot. On constate également une différence de durée substantielle entre la première syllabe de « 'ABstract » (221 ms) et la première syllabe de « ab'STRACT » (126 ms), quasiment deux fois plus courte et une différence d'intensité (intensité plus grande sur la première syllabe dans « 'ABstract » et inversement pour « ab'STRACT »). Mais selon Faure (1962 : 144),

« la pression exercée sur notre jugement en matière d'accent par les éléments mélodiques, du fait de notre aptitude exceptionnelle à en saisir les plus petites nuances, fait que le seul changement de ton, à l'exception de toute altération de l'intensité et de la durée, suffit par exemple, à opposer « 'permit » au verbe « per'mit » ».

Le passage de l'un à l'autre mot provoque un changement de sens. Puisqu'en changeant la place

de l'accent au sein d'un mot, on change le sens de ce mot, l'accent a donc une valeur phonologique en anglais.

2.1.5. Langue non accentuée vs langue accentuée ?

La « faiblesse » acoustique de l'accentuation en français, sa régularité syllabique et son caractère non distinctif ont apporté de nombreux arguments à des chercheurs tels que Hjelmslev, Torgueby et Pilch pour dire que le français était une langue sans accent.

Le synchronisme entre accent et intonation a été également avancé pour dire que l'accent n'avait pas de fonction propre, les syllabes accentuées étant également le point où le contour de hauteur est réalisé lors d'un groupe accentuel final dans une unité d'intonation. D'après Beckman (1992), cité par Di Cristo (1999 : 151) :

« les propriétés de l'organisation prosodique du français incitent à considérer l'accent final comme un indicateur de la limite du groupe rythmique (edge-marking). Il diffère donc fondamentalement, selon cet auteur, de l'accent de l'anglais et du suédois qui peut être interprété comme un indicateur de la tête du pied accentuel (head marking), complètement indépendant de tout effet d'allongement final de syntagme ». Selon Di Cristo, « ce point de vue tend à promouvoir l'idée que, en définitive, le français ne posséderait pas d'accent à proprement parler, mais seulement des indices démarcatifs des limites des mots ou des syntagmes ».

Ce point de vue est remis en question par Ladd (1996), qui pense au contraire que les syllabes finales de groupe rythmique sont bien en français des syllabes métriquement fortes. Ce point de vue est partagé par Di Cristo (1999 : 201) :

« l'accent initial, comme l'accent final peuvent être considérés comme les têtes métriques des pieds accentuels (ou unité rythmique minimale) en français ».

Par rapport à l'anglais, Vaissière pense que l'accent a bien une fonction en français puisqu'il permet de délimiter les groupes rythmiques les uns par rapport aux autres. Ainsi, elle parle de langue de frontière (par opposition à une langue accentuelle comme l'anglais). L'accent a donc

une valeur démarcative. Pour Llorca (1987), il a en outre une fonction unitaire, dans le sens où il assure la cohésion des éléments du groupe.

2.2. Intonation

2.2.1. L'accent de phrase

2.2.1.1. L'accent de phrase en anglais

Faure (1962 : 125) distingue l'accent de mot,

« celui qui détache une syllabe donnée dans un mot isolé » de l'accent de phrase qui permet « le renforcement de l'accent de mot, ou de l'accentuation volontaire d'une monosyllabe habituellement inaccentuée ».

Cet accent de phrase porte sur le mot le plus important de l'énoncé et dépend de l'information que le locuteur veut faire passer à son interlocuteur.

« On pourrait parler ici d'accent d'insistance : un mot reçoit un supplément d'intensité au détriment des autres mots de la phrase. Tout énoncé anglais peut ainsi recevoir presque autant de sens différents qu'il comporte de mots » (Keen et Adamczewski, 1973 : 189).

Pour désigner l'accent de phrase, on parle également de focus, notion définie par Rossi (1999 : 205) comme l'« *item porté au premier plan par l'accent de focalisation* ». Selon Hirst (1998 : 66), ce terme a remplacé le terme plus traditionnel d'emphase, bien que les deux termes ne désignent pas tout-à-fait la même réalité : tandis que « *focus* » est une notion syntagmatique, qui désigne un élément de la séquence, « *emphase* » est une notion paradigmatique, qui désigne chaque élément comme emphatique ou non emphatique. Cette notion s'applique mieux à l'anglais britannique selon Hirst.

Dans la phrase « *She was in the bedroom* », l'information apportée (le rhème) concerne le lieu où la personne dont on parle se trouve, c'est-à-dire la chambre à coucher. En revanche, dans la phrase « *She wasn't in the bedroom* », l'information primordiale n'est pas le lieu où la personne

se trouve, mais le fait qu'elle ne s'y trouve plus. En conséquence, c'est le mot « *wasn't* » qui porte l'accent de phrase. L'accentuation doit en effet être limitée au point nouveau portée par le message. L'accent de phrase se matérialise par un pic ou un sommet intonatif : dans la phrase « *She was in the bedroom* », ce pic correspond à la syllabe « *bed* » (figure n°56 et **extrait sonore n°24**) et dans la phrase « *She wasn't in the bedroom* », il correspond à la syllabe « *wasn't* » (figure n°57 et **extrait sonore n°25**).

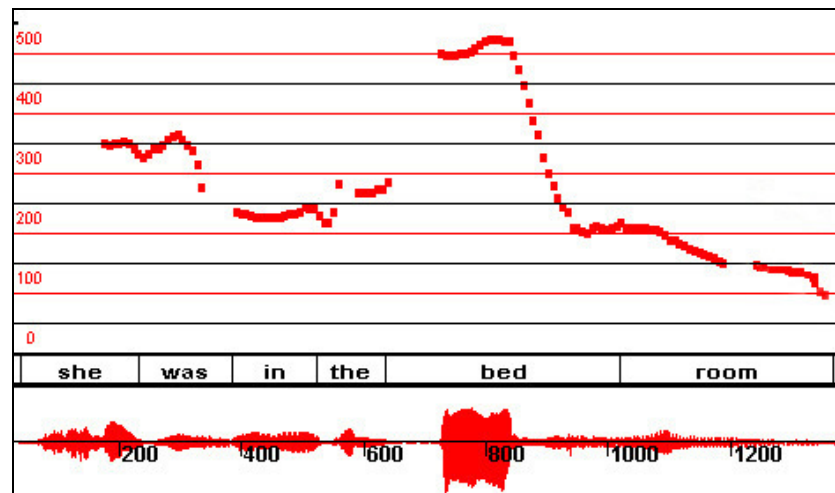


Figure n°56 : courbe mélodique de la phrase « *She was in the bedroom* » prononcée par une locutrice anglophone (logiciel Winsnoori).

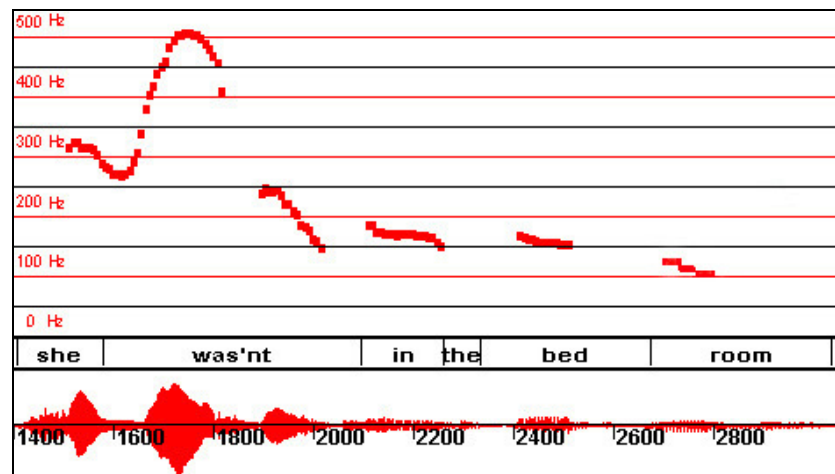


Figure n°57 : courbe mélodique de la phrase « *She wasn't in the bedroom* » prononcée par une locutrice anglophone (logiciel Winsnoori).

On constate également un accroissement de la durée de ces syllabes, mais l'indice le plus

saillant semble bien être l'indice mélodique.

2.2.1.2. L'accent de phrase en français

L'accent de phrase se manifeste différemment en anglais et en français, puisque nous avons vu qu'en anglais, il affecte le mot le plus important, le mot sur lequel il faut insister. Il se rapproche donc plutôt de la notion d'accent d'insistance en français. Mais, contrairement à l'anglais, dans un contexte neutre, le français ne privilégie pas une syllabe se trouvant à l'intérieur du groupe rythmique, mais bien la syllabe finale du groupe rythmique, qui sera marquée par un allongement de la durée et un glissando final. On observe pourtant un sommet intonatif, qui correspond au point le plus haut de l'énoncé.

Sa localisation n'est cependant pas dictée par le besoin de mettre en relief un mot particulier au sein de l'énoncé, mais bien par la structure de cet énoncé.

2.2.1.3. La notion de pivot

Le pivot ou le centre intonatif de l'énoncé est le point central de l'énoncé en anglais, son centre intonatif, appelé également « *nucleus* ». Le pivot se place sur le dernier ton accentué, c'est-à-dire sur la « *syllabe accentuée du dernier mot pleinement accentué de l'énoncé* » (Ginésy, 1995 : 204). Ce ton est en général mélodique ou constitué d'une combinaison de tons statiques suivie d'une remontée de la fréquence fondamentale sur les syllabes atones suivantes.

Cette suite de syllabes atones localisée après le pivot constitue la queue de l'énoncé ou « *tail* » en anglais. L'énoncé doit comporter un pivot pour être complet, sinon, l'auditeur a l'impression qu'il a été interrompu, qu'il est resté en suspens. La première syllabe accentuée de l'énoncé est appelée tonique de tête (ou « *head* ») en anglais. Elle doit supporter une rupture mélodique suffisamment importante pour être distinguée des syllabes atones voisines. Les syllabes atones précédant la tête constituent ce qu'on appelle la « *pre-head* » en anglais. Toutes les syllabes qui se trouvent entre la tonique de tête et le pivot constituent le corps de l'énoncé (« *body* », en anglais). Leur nombre est indéterminé, il peut même ne pas y en avoir.

2.2.2. Les tons mélodiques

2.2.2.1. La richesse musicale de l'anglais

Un ton mélodique consiste en une variation de la hauteur sur la même syllabe, associée à une augmentation de la durée et accessoirement, à un accroissement de l'intensité. La variation de hauteur est un élément essentiel pour assurer la mise en avant des syllabes accentuées en anglais. Ainsi, les tons mélodiques sont généralement utilisés pour mettre en relief le pivot. Il existe cinq tons mélodiques différents en anglais, dont trois qu'on pourrait qualifier de simples, car

leur mouvement mélodique consiste en une montée ou une descente. On les désigne par leur forme mélodique, c'est-à-dire « petite chute », « grande chute » (« High Jump ») et « montée » (voir figure n°58). La petite chute traduit le caractère final d'un énoncé et l'objectivité du locuteur par rapport à cet énoncé (ton de l'assertion).

Mais, lorsque cette chute est mise en relief par une pente beaucoup plus marquée, ce n'est plus l'objectivité qui est exprimée, mais l'implication ou la volonté d'insistance du locuteur : la grande chute est en effet utilisée pour l'exclamation, l'ordre, l'indignation... A l'inverse du pattern de descente, celui de montée traduit le caractère non final, suspensif d'un énoncé (inachèvement d'une phrase, appel à la réponse à une question de type oui ou non). Il est également utilisé pour exprimer un sous-entendu ou l'implication du locuteur par rapport à son interlocuteur (amitié et chaleur humaine, encouragement, politesse, courtoisie).

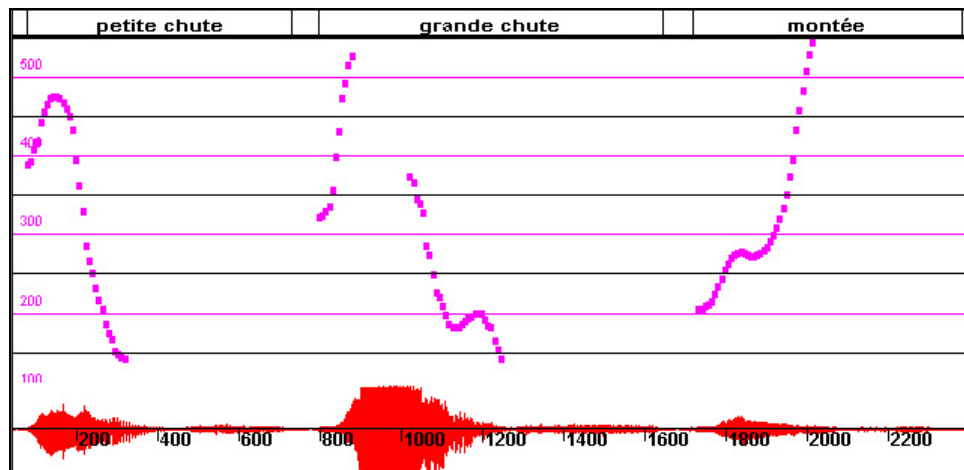


Figure n°58 : les 3 tons mélodiques simples de l'anglais sur le mot « Yes »
prononcée par une locutrice anglophone (logiciel Winsnoori)

La combinaison de ces traits donne naissance à des tons plus complexes, que sont le ton creusé (« *fall-rise* ») et le ton circonflexe (« *rise-fall* ») (voir figure n°59). Ces tons complexes se trouvent toujours en position terminale d'énoncé. Le ton creusé « *traduit d'abord une affirmation (au travers de la chute) suivie de réserves* » (Ginésy, 1995 : 213). Alors que la partie descendante est assertive, cette affirmation est nuancée par la remontée (implication, sous-entendu, non-dit). A l'inverse, le ton circonflexe

« *marque d'abord une incertitude de la part du locuteur (c'est la montée), incertitude vite effacée par l'affirmation portée par la chute. Ce ton est plus rare que le fall-rise en anglais et marque*

souvent l'irritation ou l'impatience ou l'ironie du locuteur. Mais il peut aussi marquer une chaleur, un intérêt véritables pour le sujet ou l'interlocuteur » (Ginésy, 1995 : 213).

L'**extrait sonore n°26** nous donne la suite des cinq tons mélodiques dans l'ordre où nous les avons présentés, réalisés par une locutrice anglophone.

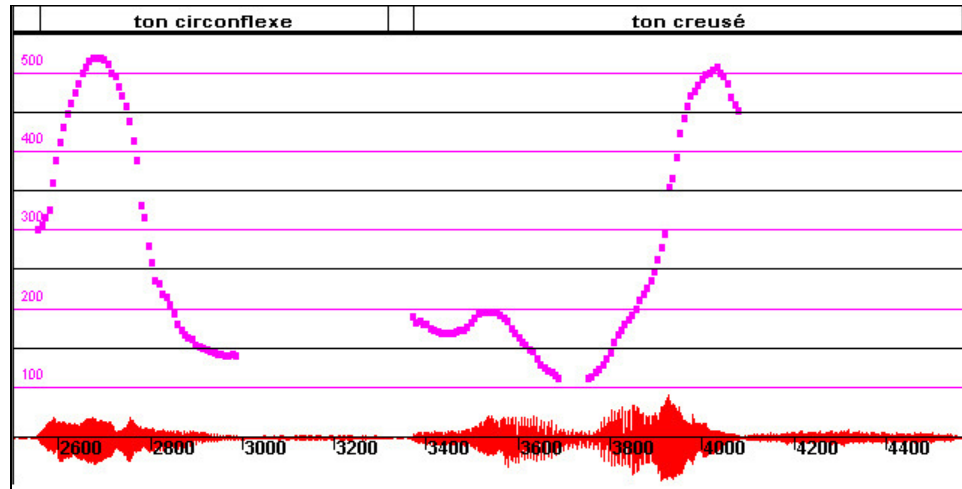


Figure n°59 : les 2 tons mélodiques complexes de l'anglais sur le mot « Yes » prononcée de manière très emphatique par une locutrice anglophone (logiciel Winsnoori).

2.2.2.2. Des tons mélodiques en français ?

Les locuteurs francophones modulent très rarement la hauteur sur une même syllabe et ont le plus souvent recours aux tons statiques, à cause de la grande tension articulatoire qui caractérise leur langue maternelle. Cette tension articulatoire empêche la modulation de la hauteur sur les syllabes, en plus de contraindre à la régularité syllabique. On observe ainsi une succession de tons statiques (paliers horizontaux) de même durée : pour une oreille étrangère, l'effet produit est proche du staccato en musique et ressemble au bruit d'une mitrailleuse. Pour cette raison, les anglais parlent de « machine gun effect » (effet mitrailleuse). Faure (1962) affirme qu'« *on peut représenter toutes les nuances mélodiques en se limitant à quelques tons statiques et à trois tons mélodiques* » :

Un ton mélodique descendant, comme dans « *Viens avec nous* ».

Un ton mélodique descendant, comme dans « *C'est absolument ridicule !* ».

Un ton mélodique ascendant, comme dans « *Il y a de l'argent !* ».

Il n'existe donc pas de tons complexes. Mais, en réalité, l'utilisation des tons mélodiques en français est surtout liée à l'accent d'insistance, qui marque l'emphase, l'affectivité (c'est le cas dans les deux dernières phrases). Le procédé de transcription adopté par Faure (1948 : 119) révèle de manière flagrante la différence entre les deux langues (différence de durée affectant les syllabes et tons mélodiques vs tons statiques) :

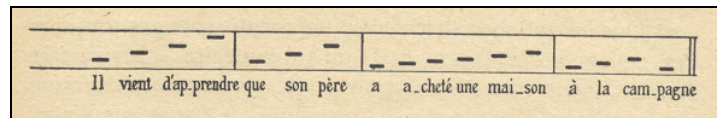


Figure n°60 : transcription de la phrase « Il vient d'apprendre que son père a acheté une maison à la campagne » (Faure, 1948 : 117).

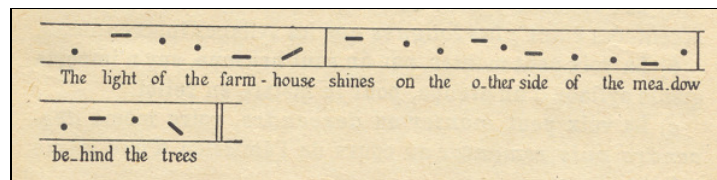


Figure n°61 : transcription de la phrase « The light of the farmhouse shines on the other side of the meadow behind the trees » (Faure, 1948 : 117).

2.2.3. Description comparée de l'intonation déclarative en anglais et en français :

Nous nous sommes limitée à la description du pattern d'intonation déclarative car le niveau en anglais des enfants est débutant. Dans cette étude, nous n'avons donc pas besoin de connaître toutes les ressources du système intonatif de l'anglais (se reporter aux descriptions de Faure (1962), Ginésy (1995), Hirst (1998)).

2.2.3.1. Finalité

En anglais, le schéma qui correspond à l'assertion est appelé « *glide down* ». Par exemple, dans la phrase « *I didn't expect you to know her* » (phrase empruntée à Ginésy, 1995 : 208), le pivot «

know » porte l'accent mélodique (**extrait sonore n°27**). Le ton mélodique associé est la petite chute (voir figure n°62) : il exprime la neutralité et traduit souvent la certitude ou l'achèvement. C'est une intonation terminale utilisée dans l'affirmation simple, pour un contenu informatif, sans sous entendu, ni emphase.

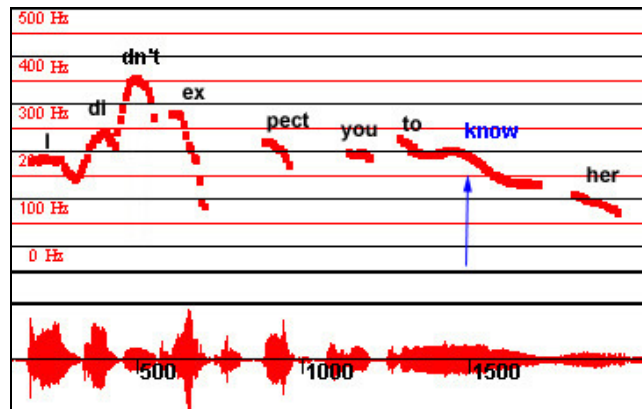


Figure n°62 : Modalité neutre en anglais : « petite chute » produite sur le pivot « know »

Le sommet intonatif de l'énoncé porte sur la première syllabe accentuée (« *didn't* » dans notre exemple), c'est-à-dire la tête de l'énoncé, mais aussi sur le mot qui porte l'information la plus importante. A sa suite, la ligne mélodique s'abaisse progressivement (« *glide* ») jusqu'au pivot, matérialisé par une petite chute (« *down* ») (effet global de déclinaison).

En français, la finalité est également traduite par une courbe mélodique descendante, mais sa structuration est différente.

« La phrase énonciative, objective, contient un sommet intonatif (le point le plus haut de la phrase autour duquel s'articulent les différents groupes qui la composent. La ligne générale de l'intonation est descendante, mais chaque groupe contient l'intonation suivante : ligne presque plate (légèrement montante par paliers statiques) et fin de groupe montante. [...] Les montées françaises ne se font pas à l'aide de tons mélodiques, mais par le biais d'un rejet vers le haut de la syllabe accentuée » (Ginésy, 1995 : 223).

Dans la phrase correspondante à celle de l'anglais, « *Je ne savais pas que tu la connaissais* », on observe un pattern circonflexe, constitué d'une montée et d'une descente (voir figure n°62). Selon Hirst et Di Cristo (1998), le pattern de hauteur typique d'une phrase déclarative finit par sur une hauteur basse finale, contenant une montée à la fin de chaque groupe accentuel.

2.2.3.2. Continuation

Nous avons vu dans le paragraphe 2.1.1.3. (la régularité syllabique du français) qu'il existait une autre proéminence en français, en dehors de l'accent final, c'est l'accent secondaire. Cet accent est marqué non point par l'allongement de la durée syllabique, mais par un indice mélodique et plus précisément par un sommet mélodique. Il permet de délimiter à droite une unité plus petite que le groupe rythmique, le pied métrique. Ainsi, notre phrase est constituée d'un seul groupe rythmique « *Je ne pensais pas que tu la connaissais* » mais de deux pieds

métriques. Le premier pied métrique, « *Je ne pensais pas* » est marqué par une ligne mélodique légèrement concave, suivie d'une montée sur la dernière syllabe « *pas* », qui correspond à l'accent secondaire. Le second pied métrique, « *que tu la connaissais* » repart après un décrochement et se termine par un mouvement mélodique descendant, qui traduit l'assertion (Lhote, 1995 : 139).

C'est donc le mouvement mélodique qui marque la continuation en français, ce qu'avait relevé Delattre dès 1961 : l'intonation de la dernière syllabe d'un groupe de sens indique la continuation en prenant essentiellement la forme ascendante en français. Dans le cas où l'énoncé n'est composé que d'un seul groupe accentuel, l'intonation est globalement descendante.

« Ce qui caractérise le plus le français, c'est la descente graduelle du deuxième groupe et l'inversion de pente entre les deux groupes » (Lhote, 1995 : 140).

Par ailleurs, le sommet intonatif de l'énoncé correspond à la dernière syllabe du dernier groupe accentuel interne de l'énoncé. Dans notre exemple, ce sommet correspond donc à la syllabe « *pas* ».

En anglais, le pattern mélodique qui exprime la continuation constitue une sorte de « négatif » de celui du français : la syllabe accentuée se trouvant en position initiale, la régulation s'effectue donc à partir du début du groupe rythmique (régulation initiale). Toute la tension nécessaire à la montée mélodique qui caractérise la syllabe accentuée en anglais se relâche lors du passage à la syllabe suivante. Après la rapidité de ce saut mélodique, la descente est toute aussi brusque. Le contour mélodique est donc descendant et de forme convexe : il ressemble à un S renversé (Delattre, 1965). Sur la figure n°63 (Delattre, 1965), on peut voir que ce pattern s'oppose à celui du français, ce dernier étant marqué par un contour montant de forme concave, qui aboutit sur un plateau élevé à la dernière syllabe.

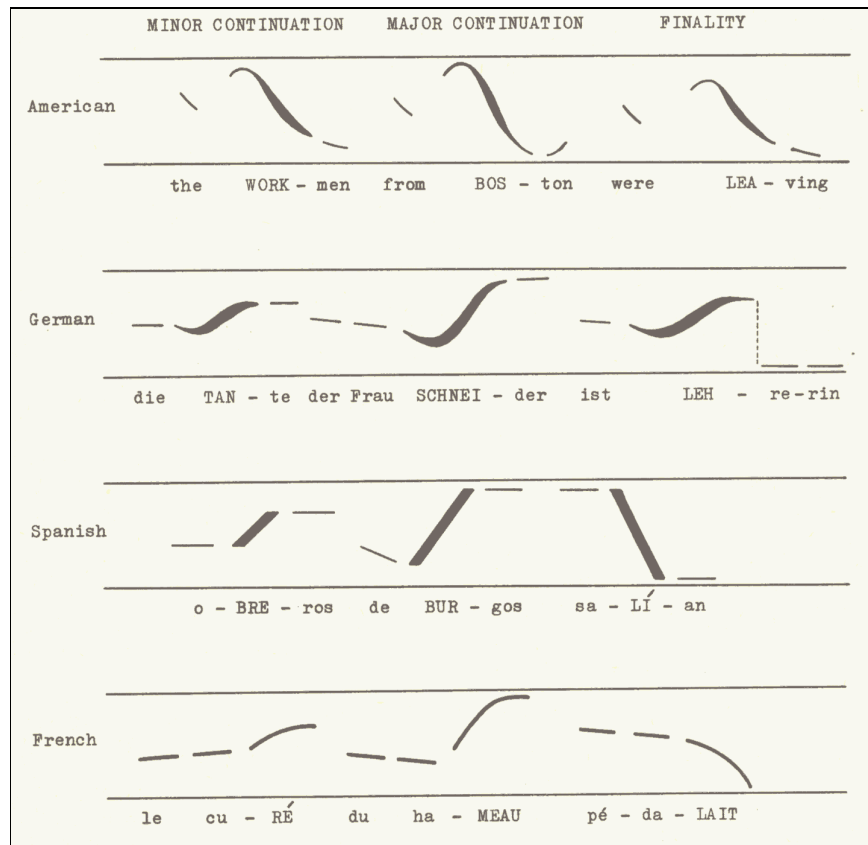


Figure n°63 : intonation caractérisant la continuation et la finalité en français et en anglais américain, Delattre (1965 : 25)

Selon Hirst et Di Cristo (1998), le paramètre de localisation de la tête (la syllabe métriquement la plus forte) dans le pied métrique est déterminant et permet de distinguer l'anglais du français sur la base d'une opposition : tête à gauche / tête à droite, respectivement pour l'anglais et le français.

Un tableau récapitulatif des principales différences prosodiques entre le français et l'anglais figure dans le Volume II, annexe n°3.1.

3. Description comparée du système vocalique de l'anglais et du système vocalique du français

3.1. Caractéristiques générales

3.1.1. Conséquences de la différence d'organisation rythmique sur le niveau segmental

Nous avons vu que l'anglais possédait une organisation rythmique radicalement différente de celle du français. Or, ce cadre rythmique influence directement la prononciation du niveau segmental, et plus particulièrement celle des voyelles.

En anglais, l'organisation rythmique se fait autour du retour régulier de l'accent tonique. La conséquence de cette rythmicité sur la prononciation des voyelles anglaises est déterminante. Les syllabes accentuées sont en effet beaucoup plus audibles que les syllabes atones, car elles sont mises en relief par les paramètres acoustiques que sont l'amplitude (leur intensité est plus forte), la fréquence fondamentale (correspondant à un ton mélodique ; à cause de ces sommets mélodiques, Faure (1962) va jusqu'à qualifier l'anglais, de façon volontairement provocante, de langue à tons) et quelquefois la durée (allongement, notamment en A.N.A.). Le timbre des voyelles « toniques » a le temps de se réaliser pleinement puisque ces voyelles constituent le noyau des syllabes accentuées.

En revanche, ce n'est pas le cas du timbre des voyelles « atones », constituant le noyau des syllabes non accentuées. N'ayant pas le temps de se réaliser, elles ont tendance à perdre leur articulation pleine et à se centraliser. Selon Delattre (1965), cette position centrale constitue la caractéristique distinctive de l'anglais, puisque 90 % des voyelles non accentuées sont proches de /ə/ et que la proportion de ces voyelles parmi toutes les voyelles de l'anglais est de 61 %. En résumé, le timbre d'une voyelle accentuée sera donc toujours pleinement réalisé tandis que celui d'une voyelle atone se centralisera ; une voyelle accentuée sera par ailleurs plus longue, plus intense et plus haute qu'une voyelle atone.

L'organisation rythmique de l'anglais a donc d'énormes conséquences sur la prononciation des voyelles. Elle est d'autant plus difficile à maîtriser pour un francophone que l'organisation rythmique de sa LM est opposée à celle de l'anglais. En français, la régularité n'est pas fondée sur le retour régulier de l'accent, mais sur le retour régulier de la syllabe (cf. 2.1.3.1.). En raison de cette régularité syllabique, le timbre des voyelles du français est toujours clair. Il y a bien un

[ə] caduc, mais

« le schwa a un statut de phonème en français et son occurrence n'est pas une simple conséquence d'un autre phénomène tel que l'accent » (Ostiguy et al, 1999 : 151).

A cause du régime de tension croissante du français (voir 2.1.3.3.), l'articulation est très tendue, et empêche la diphtongaison, typique de l'anglais. La diphtongaison s'explique en effet par une mobilité des articulateurs pendant l'émission. Lorsque les cavités supra-glottiques changent de configuration, elles provoquent un changement de la fonction de transfert, et en conséquence un changement de la qualité du timbre.

Le phénomène de diphtongaison dépend donc fortement de la durée. Les voyelles longues, nous le verrons, ont tendance à se diphtonguer. Quant aux noyaux vocaliques complexes que sont les diphtongues et les triphthongues, ils figurent parmi les noyaux les plus allongés. Enfin, l'anglais se caractérise par une certaine lenteur à ouvrir et fermer les articulateurs se traduisant au niveau acoustique par de longues transitions entre les cibles vocaliques, ce qui n'est pas le cas du français.

Les problèmes de centralisation, de durée et de diphtongaison sont donc très dépendants de l'organisation rythmique de la langue et on ne peut étudier le système vocalique d'une langue sans avoir au préalable établi son cadre rythmique.

3.1.2. Différences d'organisation de l'espace articulatoire

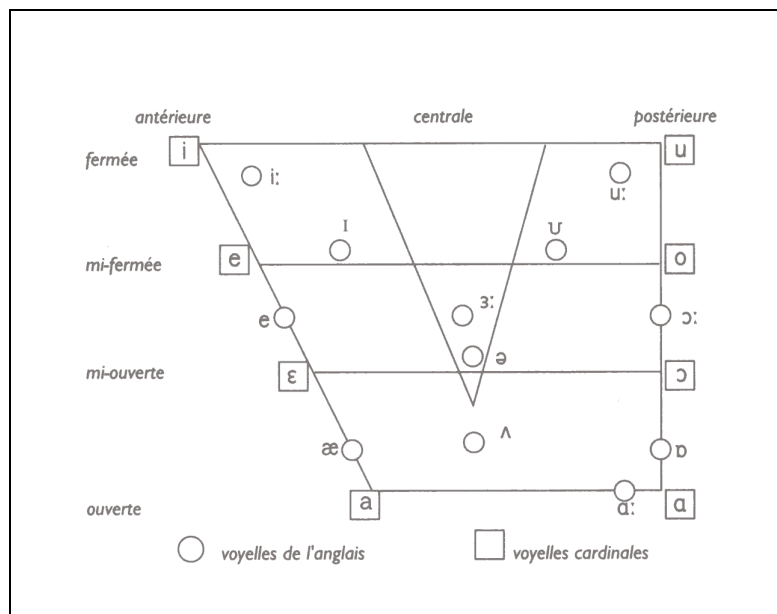


Figure n°64: trapèze vocalique comparatif des voyelles françaises et des monophthongues anglaises
(emprunté à Ginésy, 1995 : 15).

Un français qui va apprendre l'anglais sera confronté à un système vocalique beaucoup plus riche et plus complexe que celui de sa LM. En français, le système vocalique est composé de

seize monophthongues, dont quatre voyelles oralo-nasales.

En anglais britannique standard (désormais R.P.), on trouve vingt-cinq noyaux vocaliques, dont douze monophthongues (et parmi elles, cinq voyelles diphtonguées), huit diphtongues et cinq triphthongues.

A titre de comparaison, l'A.N.A. (désormais A.N.A.) ne possède que seize noyaux vocaliques (Lehiste et Peterson, 1960), dont dix monophthongues et six noyaux complexes (trois « glides » et trois diphtongues).

Au niveau de l'articulation, le français et l'anglais combinent différemment les traits articulatoires et leur système vocalique occupe différemment l'espace articulatoire (figure n°64, trapèze vocalique comparatif).

Le système vocalique du français, délimité à ses extrémités selon l'axe vertical par la voyelle fermée /i/ et la voyelle ouverte /a/ montre une grande amplitude d'aperture. En raison de leur « extrémité articulatoire », ces voyelles « cardinales » ont été utilisées comme points de référence par Daniel Jones pour établir des chartes vocaliques. En raison de l'asymétrie des organes articulatoires,

« la mâchoire s'écarte beaucoup plus pour les voyelles d'avant que pour les celles d'arrière puisqu'elle pivote sur un axe postérieur » (Léon, 1992 : 80).

Le français joue beaucoup de cette amplitude d'aperture car la plupart de ses voyelles sont articulées sur l'avant de la bouche : deux séries antérieures sur trois, soit dix voyelles sur seize (cette configuration donne au français une « *résonance vocalique claire* », Léon, 1992 : 82).

Au contraire, l'anglais se caractérise par une plus grande proportion de voyelles postérieures (deux séries sur trois, soit huit monophthongues sur douze) et donc une amplitude d'aperture plus réduite. Malgré cela, la proportion de voyelles dites ouvertes est plus importante qu'en français et le centre de gravité, plus bas (Delattre, 1965). On remarque également plus de centralisation (trois voyelles centralisées /əɜːʌ/).

Le français est dominé par l'articulation labiale, puisque deux tiers des voyelles sont labialisées, soit onze sur seize (Delattre, 1965). La proportion s'inverse en anglais où l'on ne trouve qu'une

série labialisé et pas de série antérieure labialisée, équivalente à la série /iyø/ en français. Par ailleurs, le degré de labialisation est beaucoup plus fort en français, les lèvres étant véritablement arrondies, alors qu'en anglais, on n'observe qu'une légère projection des lèvres.

Enfin, on trouve en français quatre phonèmes vocaliques très fortement nasalisés. On parle d'ailleurs de voyelles oralo-nasales, en raison d'un fort degré de couplage entre la cavité orale et la cavité nasale. En R.P, le trait de nasalité ne permet pas d'opposer des voyelles entre elles ; il n'est donc pas phonologique, mais phonétique. Pour cette raison, nous n'en ferons pas la description, car il ne peut pas y avoir d'interférences entre les deux langues à ce sujet.

3.1.3. Conséquences pour un apprenant francophone

3.1.3.1. Prédiction d'une typologie des fautes découlant de ces différences

L'anglais se démarque donc principalement du français par sa tendance à la diphtongaison, par l'utilisation qu'il fait de la durée et par son articulation plus postérieure, plus centralisée et moins labialisée. En fonction de ces différences, on peut prédire le type de fautes de prononciation qu'un francophone réalisera en anglais.

L'excès de tension articulatoire que le locuteur francophone transpose de sa LM à l'anglais est à l'origine de toute une série de fautes :

- elle va entraver la **tendance à la diphtongaison** des voyelles longues et la reproduction correcte des diphtongues et des triphthongues spécifiques de l'anglais ;
- elle va amener l'apprenant à produire des **transitions trop rapides entre les cibles vocaliques** qui composent une diphtongue, avec pour conséquence une tendance à produire une succession de plusieurs voyelles au timbre stable, plutôt qu'un glissement lent et progressif d'une cible vocalique à une autre. Or, une « vraie » diphtongue est perçue comme un glissement vers la seconde cible et non pas comme la succession de deux voyelles ;
- elle va l'amener à **raccourcir la durée des syllabes**, ce qui aggrave les deux défauts

précédents. En effet, il faut une certaine durée pour réaliser un son diphtongué ;

- elle va l'empêcher de reproduire correctement la **réduction vocalique qui caractérise les voyelles non accentuées de l'anglais**, car celui-ci se caractérise par un grand relâchement articulaire.

Les différences articulatoires entre le système vocalique du français et de l'anglais amènent l'apprenant à faire les erreurs suivantes :

- Le français étant une langue dont l'articulation est antérieure, l'apprenant aura le réflexe de **produire les voyelles de l'anglais trop en avant**, surtout si leur timbre est proche de certaines voyelles françaises plus antérieures. C'est le cas de [ɪ] par exemple qui aura tendance à être prononcé « à la française », comme [i] ou [e].
- Les voyelles françaises étant pour la plupart labialisées, l'apprenant aura tendance à **trop labialiser les voyelles anglaises** dont le timbre se rapproche de ces voyelles labialisées. Par exemple, la voyelle [ʊ] de « *book* », qui n'est que faiblement arrondie en anglais, aura tendance à être prononcée comme [u] français, et à être en conséquence, fortement labialisée.
- Les problèmes d'interférences naissent lorsque le timbre des sons non natifs est proche de celui de certaines voyelles de la LM (crible phonologique). Dans ce cas, elles sont analysées comme des exemplaires de sons natifs, éventuellement légèrement déviants. Bien que l'anglais possède un système vocalique très différent de celui du français, certaines de ses voyelles non diphtonguées ont un timbre proche de celui de certaines voyelles du français. Les catégories natives et non natives se chevauchant partiellement, il risque en conséquence d'y avoir **beaucoup d'interférences au niveau perceptif**.

3.1.3.2. Paramètres importants nécessaires à la description de l'anglais pour un apprenant

francophone

3.1.3.2.1. Durée

Nous venons de voir que l'excès de tension articulatoire allait, entre autres choses, amener l'apprenant francophone à produire des noyaux vocaliques trop courts. Or, la durée des noyaux vocaliques est un facteur très important pour décrire les différents noyaux vocaliques de l'anglais. Ainsi, Lehiste et Peterson (1960) ont décrit le système vocalique de l'A.N.A.⁸ en fondant en grande partie leur classement sur ce critère de durée. Ils distinguent parmi les monophthongues, les noyaux simples et courts (monophthongues d'une durée de 180 à 200 ms), des noyaux simples et longs (monophthongues d'une durée, de 240 à 260 ms).

⁸ Analyse acoustique de 1263 monophthongues (70 mots prononcés par un locuteur de l'A.N.A.).

Il est probable que la différence entre ces deux catégories n'apparaîtra pas dans les productions d'un locuteur francophone, car les noyaux longs seront trop réduits. Par ailleurs, l'effet de crible phonologique risque de renforcer cette tendance. En effet, les voyelles anglaises auront tendance à être assimilées aux voyelles françaises de timbre « correspondant ». Par exemple le [i:] anglais aura tendance à être assimilé à la voyelle française [i] et produit comme un [i]. Or, la durée de cette voyelle est beaucoup plus courte que celle de son homologue anglais. Il apparaît donc pertinent de relever la durée absolue et la durée relative (par rapport à la durée totale de la syllabe) des noyaux vocaliques produits par un locuteur francophone en anglais. Elle fournira en effet des informations sur les effets du crible prosodique (et en particulier rythmique) et sur les effets du crible phonologique.

3.1.3.2.2. Structure interne du noyau vocalique

En plus de cette différence de durée intrinsèque qui affecte le noyau vocalique, les travaux de Lehiste et Peterson (1960) révèlent l'importance de la durée relative des événements composant ce noyau. Toute voyelle est constituée théoriquement d'une tête (« *on glide* »), d'une partie stable (« *target* ») et d'une queue (« *off glide* »), la relation entre les trois événements lors de la durée totale du noyau vocalique restant constante, indépendamment de la consonne suivante. Selon les auteurs, le temps mis pour produire un changement total de timbre (durée relative par rapport à la durée totale de la syllabe) pourrait constituer un critère pertinent pour dresser une typologie fine des différents types de noyaux vocaliques en anglais (1960 : 292). Or l'excès de tension articulaire qui caractérise les productions francophones va justement affecter la durée du changement de timbre et bouleverser la structure interne du noyau vocalique. En effet, la durée des phases de transition sera réduite, par rapport à la durée de la zone de stabilité de timbre, qui elle sera démesurée. Cette trop grande stabilité constitue un obstacle à la production des nombreux sons diptongués de l'anglais, ainsi que des voyelles longues.

Il paraît donc pertinent d'étudier la structure interne des noyaux vocaliques produits par les locuteurs francophones en anglais. Ce critère pourrait servir d'indice pour évaluer la progression de leur maîtrise de la prononciation.

3.1.3.2.3. Timbre

Enfin, il est important de déterminer le timbre des noyaux vocaliques produits par les apprenants francophones en anglais pour savoir s'ils sont soumis au phénomène de crible phonologique et si oui, pour quels types de sons. Mais, comment mesurer le timbre d'une voyelle ? Traditionnellement, les chercheurs pensent que l'information essentielle pour déterminer la qualité des voyelles réside dans la localisation des maxima spectraux correspondant aux trois premiers formants (Lonchamp, 1983 ; Schwartz, 1987). Cette localisation correspond à la notion de cible vocalique. D'après Lehiste et Peterson (1960 : 290),

« the time interval within the syllable nucleus where the formants are parallel to the time axis has been considered as the extent of a vowel target ».

L'« extension » de la cible vocalique » dont il est question correspond à la partie stable de la voyelle et elle se mesure habituellement sur une représentation spectrographique en délimitant la zone où les formants sont parallèles à l'axe du temps. Cette stabilité spectrale reflète une position articulaire stable de la langue et des lèvres. Or, la notion de cible n'a de pertinence qu'au travers de la perception (Johnson et al., 1993 ; Ladefoged, 1997) : en effet, la cible qu'essaie d'atteindre le locuteur en produisant une voyelle est avant tout définie par des propriétés auditives. Diverses procédures de mesure permettent de relever les valeurs absolues de F1, F2 et F3 en Hz en un point précis, généralement au centre de la partie stable de la voyelle. Mais, est-il vraiment judicieux d'utiliser des valeurs absolues lorsqu'on travaille du point de vue de la perception ? En effet, la corrélation entre la localisation des trois premiers formants et sa couleur perceptive est atténuée par un certain nombre de facteurs, notamment la variabilité intra-individuelle (co-articulation, variations dans le discours selon les émotions, la situation...) et inter-individuelle (différences physiologiques, socioculturelles entre individus, stratégies individuelles de production...). Les résultats de Peterson et Barney (1952) sont très

souvent cités pour illustrer la variabilité inter-locuteurs, notamment entre les voix d'enfants, de femmes et d'hommes :

« the first formants for the children are seen to be about half an octave higher than those of the men, and the second and third formants are also appreciably higher » (Peterson et Barney, 1952 : 132).

Mais cette variabilité est en grande partie éliminée si les voyelles sont caractérisées, non plus en termes de fréquences absolues, mais en termes de rapports entre leurs formants. Ainsi, les rapports de fréquences entre le premier et le second formant, et entre le second et le troisième formant pourraient servir à éliminer les différences entre les locuteurs (Peterson, 1961). A la fin du XIXe siècle, Lloyd a formulé une théorie baptisée la « *formant-ratio theory* », relatée par Miller dans un article traitant de la normalisation perceptive (1989 : 2115) :

« Like articulations produce like perceptions of vowel qualities and like articulations produce like ratios of the formants. He called his theory « the relative resonance theory » and stated that the vowel quality depends on intervals between the resonances, not on their absolute values ».

Lloyd fonde notamment sa théorie sur le rôle central des rapports de fréquence en musique. Selon Miller (1989), des travaux paraissent de temps en temps depuis les travaux de Lloyd, mais ne reconnaissent que rarement la paternité de ce dernier. Potter et Steinberg (1950⁹), dans une étude concernant les voyelles de l'A.N.A. prononcées par 10 hommes, 10 femmes et 5 enfants, ont confirmé l'hypothèse que la différence inter-locuteurs était éliminée si les voyelles étaient caractérisées par leurs rapports entre leurs trois premiers formants. Sur la figure n°X, on peut voir que pour chaque catégorie de voyelle, les rapports entre F3 et F2 (M3/M2, ligne du dessus) et entre F2 et F1 (M2/M1, ligne du dessous) sont identiques. Ces rapports sont exprimés en mel, mais Potter et Steinberg (1950) soulignent que les rapports de fréquence éliminent la variabilité inter-locuteurs exactement de la même façon que le font les mel. Cependant, cette méthode ne permet pas de distinguer les voyelles [ɑ] et [ɔ] et les voyelles [ʊ] et [u].

⁹ Travaux cités par Miller (1989 : 2115-2116).

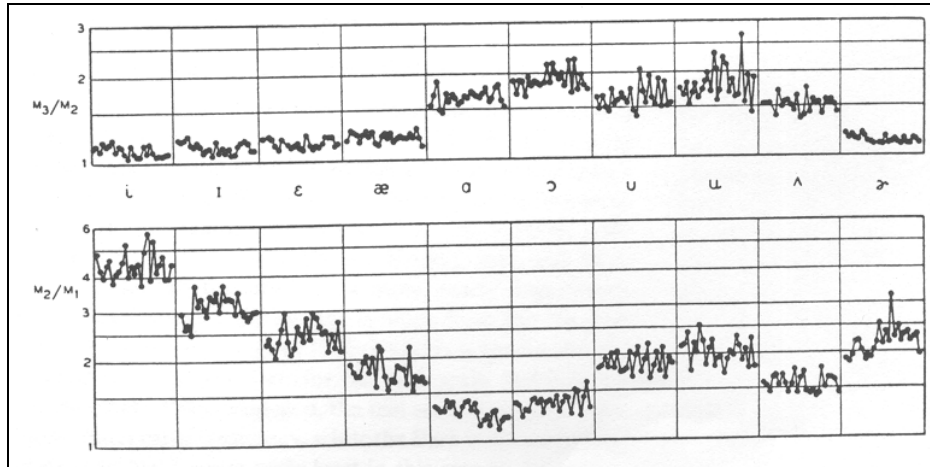


Figure n°65 : rapports $F3-F2$ (au dessus) et $F2-F1$ (en dessous) en équivalent mel
(emprunté à Miller, 1989 : 2116, figure n°13 de Potter et Steinberg, 1950).

Cette hypothèse a été confirmée par Peterson (1961) pour les rapports de fréquence (qui correspondent aux intervalles musicaux). Pour cet auteur, la force majeure des rapports de formants réside dans leur capacité à éliminer ou à réduire de manière drastique les différences relatives au locuteur, à l'âge et au sexe dans la description acoustique des voyelles. En revanche, pour certaines voyelles, les rapports sont très similaires et il est possible que d'autres facteurs que les rapports interviennent pour permettre leur différenciation. Les travaux précurseurs de Lloyd et les travaux suivants suggèrent donc tous que les rapports entre les trois premiers formants d'une voyelle jouent un rôle déterminant dans son interprétation auditive et perceptive. Une charte vocalique n'étant pas conçue pour représenter les rapports entre les trois premiers formants, nous avons mis au point une nouvelle représentation, la « charte des rapports de formants » (pour une description de cette représentation, se reporter au Volume II, annexe n°3.2.).

3.1.3.3. Nécessité d'une description plus poussée

Nous ferons donc la description du système de l'anglais en fonction du timbre de ses noyaux vocaliques, mais également de leur durée (absolue et relative) ainsi que de leur structure interne. En ce qui concerne le timbre, nous disposons des données de Peterson et Barney (1952) qui

servent de référence en la matière (valeur standard des trois premiers formants pour 33 hommes, 28 femmes et 15 enfants), mais elles concernent l'A.N.A. Or, lors de l'apprentissage de l'anglais, le modèle de prononciation donné aux enfants est celui de la R.P., c'est-à-dire de l'accent britannique standard (« BBC English »).

En ce qui relève de la durée, les données de Lehiste et Peterson (1960) ne concernent malheureusement que l'A.N.A. Or, l'A.N.A. se différencie de la R.P., notamment par un système vocalique moins riche. Sauf pour le timbre, la littérature ne nous fournit pas de description suffisante correspondant aux indices que nous voulons étudier pour la R.P.

Nous avons donc décidé :

- a) d'établir un corpus qui nous servira dans un premier temps à décrire le système vocalique de la R.P. selon nos propres critères ;
- b) de réaliser une batterie de tests exploitant les principales interférences entre les systèmes vocaliques de l'anglais et du français (voir Volume II, annexe 4.2.).

3.1.4. Réalisation d'une étude descriptive

3.1.4.1. Locutrices

Deux locutrices ont été sélectionnées, une locutrice anglaise (native speaker), professeur d'anglais à l'Université, et une locutrice française, étudiante. La prononciation de la locutrice anglophone est standard (R.P.). Quant à la locutrice francophone, elle a été choisie en fonction de son absence d'accent régional, et ce, malgré son origine franc-comtoise. Sa prononciation est standard.

3.1.4.2. Matériel de parole

Pour décrire le système vocalique de l'anglais par rapport à celui du français, le matériel de parole suivant a été enregistré :

- Pour l'anglais, une liste de 23 mots monosyllabiques de type CVC et de type CV a été enregistrée par la locutrice anglophone. Ces mots ne diffèrent que par leur noyau vocaliques et couvrent la quasi totalité des noyaux vocaliques de l'anglais (8

diphtongues, 3 triphthongues et 11 monophthongues). La voyelle [ə] et les deux triphthongues [aʊə] et [əʊə] ayant une occurrence très faible dans la langue ont été écartées.

- Pour le français, une liste de 24 mots monosyllabiques de type CVC et de type CV a été enregistrée par la locutrice francophone. Là-aussi, les mots ne diffèrent que par leur noyau vocalique. Ils couvrent les 11 monophthongues orales du français. La voyelle [ə] et les 4 voyelles nasales (en raison de l'absence de nasales en anglais) ont été écartées.

En anglais et en français, les mots CVC et CV choisis commencent par la constrictive sourde [s] et les mots CVC finissent par l'occlusive sourde [t]. Nous avons retenus les deux types de syllabes afin de respecter les structures syllabiques de l'anglais : syllabes CVC pour les monophthongues et CV pour les diphtongues et les triphthongues. L'agencement des phonèmes respecte les règles combinatoires de l'anglais. En revanche, pour le français, la liste a été conçue en fonction des interférences possibles avec les voyelles anglaises de timbres proches. De fait, certaines combinaisons retenues n'existent pas en français (comme [set] par exemple).

3.1.4.3. Conditions d'enregistrement

Pour une qualité optimale, les enregistrements ont été réalisés en chambre sourde au Laboratoire de Phonétique de Besançon, avec un magnétophone D.A.T. « Aiwa HD-S1 » et un microphone « Aiwa ». Les locutrices ont pris tout le temps nécessaire pour se préparer et s'approcher de la prononciation la plus standard possible. Lorsqu'elles estimaient avoir échoué, les enregistrements étaient refaits. La consigne leur était donnée de garder le même schéma prosodique et de lire la liste de mots en gardant un débit constant (quelques mots supplémentaires figuraient en début de liste pour que la locutrice s'habitue et en fin de liste afin d'éviter la baisse de l'intonation correspondant à la fin de l'énonciation).

3.1.4.4. Analyses

Pour la description de la longueur des différents noyaux vocaliques produits par la locutrice francophone et par la locutrice francophone, la durée absolue de chaque noyau vocalique a été

relevée (en ms), ainsi que la durée absolue de la syllabe contenant le noyau (en ms). Cette dernière mesure permettra de calculer la durée relative du noyau vocalique au sein de la syllabe (exprimée en pourcentage).

Pour la description de la structure interne des noyaux vocaliques, 47 fichiers ont été édités¹⁰ pour chacun des noyau vocalique des 47 mots prononcés par les deux locutrices. Ces fichiers comportent le suivi des trois premiers formants sur la durée totale de chaque noyau vocalique. Ils ont ensuite été traités par le programme informatique réalisé dans le cadre de l'étude n°5 (pour une description de la méthode et une description détaillée de la conception du programme, se reporter au Volume II, annexe n°11.1.). Ce programme fournit la durée absolue (en ms) et relative (exprimée en pourcentage) de la durée des différents événements composant le noyau vocalique, c'est-à-dire la transition initiale, la zone de stabilité de timbre et la transition finale. Au préalable, chaque fichier a été vérifié manuellement, de façon à ce que les résultats fournis par le programme soient totalement fiables.

Enfin, pour l'étude du timbre, la fréquence des trois premiers formants a été prélevée par le programme au milieu de la zone de stabilité de timbre. Une fois ces valeurs transposées en huitièmes de tons¹¹, les intervalles entre F2 et F1 (F2-F1) et entre F3 et F2 (F3-F2) ont été calculés. Puisque le timbre des voyelle est présenté en terme de rapports de formants, la représentation classique de type « charte vocalique » ne pourra pas être utilisée.

3.2. Caractéristiques détaillées

¹⁰ Fichiers édités à l'aide du logiciel Winsnoori, logiciel distribué par Babel Technologies (<http://www.babeltech.com>), mis au point par Yves Laprie, LORIA, Nancy. Version utilisée v.1.2. (1998-1999). Actuellement, la dernière version du logiciel est la version 1.3.

¹¹ Selon la formule $S = (\log(X) - \log(220)) / (\log(2) / 48)$, X étant la valeur en Hz à convertir et 220, la fréquence de référence.

3.2.1. Les noyaux vocaliques simples de l'anglais et du français

3.2.1.1. Durée

En anglais R.P., il existe 12 monophthongues que nous qualifierons de voyelles simples (Ginésy, 1995 : 16). On peut les répartir en deux catégories en prenant pour critère leur durée intrinsèque : 7 voyelles brèves [ɪ], [e], [æ], [ʌ], [ʊ], [ɒ] (**extrait sonore n°28**) et 5 voyelles longues [i:], [ɜ:], [u:], [ɔ:] et [ɑ:] (**extrait sonore n°29**).

3.2.1.1.1. Les voyelles brèves

Lehiste et Peterson (1960) ont montré que les noyaux simples et « brefs » de l'A.N.A. ont une durée de 180 à 200 ms. Dans notre corpus, **la durée absolue des 7 noyaux étudiés n'excède pas 209 ms** pour la voyelle [æ], ce qui correspond aux données de Lehiste et Peterson.

Quant à leur **durée relative**, celle-ci est **courte** (de 25 à 29 % de la syllabe, voir tableau n°5).

En ce qui concerne la répartition des différents événements composant le noyau vocalique, c'est-à-dire la phase de transition initiale (tête), la partie stable et la phase de transition finale (queue), on remarque :

- que la **durée de la partie stable est longue** (de 75 % à 91 % de la durée du noyau, voir exemple de la voyelle [e], figure n°66), sauf dans le cas de la voyelle [ɒ], dont la partie stable, plus courte (58,33 %), occupe tout de même plus de la moitié du noyau. En outre, plus les voyelles sont postérieures, plus leur partie stable est réduite. C'est notamment le cas des voyelles [ɒ] et [ʊ] ;
- la **présence d'une tête dans tous les noyaux** avec une durée relative approximative de 20 %. Cette durée est cependant plus courte pour les voyelles [ɪ] (12,78 %) et [æ] (8,82 %) ;

- **l'absence de queue dans la plupart des noyaux**, sauf dans le cas des voyelles [ɪ] (9,02 %) et [ɒ] (16,11 %). Néanmoins, pour ces deux voyelles, la durée relative de la queue reste courte.

	Durée de la tête (DT)	Durée de la partie stable (DPS)	Durée de la queue (DQ)	Durée de la voyelle par rapport à la syllabe
[ɪ]	12,78 %	78,19 %	9,02 %	24,47 % (139 ms)
[e]	20 %	80 %	0 %	24,62 % (180 ms)
[æ]	8,82 %	91,17 %	0 %	29,52 % (209 ms)
[ʌ]	22,29 %	77,70 %	0 %	25,31 % (163 ms)
[ʊ]	25,15 %	74,84 %	0 %	25,11 % (169 ms)
[ɒ]	25,55 %	58,33 %	16,11 %	27,39 % (186 ms)

Tableau n°5 : durée relative des différents événements composant les 7 noyaux vocaliques brefs de la locutrice anglophone (exprimée en pourcentage).

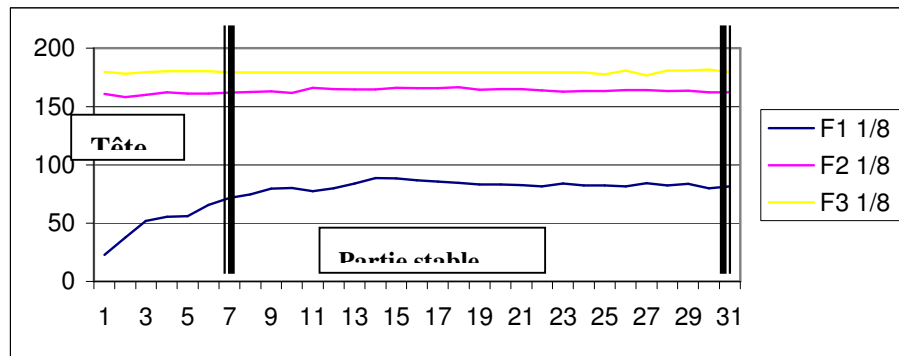


Figure n°66 : évolution des trois premiers formants de la voyelle brève [e] dans le mot « set » prononcé par la locutrice anglophone (exprimée en 1/8^{ème} de tons) ; la partie stable s'étend de la mesure n°7 à la mesure n°31.

En résumé, les voyelles brèves de notre corpus se caractérisent par une durée absolue (<210 ms) et relative réduite (25 à 29 % de la durée de la syllabe), une grande partie stable (plus de 50 % de la durée de la voyelle), la présence d'une tête (approximativement 20 % de la durée de la syllabe) et une queue presque toujours absente ou alors très réduite.

3.2.1.1.2. Les voyelles longues

D'après Lehiste et Peterson (1960), les noyaux vocaliques simples et longs de l'A.N.A. se caractérisent par une durée absolue comprise entre 240 et 260 ms. Dans notre corpus, la durée

absolue des 5 noyaux étudiés se répartit entre 233 et 325 ms (voir tableau n°6), la voyelle [i:] étant la plus courte (180 ms). La durée relative du noyau vocalique, exprimée en pourcentage de la durée totale de la syllabe, est plus importante pour les voyelles longues que pour les voyelles brèves (aux environs de 35 %), à cause de leur plus grande durée intrinsèque sauf dans le cas de la voyelle [i:]. Cette dernière se caractérise par une grande phase de stabilité (63,79 %) qui la rapproche des voyelles brèves. Cependant, elle se démarque de la voyelle brève la plus proche de timbre [ɪ] par une phase de transition initiale beaucoup plus longue (33,33 % contre 12,78 %). Bien que la voyelle [ɜ:] ait une durée plus longue que la voyelle [i:], elle se caractérise également par une grande phase de stabilité (59,42 %).

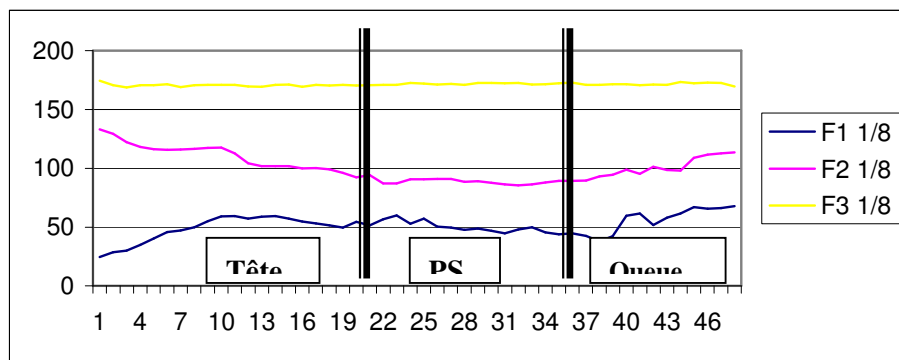
En revanche, les trois autres voyelles longues, c'est-à-dire [u:], [ɔ:] et [ɑ:] se caractérisent par une forte instabilité (respectivement 74,44 %, 72,15 %, 61,74 % de transition si on ajoute la phase de transition initiale à la phase de transition finale). La phase de transition finale constitue le segment le plus long dans les voyelles [ɑ:] et [u:] (respectivement 51,09 % et 56,38 %). A l'inverse, dans la voyelle [ɔ:] (voir figure n°67), c'est la phase de transition initiale qui couvre la durée la plus grande (48,71 %).

La grande instabilité qui caractérise ces voyelles s'explique par la tendance à la diphtongaison qui caractérise la production des sons vocaliques de longue durée en anglais. Cette tendance provient directement de l'organisation rythmique de l'anglais (cf. 3.1.1.). Cette instabilité semble plus particulièrement affecter les voyelles postérieures.

En résumé, il existe une assez grande différence entre les voyelles longues antérieures et les voyelles longues postérieures. Ces dernières se caractérisent par une grande phase d'instabilité (supérieure à 60 % de la voyelle), une grande durée absolue (de 279 à 325 ms) et une grande durée relative (de 32,63% à 42,09 % de la durée totale de la syllabe). Les caractéristiques des voyelles [i:] et [ɜ:] se caractérisent au contraire par une grande phase de stabilité et une répartition des événements qui les rapprochent des voyelles brèves.

	Durée de la tête (DT)	Durée de la partie stable (DPS)	Durée de la queue (DQ)	Durée de la voyelle par rapport à la syllabe
[i:]	33,33 %	63,79 %	2,87 %	27,90 % (180 ms)
[a:]	10,65 %	38,24 %	51,09 %	42,09 % (325 ms)
[ɔ:]	48,71 %	27,83 %	23,44 %	36,66 % (279 ms)
[u:]	18,06 %	25,55 %	56,38 %	32,63 % (233 ms)
[ɜ:]	23,96 %	59,42 %	16,61 %	35,68 % (319 ms)

Tableau n°6 : durée relative des événements composant les 5 noyaux vocaliques longs de la locutrice anglophone.



3.2.1.1.3. Les voyelles du français

A cause de sa rythmique, le français se caractérise par une grande tension articulatoire qui empêche toute tendance à la diphtongaison. En conséquence, les voyelles ont un timbre précis, c'est-à-dire qu'il ne change pas en cours d'émission, et ce, même si elles se trouvent dans une syllabe inaccentuée.

Si nous laissons de côté les voyelles oralo-nasales et la voyelle [ə], le français comporte 11 voyelles orales, [i], [e], [y], [ɛ], [a], [ø], [u], [o], [ɔ], [ɑ] et [œ].

La durée absolue du noyau vocalique n'excède pas 139 ms pour les voyelles les plus « brèves » du français, c'est-à-dire les voyelles [i], [e], [ɛ], [a], [ɔ], [u] et [œ]. Cette durée est plus courte que pour les voyelles brèves de l'anglais dont la durée s'étend jusqu'à 209 ms (tableau n°7). La durée relative du noyau au sein de la syllabe n'excède pas 21,95 %, ce qui est également plus court que pour les voyelles brèves de l'anglais (25 à 29 %). On note une grande stabilité de timbre (de 52,25 à 100 % ; voir pour la voyelle [ɔ], figure n°68), comme pour les voyelles brèves de l'anglais (de 58 à 91 %).

La durée absolue des voyelles fortement labialisées comme [o] et [ø] est plus longue (respectivement 180 et 203 ms), mais, même avec cette longueur, ces voyelles sont aussi longues que les voyelles brèves de l'anglais. Leur durée relative (respectivement 31,51 % et 35,97 %) est également beaucoup plus importante, comme celle des voyelles longues de l'anglais, mais elles sont beaucoup plus stables que ces dernières ; on relève une plus grande stabilité de timbre (zone de stabilité de 67,77 % pour [o] et de 91,13 % pour [ø]), en raison de l'articulation fortement tendue du français.

La voyelle [ɑ] est la plus longue des voyelles produites par la locutrice francophone, que ce soit du point de vue de sa durée absolue (301 ms), que de sa durée relative (46,79 % de la durée totale de la syllabe). On remarque cependant une zone de stabilité plus courte (42,52 % de la durée totale de la voyelle), au profit d'une grande transition initiale (51,82 %). On relève une répartition quasiment identique pour la voyelle [a] (55,17 % de stabilité et transition initiale de 55,17 %), bien que sa durée absolue et sa durée relative soient courtes (respectivement 116 ms et 18,59 % de la durée totale de la syllabe).

Mais, dans tous les cas, sauf pour [a] et [ɑ], la partie stable est toujours le segment le plus long du noyau vocalique. Des voyelles comme [i], [y], [o], [ø], [œ] se caractérisent notamment par une phase de stabilité qui couvre presque la totalité de leur durée (plus de 85 % de zone de stabilité par rapport à leur durée totale). Les voyelles orales du français se caractérisent donc par grande stabilité et une durée absolue généralement plus courte (< 203 ms) que les voyelles brèves de l'anglais, exception faite de la voyelle [ɑ].

Leur durée relative avoisine les 20 % ce qui est également plus court que les voyelles brèves de l'anglais. Les voyelles les plus fermées se caractérisent par une plus grande durée relative (de 31,51 à 46,79 % de la durée totale de la syllabe).

	Durée de la tête (DT)	Durée de la partie stable (DPS)	Durée de la queue (DQ)	Durée de la voyelle par rapport à la syllabe
[i]	0 %	100 %	0%	14,70 % (111 ms)
[e]	7,91 %	71,22 %	20,86 %	20,13 % (139 ms)
[ɛ]	8,27 %	57,14 %	34,58 %	21,95 % (133 ms)
[a]	55,17 %	35,34 %	9,48 %	18,59 % (116 ms)
[y]	6,89 %	93,10 %	0 %	15,70 % (87 ms)
[ø]	2,95 %	91,13 %	5,91 %	35,97 % (203 ms)
[œ]	6,89 %	86,20 %	6,89 %	18,19 % (87 ms)
[u]	0 %	68,18 %	31,81 %	18,68 % (110 ms)
[o]	16,11 %	67,77 %	16,11 %	31,51 % (180 ms)
[ɔ]	20,72 %	52,25 %	27,02 %	19,18 % (111 ms)
[ɑ]	51,82 %	42,52 %	5,64 %	46,79 % (301 ms)

Tableau n°7 : durée relative des événements composant les 11 monophthongues de la locutrice francophone.

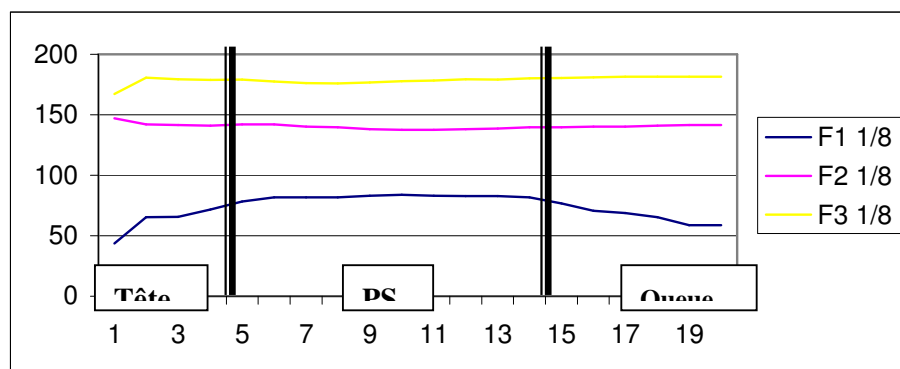


Figure n°68 : évolution des trois premiers formants (en 1/8^{ème} de ton) de la voyelle [ɔ] dans le mot « sotte » prononcé par la locutrice francophone ; la partie stable s'étend de la mesure n°5 à la mesure n°15.

3.2.1.1.4. Interférences de durée du français vers l'anglais

D'après nos résultats (pour un récapitulatif, voir tableau n°8), les voyelles brèves de l'anglais ont généralement une durée absolue et relative plus longue que celle des voyelles françaises. Par ailleurs, bien qu'elles se caractérisent par une stabilité certaine, cette stabilité reste moins marquée qu'elle ne l'est dans les voyelles françaises. Cependant, malgré ces différences, leurs structurations temporelles sont semblables et cette similitude risque de favoriser les phénomènes d'assimilation entre voyelles anglaises et voyelles françaises de timbre proche.

En cas d'assimilation, il est probable qu'un locuteur francophone produise les voyelles brèves de l'anglais avec la stabilité et la durée qui caractérisent les voyelles françaises. Dans ce cas, la mesure de la durée absolue et relative de la voyelle, ainsi que la mesure de la zone de stabilité fourniront des indices sur le poids du crible phonologique chez l'apprenant.

La durée absolue et la durée relative des voyelles longues de l'anglais sont proches de celles de la voyelle française la plus longue, la voyelle [ɑ]. Elles constituent donc des sons très longs pour un locuteur francophone, qui aura tendance à les raccourcir et de ce fait, à accroître leur stabilité. Le phénomène de crible devrait être moindre pour les voyelles longues que pour les voyelles brèves de l'anglais car ces sons ont des caractéristiques « non natives » plus saillantes, donc plus repérables par l'apprenant. Cependant, si malgré tout une assimilation entre voyelles de timbre correspondant se produit, les français vont transposer la tension articulaire aux voyelles longues de l'anglais, qui se caractériseront par un excès de stabilité. Selon Faure (1948 : 79),

« les Français ont généralement le défaut de ne pas allonger assez les voyelles longues et les diphtongues, à la fin des mots ou devant une consonne sonore, et, donc, de prononcer des mots comme : sea, see, car, door, d'une façon trop brève [...]. Ils doivent tendre [...] à leur donner une longueur beaucoup plus considérable qu'ils ne le font d'ordinaire ».

Langues	Type de voyelles	Caractéristiques
Anglais	7 voyelles brèves [ɪæʌʊɒə]	DA < 209 ms 25 % < DR < 29 % Grande PS (de 58 à 91 %) > T ou Q Petite tête (de 8,82 à 20 %) Absence de queue, sauf pour [ɪ] et [ɒ].
	5 voyelles longues ou « diphtonguées » [i:ɑ:ɔ:u:ɜ:]	233 < DA < 325 ms (180 ms pour [i:]) DR = 35 % (sauf pour [i:]) PS < T ou Q, sauf pour [i:] et [ɜ:]
Français	8 voyelles brèves [iɛɛauɔœə]	DA < 139 ms DR < 21,95 % Grande PS (de 52,25 à 100 %) > T ou Q Tête, de à 20,72 % Queue, de à 34,58 % (plus longue pour [œuɔ])

	3 voyelles longues [oø]	180 < DA < 203 ms 31,51% < DR < 35,97 % Tête et queue réduites.
	[a]	DA=301 ms DR=46,79 % PS (42,52 %) < T (51,82 %), comme pour [a] (PS=35,34 % et T=55,17 %).

Tableau n°8 : récapitulatif des caractéristiques de durée des différents types de voyelles de l'anglais et du français (DA pour durée absolue, DR pour durée relative, PS pour partie stable, T pour tête et Q pour queue).

3.2.1.2. Timbre

Les différents types d'interférences

Certains contrastes de la langue cible dits contrastes non-natifs (désormais NN) sont absents dans la LM, mais montrent une similitude de timbre avec certains contrastes natifs (désormais N). En raison du phénomène de crible phonologique (voir partie 1.2.), il sera très difficile pour l'auditeur de les différencier et il risque d'assimiler et de produire le contraste NN à un contraste N ayant un timbre proche.

Un autre type d'interférence peut intervenir, lorsque deux voyelles NN, absentes du système vocalique maternel montrent une similitude de timbre. L'auditeur aura alors beaucoup de difficultés à les différencier et les deux contrastes seront généralement assimilés l'un à l'autre, ou assimilés tous les deux à un contraste N de timbre proche (catégorie SC dans la typologie de Best, voir 1.2.).

La graphie peut, elle aussi, favoriser des interférences de timbre. Prenons un mot anglais comme « *lip* ». Dans ce mot, la graphie de la voyelle [ɪ] peut favoriser l'assimilation de ce son avec la voyelle française [i].

La similitude de forme graphique favorise également les interférences. Le mot « *orange* » s'écrit de manière identique et possède une signification similaire dans les deux langues. En fait, il ne diffère que par sa prononciation : [ɔʁɑ̃ʒ] en français et [ˈɒrɪndʒ] en anglais. Dans cet exemple, non seulement la graphie favorise l'assimilation de la voyelle anglaise [ɒ] à la voyelle

française [ɔ] de la même façon que dans l'exemple du mot « *lip* », mais de plus, cette assimilation est renforcée par la similitude de forme entre les deux mots. Dans le cas d'« *orange* », cette similitude est totale, puisqu'il s'agit du même mot et l'apprenant aura tendance à prononcer le mot « *orange* » à la française.

Par ailleurs, deux voyelles NN peuvent se transcrire de la même manière. C'est le cas de la voyelle [ɑ:] dans le mot « *car* » et de la voyelle [æ] dans le mot « *cat* ». L'apprenant aura donc tendance à produire la même voyelle NN, soit [æ], soit [ɑ:] dans les deux mots ou alors il produira à la place la voyelle française [a], qui s'écrit de la même manière et qui possède une similitude de timbre avec les deux voyelles NN.

Enfin, les interférences avec les voyelles françaises peuvent être favorisées lorsque les mots anglais sont familiers et prononcés « à la française » par l'entourage de l'apprenant. Ce phénomène est accentué par le fait que l'anglais est une langue très répandue et les emprunts nombreux. Même les locuteurs francophones qui ne parlent pas anglais connaissent et utilisent des termes anglais. Ainsi, des mots « *book* », « *park* » (qu'on trouve dans le mot français « *parking* »), « *dog* », « *car* », « *shop* »...

Prédiction de ces interférences à partir de l'étude comparée des systèmes vocaliques de la locutrice anglophone et de la locutrice francophone

Les fréquences des trois premiers formants des voyelles composant le système vocalique de la locutrice anglophone et celui de la locutrice francophone ont été relevées (Volume II, annexe n°3.3., tableaux n° 1 et n°2). Après avoir calculé les intervalles séparant F2 et F1 et F3 et F2 pour chaque voyelle, une charte des rapports de formants a été établie pour chaque locutrice (figure n°69 et figure n°70). La comparaison de ces chartes a permis de déterminer les proximités de timbre entre voyelles anglaises et voyelles françaises et donc, les risques

d'assimilation pouvant intervenir entre les deux langues. La figure n°71 propose une vision synthétique de ces interférences.

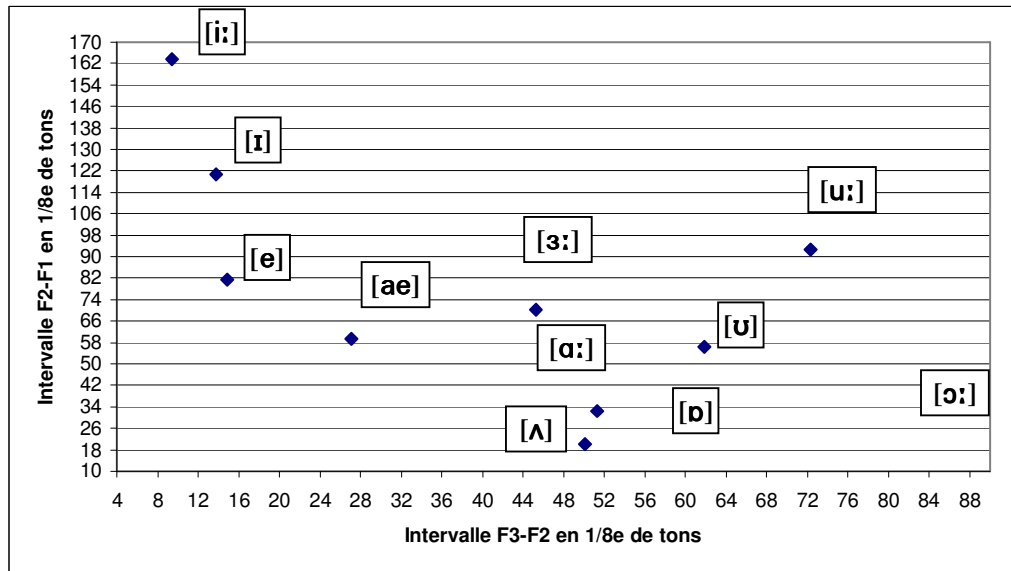


Figure n°69 : charte des rapports de formants de la locutrice anglophone.

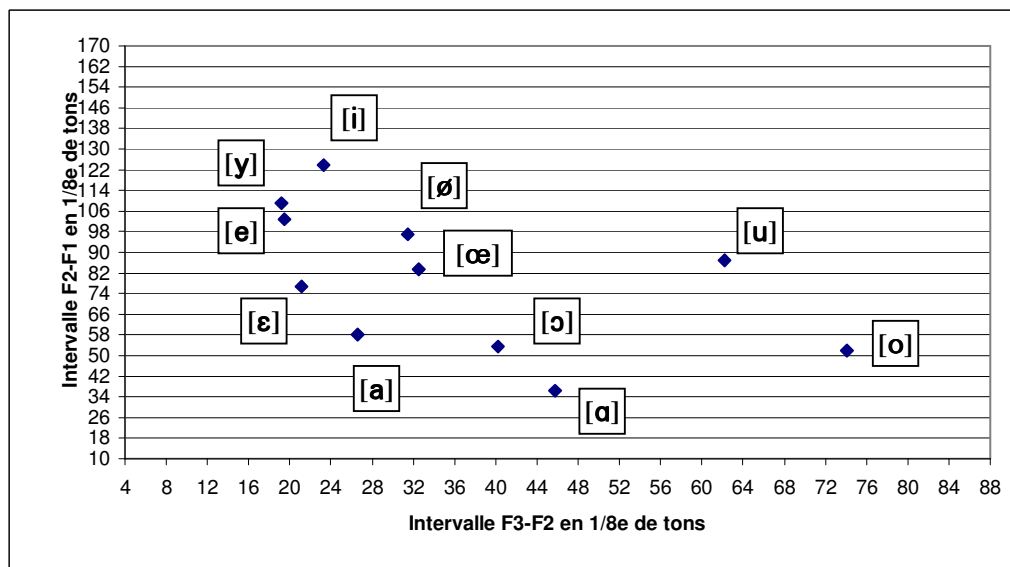


Figure n°70 : charte des rapports de formants de la locutrice francophone.

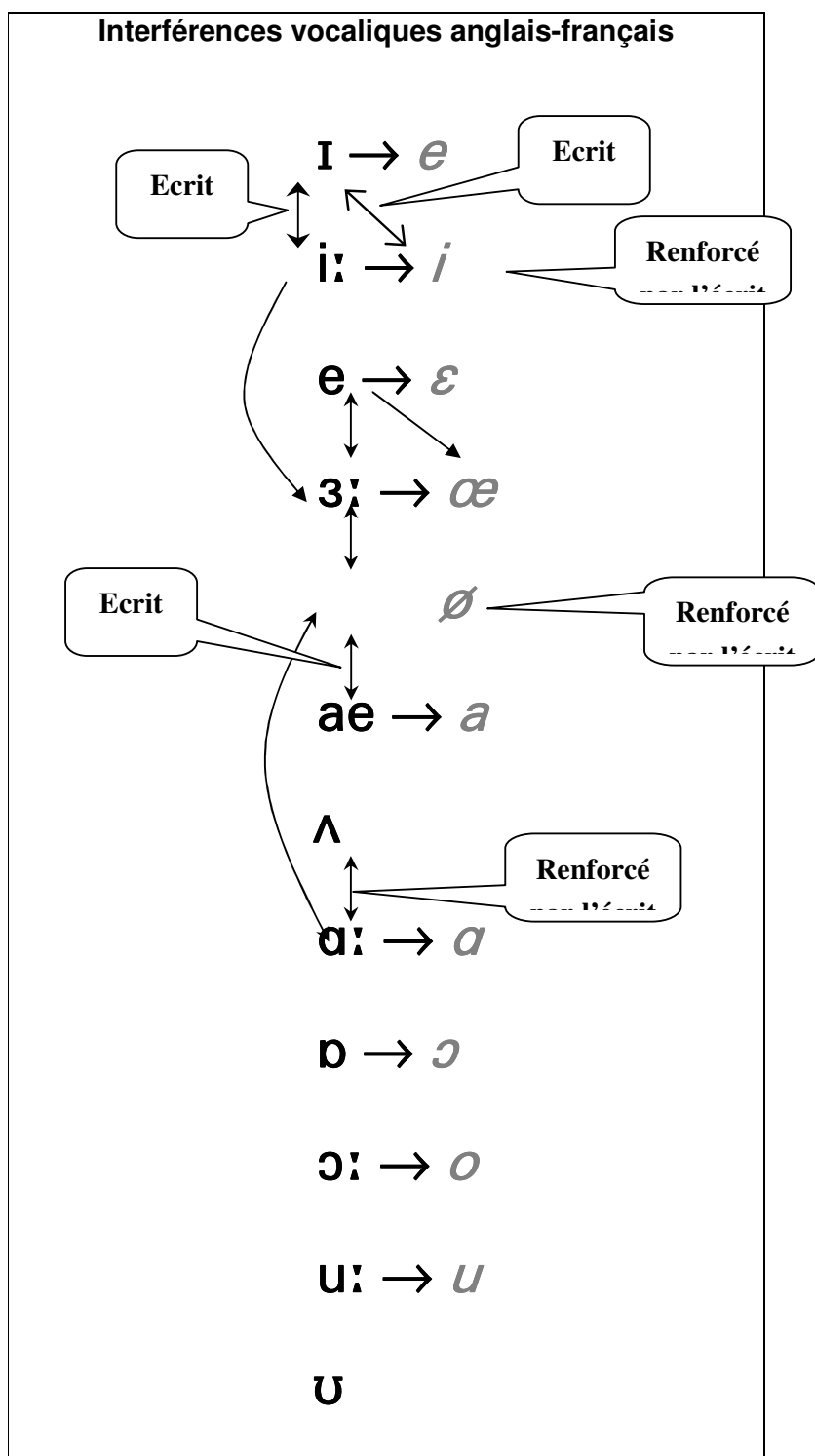


Figure n°71 : relevé des principales assimilations des contrastes non natifs de l'anglais à d'autres contrastes de l'anglais et aux contrastes natifs du français (en gris).

3.2.2. Les noyaux vocaliques complexes en anglais

Rappelons qu'en anglais R.P., il existe 13 noyaux vocaliques complexes dont 8 diphtongues et 5 triphthongues.

3.2.2.1. Les diphtongues

3.2.2.1.1. Définition

Au niveau de la production, on peut définir la diphtongue comme le noyau d'une syllabe contenant deux cibles vocaliques. Bien qu'elle soit transcrite en A.P.I. par un double symbole phonétique ([aɪ] par exemple), elle fonctionne comme une seule unité vocalique appartenant à la même syllabe (Gynésy, 1995 : 34) et les symboles phonétiques utilisés ne correspondent pas à la qualité de la première, ni de la seconde cible :

« They can be described as movements from one vowel to another [...]. Furthermore, contrary to the traditional transcriptions, the diphthongs often do not begin and end with any of the sounds that occur in simple vowels » (Ladefoged, 1975 : 69).

Sur une représentation spectrographique, une diphtongue se caractérise par un changement de trajectoire des trois premiers formants en cours d'émission. Les différents événements qui composent une diphtongue sont, dans l'ordre d'apparition, une phase de transition initiale (désormais TE pour « tête »), une première partie stable (désormais PS1) correspondant à la première cible, une zone de transition centrale (désormais TC), une deuxième partie stable (désormais PS2) correspondant à la deuxième cible et une zone de transition finale (désormais Q pour « queue »). Le tableau n°9 en schématise la structure. Les deux cibles sont clairement identifiables (voir figure n°71, diphtongue [ɔɪ]). Il est rare que les deux segments cibles aient les mêmes caractéristiques de durée et d'intensité. En anglais,

« the first part of the diphthong is usually more prominent than the last. In fact, the last part is often so brief and transitory that is difficult to determine its exact quality » (Ladefoged, 1975 : 69).

La diphtongue commence donc plutôt énergiquement, puis elle s'affaiblit. Pour désigner la partie « proéminente », Catford (1988 : 115-116) parle de « *stress pulse* ». Celui-ci se repère facilement si une représentation de la mélodie est superposée à une représentation spectrographique.

En effet, on observe une correspondance du pic mélodique du contour de Fo (voir figure n°72 pour la diphtongue [ɔɪ] où la flèche indique une correspondance entre le pic de Fo et le premier élément de la diphtongue). Lehiste et Peterson (1960 : 290) ont utilisé cet indice pour effectuer leurs prises de mesure :

« the intonation pattern was often helpful in determining the position of the target, since the peak of the intonation contour associated with the CNC word usually coincided with the target of the syllable nucleus ».

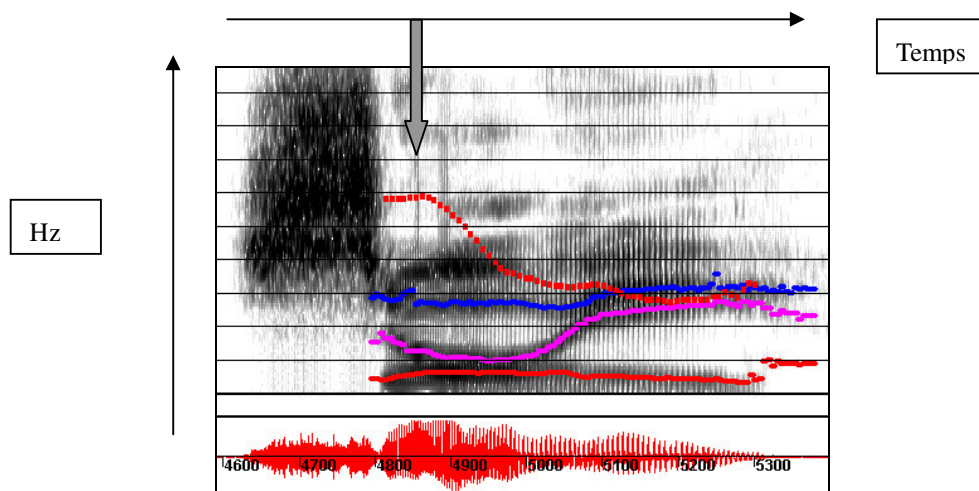


Figure n°72 : représentation spectrographique à bandes larges de la diphtongue [ɔɪ] du non mot « soy » prononcé par la locutrice anglophone avec superposition de l'évolution des trois premiers formants (lignes continues) et du contour de Fo (ligne en pointillés).

Du point de vue de la perception, dans une diphtongue, les deux cibles vocaliques, c'est-à-dire le point de départ et le point d'arrivée ne sont pas perçus comme des voyelles séparées, mais comme un glissement vers le second élément.

« La durée des segments initiaux et finaux et surtout l'angle que forme F2 en direction de la cible avec l'horizontale semblent constituer les deux facteurs de perception les plus importants pour la distinction des diphtongues » (Carton, 1974 : 44).

Dans une étude sur le français parlé à Montréal, Yaeger (1979) a identifié les principaux indices acoustiques permettant d'identifier des allophones diphtongués : une variation de F2 de plus de 200 Hz pour toutes les voyelles ou un changement d'au moins 100 Hz pour F1, même si F2 est stable pour les voyelles postérieures.

TE	PS1	TC	PS2	Q

Tableau n°9 : les différents événements composant la structure d'une diphtongue (TE pour tête, PS1 pour première partie stable, TC pour transition centrale, PS2 pour deuxième partie stable et Q pour queue).

3.2.2.1.2. Description des diphtongues de l'anglais

Il existe différents types de diphtongues en anglais :

- 5 diphtongues fermantes (dont la trajectoire monte dans l'espace vocalique, c'est-à-dire que l'aperture se réduit au cours du temps. C'est le cas pour [aɪ], [eɪ], [ɔɪ], [əʊ] et [aʊ] (**extrait sonore n°31**) ;
- 3 diphtongues centralisantes dont la trajectoire commence dans la périphérie de l'espace vocalique et se dirige ensuite vers son centre. C'est le cas pour [ɪə], [ʊə] et [eə] (**extrait sonore n°32**). Ginésy (1995 : 34) souligne cependant que la diphtongue [eə] n'existe pratiquement plus en R.P.

Cibles et durées

Dans cette description, nous ne séparerons pas les caractéristiques de timbre de celles de durée comme nous l'avons fait pour les voyelles, pour la simple et bonne raison qu'elles sont intimement liées dans le phénomène de diphtongaison. En effet, la durée relative des événements composant une diphtongue dépend essentiellement de la cible à atteindre et tous les gestes articulatoires ne demandent pas le même effort. C'est donc en fonction de la plus ou moins grande proximité de la cible à atteindre, que les gestes articulatoires seront plus ou moins complexes à produire et que la durée nécessaire à leur réalisation sera plus ou moins longue.

Les diphtongues fermantes [aɪ], [eɪ], [ɔɪ], [əʊ] et [aʊ]

La trajectoire des diphtongues fermantes monte dans l'espace vocalique comme on peut le voir sur la charte des rapports de formant (figure n°73), à cause d'une réduction de l'aperture en cours d'émission. Au niveau articulatoire, l'aperture dépend de la position plus ou moins écartée de la mâchoire inférieure et de l'élévation subséquente de la langue par rapport au palais. Au niveau acoustique, une réduction d'aperture se manifeste par la baisse du premier formant (F1), provoquant une augmentation de l'intervalle entre F2 et F1.

La diphtongue [ɔɪ] se caractérise par la trajectoire la plus longue (figure n°73), une grande durée absolue (493 ms), ainsi qu'une grande instabilité (42,32 % de stabilité, en additionnant PS1 et PS2).

Cette instabilité peut s'expliquer par le fait qu'il faut du temps pour passer de la voyelle postérieure et ouverte correspondant à la première cible (proche de [ɔ:]) à la voyelle antérieure et fermée correspondant à la seconde cible (proche de [i:]), deux extrêmes sur la charte des rapports de formants. La diphtongue [eɪ] est également très instable (44,74 % de stabilité). Comme [ɔɪ], elle se caractérise par une grande durée absolue (510 ms) par rapport à des diphtongues comme [əʊ] (278 ms) ou [ʊə] (261 ms), dans le même contexte syllabique (CV).

En revanche, sa trajectoire est moins grande, car les deux cibles sont moins éloignées l'une de l'autre. On relève cependant la plus grande phase de transition centrale (TC) de toutes les diphtongues (31,08 %), après celle de la diphtongue [ɪɛ] (34,71 %). Alors que la diphtongue [ɔɪ] connaît un grand changement d'aperture et un grand changement de lieu d'articulation pendant son émission, la diphtongue [eɪ] ne connaît qu'un changement d'aperture (accompagné il est vrai d'une légère antériorisation, mais moindre par rapport à [ɔɪ]). Une transition centrale plus longue pourrait permettre à l'auditeur de bien faire la différence entre les deux cibles. Il est possible, par ailleurs, qu'il soit plus difficile, au niveau articulatoire, de produire un changement d'aperture non accompagné de changement de lieu d'articulation. Dans la diphtongue [aɪ], la transition centrale est au contraire très brève (6,41 %). Il y a peu de changement d'aperture et la trajectoire se caractérise plutôt par un changement de lieu articulatoire. C'est une diphtongue beaucoup plus stable (52,01 % de stabilité) que les deux précédentes, avec une PS1 plus grande (28,57 %).

Les diphtongues [aʊ] et [əʊ] sont encore plus stables que [aɪ] (respectivement 66,88 % et 80,94 % de stabilité). En effet, dans [aʊ] on relève une grande PS1 (51,01 %) et dans [əʊ], une grande PS2 (57,50 %). Leurs durées absolues sont moins grandes car elles apparaissent dans un contexte syllabique CVC (278 ms et 296 ms). Elles ont toutes les deux le même point de départ, une première cible très centralisée, puis la trajectoire de [aʊ] évolue vers une seconde cible très postérieure (proche de [ɔː]), tandis que la trajectoire de [əʊ], beaucoup moins marquée, évolue vers seconde cible moins éloignée (proche de [uː]). Mais, dans les deux cas, il y a peu de changement d'aperture et comme ces deux diphtongues sont très stables, il semblerait que le taux d'instabilité soit lié au degré d'aperture.

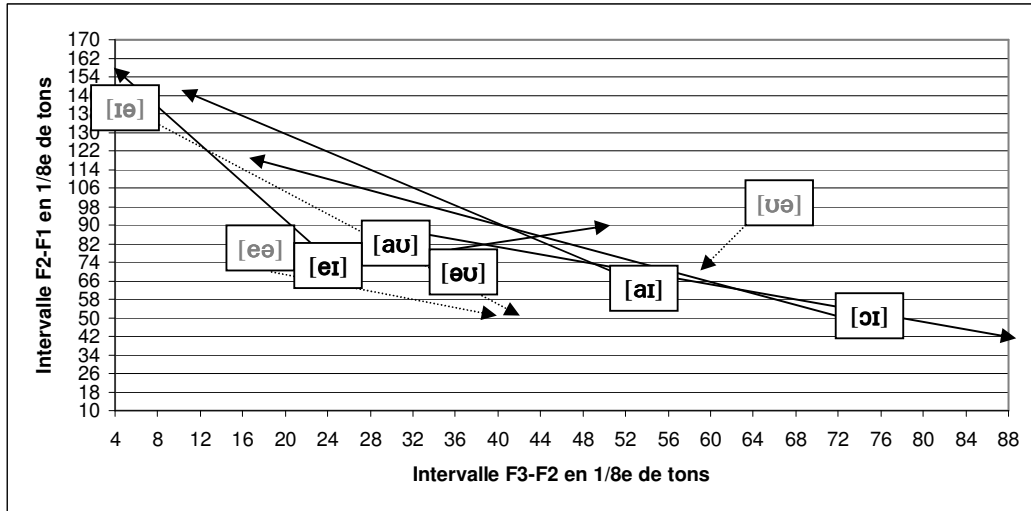


Figure n°73 : trajectoires des diphtongues de la locutrice anglophone

(en gris, diphtongues centralisantes et trajectoires en pointillé ; en noir, diphtongues fermantes et trajectoires en ligne continue).

Les diphtongues centralisantes [ɪə], [ʊə] et [eə]

La diphtongue [ɪə] se caractérise par une grande stabilité (65,28 %), mais également, comme nous l'avons déjà vu, par une forte TC entre les deux cibles (34,71 %). Sa trajectoire s'illustre notamment par un changement d'aperture plus grand que dans la diphtongue fermante [əɪ]. La diphtongue [eə] se caractérise par une très forte stabilité (77,53 %), notamment sur la première cible (49,19 %, ce qui représente quasiment la moitié de la durée totale de la diphtongue). La seconde cible occupe 28,34 % de la diphtongue. Entre ces deux cibles, la TC est réduite (7,08 %). La diphtongue [ʊə] se démarque en revanche par la plus forte instabilité de toutes les diphtongues étudiées (34,1 % de stabilité). Sa durée absolue est également la plus courte (261 ms) alors qu'elle est produite dans une syllabe CV. Sa trajectoire est également très courte et les zones de transition sont donc très marquées, et notamment la tête (20,39 %) et la queue, qui occupent presque la moitié de la durée totale de la diphtongue (43,13 %). En fait, ce son est très difficile à prononcer. La consonne le précédant est la constrictive dentale [s], qui s'articule à l'avant de la bouche. Pour atteindre la première cible de la diphtongue [ʊ], il faut aller à

l'opposé de l'espace vocalique, ce qui explique la longueur de la tête et le fait que la cible soit plus fermée que la voyelle correspondante.

	DT	DPS1	DTC	DPS2	DQ	Durée par rapport à la syllabe
[eɪ]	9,30 %	16,03%	31,08 %	28,71 %	14,85 %	75,78 % (510 ms) CV
[aɪ]	3,29 %	28,57 %	6,41 %	23,44 %	38,27 %	75,20 % (546 ms) CV
[ɔɪ]	13,11 %	23,77 %	15,57 %	19,05 %	28,48 %	73,25 % (493 ms) CV
[əʊ]	8,79 %	23,44 %	1,83 %	57,50 %	8,42 %	33,25 % (278 ms) CVC
[aʊ]	4,05 %	51,01 %	17,56 %	15,87 %	11,48 %	34,18 % (296 ms) CVC
[ɪə]	0 %	28,04 %	34,71 %	37,24 %	0 %	67,84 % (441 ms) CV
[eə]	15,38 %	49,19 %	7,08 %	28,34 %	0 %	72,35 % (500 ms) CV
[ʊə]	20,39 %	16,07 %	2,35 %	18,03 %	43,13 %	63,34 % (261 ms) CV

Tableau n°10 : durée relative des événements composant les 8 diphtongues prononcées par la locutrice anglophone (DT pour durée totale, DPS1 pour durée de la première partie stable, DTC pour durée de la transition centrale, DPS2 pour durée de la deuxième partie stable, DQ pour durée de la queue, CV pour syllabe consonne-voyelle et CVC, pour syllabe consonne-voyelle-consonne).

3.2.2.2. La triphongue

3.2.2.2.1. Définition

Du point de vue de la production, on peut définir une triphongue comme le noyau d'une syllabe contenant trois cibles vocaliques. Bien qu'elle soit transcrite en A.P.I. par un triple symbole phonétique (identifiant le point de départ, la cible à mi-chemin et la cible finale, comme dans [aɪə] par exemple), elle fonctionne comme une seule unité vocalique appartenant à la même syllabe. Les différents événements qui composent une triphongue sont, dans l'ordre d'apparition, une phase de transition initiale (TE), une première partie stable (PS1) correspondant à la première cible, une première zone de transition centrale (TC1), une deuxième partie stable (PS2) correspondant à la deuxième cible (partie stable souvent réduite, voire absente), une deuxième zone de transition centrale (TC2), une troisième partie stable

(PS3) correspondant à la troisième cible et une zone de transition finale (Q). Le tableau n°11 en schématise la structure.

Sur une représentation spectrographique, on observe une trajectoire complexe des trois premiers formants, avec un changement de direction à mi-course (voir figure n°74). Le premier élément de la triphthongue, le plus proéminent, correspond à la localisation du pic mélodique du contour de Fo (voir figure n°74 où le pic de Fo concorde avec le premier élément de la triphthongue [ɔɪə], à l'endroit indiqué par une flèche).

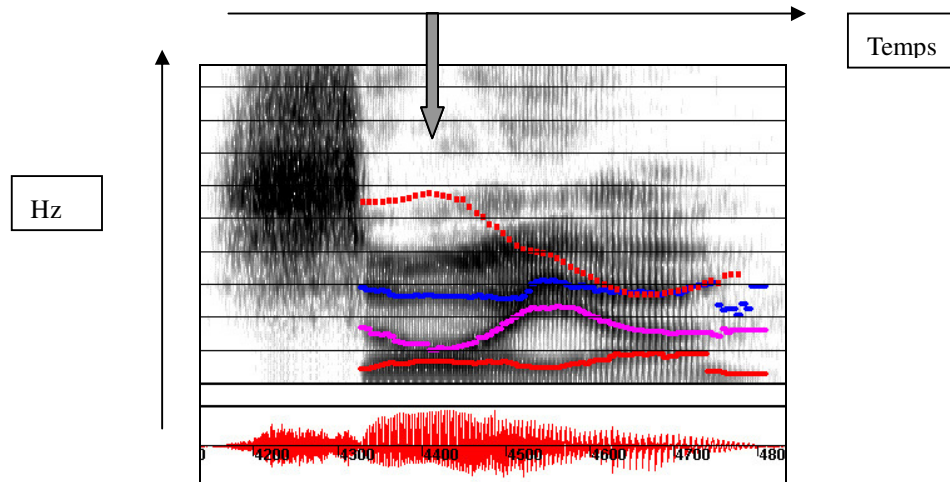


Figure n°74 : représentation spectrographique à bandes larges de la triphthongue [ɔɪə] du non mot « soya » prononcé par la locutrice anglophone avec superposition de l'évolution des trois premiers formants (lignes continues) et du contour de Fo (ligne en pointillés).

Du point de vue de la perception, dans une triphthongue, les trois cibles vocaliques, c'est-à-dire le point de départ, la cible à mi-chemin et le point d'arrivée ne sont pas perçus comme trois voyelles séparées, mais comme un glissement vers le second élément dans un premier temps et comme un glissement vers le troisième élément dans un second temps.

TE	PS1	TC1	PS2	TC2	PS3	Q

Tableau n°11 : les différents événements composant la structure d'une triphthongue (TE pour tête, PS1 pour première partie stable, TC1 pour première transition centrale, PS2 pour deuxième partie stable, TC2 pour deuxième transition centrale, PS3 pour troisième partie stable et Q pour queue).

3.2.2.2.2. Description des triphthongues de l'anglais

Selon Ginésy (1995), les vraies triphthongues sont relativement rares en anglais. Elles apparaissent dans des mots indécomposables tels que « *fire* » [faɪə], « *choir* » [kwaiə], « *shower* » [ʃaʊə], etc. En anglais R.P., on en dénombre 5 :

- 3 triphthongues ayant comme cible centrale [ɪ] : [aɪə], [eɪə] et [ɔɪə] (**extrait sonore n°33**) ;
- 2 triphthongues ayant comme cible centrale [ʊ] : [aʊə] et [əʊə] ; nous les décrirons pas car, comme nous l'avons déjà précisé, elles apparaissent très rarement en anglais R.P.

La tendance dominante est à la simplification et les triphthongues se réduisent souvent à une diphtongue, la cible centrale n'étant pas réalisée : « *les sons [ɪ] et [ʊ], qui sont les éléments centraux de ces triphthongues sont rarement atteints mais sont plutôt effleurés* » (Ginésy, 1995). Il arrive même qu'elles soient réduites à une longue monophthongue.

Du point de vue de leur structuration temporelle, les deux triphthongues [eɪə] et [ɔɪə] sont semblables, bien que la trajectoire de [ɔɪə] soit beaucoup plus marquée (voir ci-dessous, sur la charte de rapports de formants, figure n°75). Cette différence s'exprime dans la durée de PS2, plus réduite pour [eɪə] (1,39 %), que pour [ɔɪə] (8,93 %) (voir tableau n°12). Dans [eɪə], la première cible [e] est proche de la seconde cible [ɪ], ce qui réduit le trajet articulatoire pour aller de l'une à l'autre. Une fois la seconde cible atteinte, la trajectoire s'infléchit immédiatement en direction de la troisième cible. En revanche, dans la triphthongue [ɔɪə], la distance articulatoire entre les deux premières cibles est plus élevée. La trajectoire s'attarde donc un peu plus longtemps sur la cible centrale, avant de s'infléchir en direction de la troisième cible. On relève pour ces deux triphthongues de grandes phases de transition entre les cibles (DTC1 : 21,67 % et DTC2 : 18,88 % pour [eɪə] ; DTC1 : 18,95 % et DTC2 : 15,25 % pour [ɔɪə]), ce qui rend ces sons très instables au niveau articulatoire, mais aussi en ce qui concerne leur timbre.

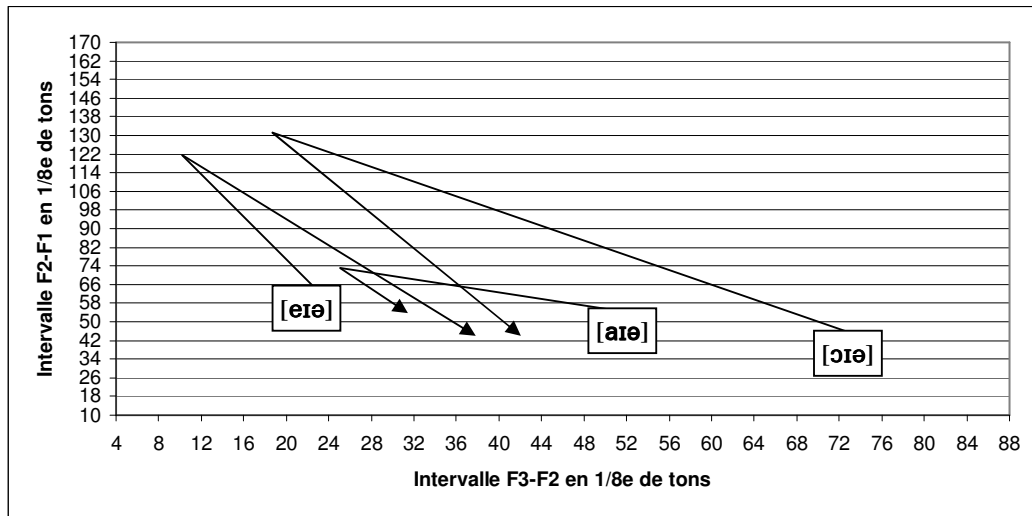


Figure n°75 : trajectoires des triptongues prononcées par la locutrice anglophone.

Comparée à ces deux diphtongues, la trajectoire de [aɪə] est plus simple. En fait, la cible centrale est « effleurée », pour reprendre les termes de Ginésy (1995 : 44), la locutrice n'étant pas allée assez loin dans l'espace articulatoire pour atteindre la cible (phénomène d'« undershooting »). En effet, [ɪ] est prononcé avec une aperture trop grande, sa valeur de F1 étant identique à celle de [a] (714 Hz). En conséquence, la durée de la transition (TR1) entre la première et la seconde cible est réduite (1,39 %). En revanche, la proportion de stabilité pour cette triptongue est plus grande que dans les deux autres triptongues (52,89 % de stabilité, contre 35,18 % pour [eɪə] et 35,29 % pour [ɔɪə], la stabilité étant calculée en additionnant le pourcentage les trois PS), la locutrice s'attardant davantage sur chacune des cibles.

En ce qui concerne le timbre des « cibles », il coïncide avec celui des voyelles désignées par les symboles phonétiques utilisés dans la transcription phonétique des trois triptongues étudiées, à l'exception de [ɪ] dans la triptongue [aɪə].

	DT	DPS1	DTC1	DPS2	DTC2	DPS3	DQ	Durée par rapport à la syllabe
[eɪə]	12,12 %	8,15 %	21,67 %	1,39 %	18,88 %	25,64 %	12,12 %	75,78 % (429 ms) CV
[aɪə]	3,96 %	13,51 %	1,39 %	10,95 %	17,48 %	28,43 %	24,24 %	78,09 % (429 ms) CV
[ɔɪə]	16,55 %	5,01 %	18,95 %	8,93 %	15,25 %	21,35 %	13,94 %	74,00 % (459 ms) CV

Tableau n°12 : durée relative des événements composant les 3 triphthongues prononcées

par la locutrice anglophone (DT pour durée totale, DPS1 pour durée de la première partie stable, DTC1 pour durée de la première transition centrale, DPS2 pour durée de la deuxième partie stable, DTC2 pour durée de la deuxième transition centrale, DPS3 pour durée de la troisième partie stable et DQ pour durée de la queue).

3.2.2.3. Difficultés rencontrées par les apprenants francophones dans la restitution des diphtongues et des triphthongues de l'anglais

Les diphtongues et triphthongues spécifiques de l'anglais sont très difficiles à reproduire par les francophones. La vraie diphtongaison s'explique physiologiquement par une inaptitude à tenir les organes en place pendant l'articulation d'une voyelle longue. Or, nous avons vu que le français se caractérisait par une grande tension articulaire. Un francophone a donc plutôt tendance à produire une succession de deux ou trois voyelles (avec des zones de transition centrale très réduites), plutôt qu'une diphtongue ou une triphthongue, car son articulation tendue l'empêche de réaliser un glissement progressif : au lieu de passer graduellement d'un son à l'autre, il a tendance à sauter de l'un à l'autre, comme il le fait dans sa LM lorsqu'il rencontre un hiatus¹² (Faure, 1948 : 30). Il peut même aller jusqu'à simplifier et ne produire qu'une seule voyelle dans le cas de la diphtongue et deux voyelles dans le cas de la triphthongue.

¹² Comme dans le mot « pays » par exemple. Faure (1948 : 30) parle d'ailleurs de « fausses diphtongues ».

La rythmique française appliquée à l'anglais renforce non seulement la tension articulatoire, mais elle empêche le locuteur francophone de produire correctement le fameux « stress pulse » dans les diphtongues, à cause de l'introduction d'un accent primaire sur la dernière syllabe (allongement final). Comme le souligne Faure (1948 : 30), « *le défaut de beaucoup de français consiste à donner trop de relief à la fin des diphtongues* ».

Dans les triphthongues, l'allongement final a comme conséquence principale l'insertion d'une coupe syllabique après la première cible. Ainsi, un énoncé comme « soya » a tendance à être prononcé de la manière suivante : [sɔ-ɪə].

Enfin, le crible phonologique contribue également à altérer la production des diphtongues et des triphthongues. Le timbre des cibles a tendance à être assimilé à celui des voyelles françaises de timbre correspondant. Or, si les cibles sont mal localisées, c'est toute la trajectoire de la diphtongue ou de la triphthongue qui est bouleversée. Par ailleurs, dans les productions des francophones, la trajectoire a tendance à glisser trop loin, jusqu'à la semi-voyelle [j] dans les diphtongues et les triphthongues contenant la cible [ɪ]. Dans les triphthongues, cette tendance est renforcée par l'insertion d'une coupe syllabique après la première cible. Un récapitulatif de ces erreurs est exposé dans le Volume II (annexe n°3.4.).

Conclusion

A partir de la description comparée de l'anglais et du français présentée dans ce chapitre, un catalogue d'interférences a été réalisé dans lequel figure la liste des principales erreurs prosodiques (Volume II, annexe n°3.5.1.) et vocaliques (Volume II, annexe n°3.5.2.) produites par les francophones en anglais. Cette description nous a également permis de créer deux batteries de tests de perception qui nous serviront d'outil d'évaluation des enfants de notre corpus (chapitre VII) et elle constitue un document de référence qui nous permettra d'analyser leurs productions en anglais (chapitre VIII). Nous avons conscience que notre description du système vocalique de l'anglais en fonction de celui du français est fondée uniquement sur l'analyse de deux locutrices. De plus, les productions de la locutrice anglophone sont souvent exagérées, car elle est professeur d'anglais (l'exagération ayant une fonction pédagogique dans la « coopération exolingue » entre l'enseignant de langue et l'apprenant ; se reporter au chapitre II, 2.4.2.1.). Mais nous cherchions une production « modèle » sur laquelle fonder notre

description et nous avons pensé que celle d'une locutrice native et professeur d'anglais constituait le meilleur modèle qui soit, d'autant qu'il correspond au type de langage dans lequel sont baignés les apprenants (R.P., B.B.C. English). Par ailleurs, les enfants de notre corpus sont débutants en langue, d'où la nécessité de limiter la complexité de la description, notamment en ce qui concerne la prosodie (en début d'apprentissage, les productions des enfants se limitent le plus souvent à des mots et à quelques phrases courtes).

Bibliographie chapitre V

- Adamczewski, Henri et Keen, Denis (1973). *Phonétique et Phonologie de l'Anglais Contemporain*. Paris : Armand Colin, 253 p.
- Astesano, Corinne (2001). Rythme et Acentuation en Français : Invariance et Variabilité. Paris : L'Harmattan, 317 p.
- Baken, Ronald J. and Daniloff, Raymond G. (1991). *Readings in Clinical Spectrography of Speech*. San Diego: Singular, 566 p.
- Beckman, Mary (1992). Evidence for speech rhythms across languages. In Tohkura, Vatikiotis-Bateson and Sagisaka, 457-463.
- Best, Catherine, T., McRoberts Gerald W. and Sithole, Nomathemb N. (1988). The phonological basis of perceptual loss of nonnative contrasts: maintenance of discrimination among Zulu clicks by English-speaking adults and infants. In *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.
- Best, Catherine, T. (1993). Emergence of language-specific constraints in perception of non-native speech a window on early phonological development. In Boysson-Bardies, Schonen, Jusczyk, MacNeilage and Morton, 289-304.
- Bot, K. de (1986). The transfer of intonation and the missing data base. Kellerman, E. and Smith, M. S., Crosslinguistic Influence. In *Second Language Acquisition*, 110-119.
- Carton, Fernand (1974). *Introduction à la Phonétique Française*. Paris : Bordas (1ère édition), 253 p.
- Catford, John C. (1988). *A Practical Introduction to Phonetics*. Oxford University Press.
- Chomsky, Noam and Halle, Morris (1968). *The Sound Pattern of English*. New-York: Harper and Row, 470 p.
- Dauer, Rebecca M. (1983). Stress-timing and syllable-timing reanalysed. In *Journal of Phonetics*, 11, 51-62.
- Delattre, Pierre (1961). La leçon d'intonation de Simone de Beauvoir. Etude d'intonation déclarative comparée. Dans *French Review*, 35 (1), 59-67.
- Delattre, Pierre (1965). *Comparing the Phonetic Features of English, French, German and Spanish*. Heidelberg: Julius Gross Verlag, 118 p.
- Di Cristo, Albert (1999). Le cadre accentuel du français contemporain : essai de modélisation. Dans *Langues*, 3/2, 184-205.
- Dodane, Christelle (2001). Les nouvelles technologies multimédias dans l'apprentissage de la prosodie d'une langue étrangère : une révolution dans la pratique ? *Actes du VIIe Congrès International de Communication Sociale*, Santiago de Cuba, Vol. I, 176-182.
- Faure, Georges (1962). *Recherches sur les Caractères et le Rôle des Eléments Musicaux dans la Prononciation*

- Anglaise* (Essai de Description Phonologique). Thèse de Doctorat d'Etat. Paris : Didier, 380 p.
- Faure, Georges (1948). *Manuel Pratique d'Anglais Parlé*. Paris : Hachette, 211 p.
- Fonagy, Yvan (1980). L'accent en français : accent probabilitaire. Dans Fonagy et Léon, 123-233.
- Fonagy, Ivan and Léon, Pierre R. (1980). *L'Accent en Français Contemporain*. Studia Phonetica, 15.
- Ginesy, Michel (1995). *Mémento de Phonétique Anglaise*. Paris : Nathan Université, 288 p.
- Grover, C., Jamieson, D. G. and Dobrovolsky, M. B. (1987). *Intonations in English, French and German: Perception and Speech*, 30/3, 277-295.
- Guierre, Lionel (1984). *Drills in English Stress-Patterns*. Paris : A. Colin-Longman, 127 p.
- Guimbretière, Elisabeth (1994). *Phonétique et Enseignement de l'Oral*. Paris : Didier / Hatier, 96 p.
- Hirst, Daniel (1998). Intonation in British English. In Hirst and Di Cristo, 56-78.
- Hirst, Daniel and Di Cristo, Albert (1984). French intonation: a parametric approach. Dans *Die Neueren Sprachen*, 83/5, 554-569.
- Hirst, Daniel and Di Cristo, Albert (1998). *Intonation Systems: a Survey of Twenty Languages*. Cambridge: University Press, 487 p.
- Institut National de la Statistique et des Etudes économiques (2001). *Note d'Information 01-18*. Edu050, avril.
- Johnson K., Ladefoged P. and Lindau M. (1993). Individual differences in vowel production. In *Journal of the Acoustical Society of America*, 94, 701-714.
- Konopczynski, Gabrielle (1990). *Le Langage Emergent : Caractéristiques Rythmiques*. Hamburg : Buske Verlag, 362 p.
- Konopczynski, Gabrielle (1991). Acquisition de la proéminence dans le langage émergent. Dans *Actes du XIIème Congrès International des Sciences Phonétiques*. Aix-en-Provence, Vol I, 333-337.
- Lacheret-Dujour Anne et Beaugendre Frédéric (1999). *La Prosodie du Français*. Paris : C.N.R.S. Langage, 353 p.
- Ladd, Robert (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladefoged, Peter (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovitch, Inc., 296 p.
- Ladefoged, Peter (1997). Linguistic Phonetic Descriptions. In Hardcastle and Laver, 589-618.
- Lado, Robert (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, Michigan, 141 p.
- Lanchec, Jean-Yvon (1976). *Psycholinguistique et Pédagogie des Langues*. Paris : PUF, 164 p.
- Landercy, Albert et Renard, Raymond (1977). *Eléments de Phonétique*. Bruxelles : Didier, 214 p.
- Lehiste Ilse and Peterson Gordon E. (1960). Transitions, glides, and diphthongs. In Baken and Daniloff (1991), 286-295.
- Léon, Pierre (1992). *Phonétisme et Prononciation du Français*. Paris : Nathan, 192 p.

- Léon, Pierre et Martin, Philippe (1971). Linguistique appliquée et enseignement de l'intonation. Dans *Etudes de Linguistique Appliquée*, 3, 36-45.
- Lhote, Elizabeth (1987). *A la Découverte des Paysages Sonores des Langues*. Besançon : Annales Littéraires de l'Université, 215 p.
- Lhote Elisabeth (1995). *Enseigner l'Oral en Interaction*. Paris : Hachette, 158 p.
- Llorca, Régine (1987). Le paysage français pour un Français : les bases rythmiques du français. In Lhote, 47-58.
- Lonchamp, François (1983). On Vowel Normalisation. *Leeds Experimental Phonetics Symposium*, Leeds.
- Meschonnic, Henri (1982). *Critique du Rythme*. Paris : Verdier, 713 p.
- Miller J. D. (1989). Auditory-perceptual interpretation of the vowel. In *Journal of Acoustical Society of America*, 85, 2114-2134.
- Montreuil, Jean-Pierre (2001). *La Phonologie de l'Anglais*. Rennes : Presses Universitaires de Rennes, 168 p.
- Ostiguy, Luc, Sarasin, Robert et Irons, Glenwood (1996). *Introduction à la Phonétique Comparée : les Sons. Le Français et l'Anglais Nord-Américain*. Laval : Presses de l'Université de Laval, 200 p.
- Pasdeloup Valérie (1991). *Modèles de Règles Rythmiques du Français Appliqué à la Synthèse de la Parole*. Thèse de doctorat, Université de Provence.
- Peterson Gordon E. (1961). Parameters of vowel quality. In *Journal of Speech and Hearing Research*, 4, 10-29.
- Peterson, Gordon E. and Barney, Harold L. (1952). Control Methods used in a study of vowels. In *Journal of the Acoustical Society of America*, 24, 175-184.
- Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press, 203 p.
- Potter R. K. and Steinberg J. C. (1950). Towards the specification of speech. In *Journal of Acoustical Society of America*, 22, 807-820.
- Pritchard, R. M. O. (1985). *The teaching of French intonation to native speakers of English*. In *International Review of Applied Linguistics*, 23/2, 117-147.
- Rossi, Mario (1985). L'intonation et l'organisation de l'énoncé. Dans *Phonetica*, 42, 135-153.
- Rossi, Mario (1999). *L'Intonation, le Système du Français : Description et Modélisation*. Paris : Ophrys, 237 p.
- Schwartz J. L. (1987). *Représentation Auditive des Spectres Vocaliques*. Thèse de Doctorat d'Etat, Université de Grenoble.
- Scott, Thomas (1987). Le français face au paysage anglais. Dans Lhote, 131-142.
- Tohkura, Vatikiotis-Bateson and Sagisaka (1992). *Speech Perception, Production and Linguistic Structure*. Tokyo: IOS Press.
- Troubetzkoy, Nicolas S. (1949). *Principes de Phonologie*. Paris : Klincksieck, 396 p. Traduit par Cantineau, J. (1986).
- Wen, and Wioland, François (1982). Is French really syllable-timed ? In *Journal of Phonetics*, 19.

Yaeger, Malcah (1979). *Context Determined Variation in Montreal French Vowels*. University of Pennsylvania, Ph.D, 353 p.

Chapitre VI

Constitution des groupes

« L'expérimentation exige un système fermé qui puisse être artificiellement isolé. Si nous voulons savoir dans quelles conditions l'eau va bouillir, nous isolons un groupe : source de chaleur, récipient, liquide (...) nous arrivons à le soustraire à la plupart des influences extérieures » (Maurois, Un art de vivre, I, 7).

Table du sixième chapitre

1) Limites de l'étude

- 1.1 Problèmes rencontrés
- 1.2 Tests musicaux
- 1.3 Réalisation d'un questionnaire

2) Définition des groupes

- 2.1 En fonction de la classe, musicale ou non
- 2.2 En fonction de la note globale obtenue aux tests musicaux
 - 2.2.1. Description des six catégories
 - 2.2.2. Résultats généraux
 - 2.2.3. Résultats détaillés
 - 2.2.3.1. Tests de discrimination
 - 2.2.3.2. Tests d'identification
 - 2.2.4. Tests les plus discriminants et nouvelle définition des groupes
 - 2.2.5. Importance des domaines harmonique et mélodique
 - 2.2.5.1. Harmonie
 - 2.2.5.2. Mélodie
- 2.3 Autres classifications
 - 2.3.1. Définition des groupes en fonction de la catégorie socioprofessionnelle
 - 2.3.2. Définition des groupes en fonction du milieu familial
 - 2.3.3. Définition des groupes en fonction de la pratique musicale et de la pratique instrumentale
- 2.4. Conclusion

Bibliographie du chapitre VI

La spécificité de notre échantillon nous a amenée à revoir la méthode que nous voulions adopter au départ et notamment à redéfinir la constitution des groupes. L'objectif de ce chapitre est de construire une échelle permettant de discriminer de manière objective les oreilles musiciennes des oreilles non musiciennes. Les enfants seront ensuite classés en fonction de cette échelle. Nous cherchons donc une sorte de « toise » qui permette objectivement de déterminer la qualité de l'oreille musicale d'un enfant.

1) Limites de l'étude

1.1 Problèmes rencontrés

Une méthodologie « idéale » ... mise à mal au contact de la réalité

Pour étudier le transfert de compétences de l'écoute musicale à l'écoute linguistique et les influences sur la production d'une langue étrangère, nous avons besoin de travailler sur des enfants qui soient à la fois en apprentissage précoce de l'anglais et de la musique. La situation expérimentale idéale aurait été d'avoir une classe expérimentale dite « musicale » (classe bénéficiant d'un enseignement musical aménagé de type Conservatoire) et une classe contrôle « non musicale ». Si l'apprentissage de l'anglais était délivré par le même enseignant dans ces deux classes, nous n'aurions qu'une seule variable indépendante, la variable « musique » à deux modalités représentée par les deux sous-groupes suivants : le groupe « musique » (classe expérimentale) et le groupe « non musique » (classe contrôle).

Difficultés de travailler avec des classes musicales

Il existe des classes musicales dans les villes possédant un Conservatoire National de Région (C.N.R.) ou une Ecole Nationale de Musique (E.N.M.). Dans la région de Franche-Comté, on en trouve à Besançon, Montbéliard, Belfort et Dole. Elles sont donc relativement peu nombreuses. Pour cette raison, elles sont fortement sollicitées pour participer à des expérimentations. Par ailleurs et surtout dans des villes universitaires comme Besançon, l'inflation des demandes d'observation de classes (stagiaires de l'I.U.F.M., étudiants de la l'U.F.R. Lettres et de l'Ecole d'Orthophonie) provoque la réticence des inspections à accepter des protocoles expérimentaux lourds. Quand elles ne refusent pas, elles imposent des contraintes qui rendent l'expérimentation très difficile à conduire.

Difficultés de trouver des classes où l'apprentissage précoce de l'anglais commence avant le CM1

L'échantillon était d'autant plus difficile à constituer qu'il fallait que tous les enfants suivent également un apprentissage précoce de l'anglais. Nous avons vu dans le chapitre IV que le « plan langues vivantes à l'école primaire » prévoit la généralisation de l'apprentissage d'une langue étrangère dès la grande section de maternelle d'ici à 2005. Mais, au moment de notre expérimentation (année 2000-2001), cette réforme n'était pas encore entrée en vigueur et à la rentrée scolaire de 2000, la priorité était donnée à la généralisation de l'enseignement des langues vivantes étrangères dans les classes de CM1. On peut donc dire que jusqu'en juin 2001, l'enseignement des langues ne concernait pas tous les enfants avant le CM1 et pendant l'année scolaire 1999-2000, le dispositif ILV ne concernait que 1,9 % des effectifs de CP et 6,2 % des effectifs de CE1, enseignement public et privé réunis¹. Il fallait donc trouver des classes en sachant que leur nombre est très réduit ou des écoles ayant pris elles-mêmes l'initiative d'instituer un apprentissage précoce « hors cadre officiel ».

A la recherche du « Graal »...

Dans ces conditions, c'était une véritable gageure de trouver une classe musicale, bénéficiant d'un apprentissage précoce commençant avant le CM1 et de trouver une classe contrôle ayant le même professeur d'anglais que la classe musicale.

Après de longues recherches, nous avons eu la chance de trouver trois classes de CE1, dont deux proviennent d'établissements privés sous contrat avec l'état : la classe de CE1 de l'école publique de Choisey (Jura), la classe de CE1 de l'école privée Ste Ursule de Dole (Jura) et la classe de CE1 de l'école privée de la Maîtrise de Dijon (Bourgogne).

Un apprentissage de l'anglais assuré par des intervenants qualifiés...

Les trois écoles concernées ont un point commun révélateur : elles ont toutes les trois mis en place un enseignement précoce des langues étrangères de leur propre initiative, quelquefois sous l'impulsion des parents d'élèves dans le cas de l'école de Choisey. Cet enseignement ne se situe dans aucun dispositif institutionnel existant. Pour cette raison, il est assuré par des intervenants extérieurs.

¹ Sources : note d'information 01-18 (avril 2001), Ministère de l'Education Nationale

Dans les trois classes composant notre échantillon, les enseignants sont des intervenants extérieurs, recrutés par les écoles en fonction de leurs compétences en anglais : une intervenante anglophone bilingue à Dole et des professeurs d'anglais de collège à Dijon et à Choisey.

... et réellement précoce

A Dole et à Choisey, le contact précoce avec l'anglais commence dès le CP, à raison d'une séance de 45 mn par semaine. A la Maîtrise de Dijon, les instituteurs commencent à sensibiliser les enfants à l'anglais dès la grande section de maternelle en utilisant des cassettes vidéo. A partir du CP, c'est un professeur de collège qui assure, comme dans les deux autres écoles, une séance par semaine. De plus, l'institutrice de la classe de CE1, formée à l'enseignement de l'anglais (certificat européen de langues obtenu à l'INFOP) assure une séance supplémentaire de 30 mn par semaine.

Ne se situant dans aucun dispositif existant, les intervenants n'ont pas de retour « qualitatif » et « officiel » sur l'enseignement qu'ils délivrent. Pour cette raison, les directeurs d'établissement, les professeurs d'anglais et les instituteurs étaient très favorables à un travail de recherche étudiant une particularité pour laquelle ils ont une forte motivation et dont ils tirent beaucoup de fierté.

Pourquoi le CE1 et pas le CP ?

Il aurait été intéressant de choisir les classes de CP pour la constitution de notre échantillon, là où commence l'enseignement de l'anglais pour au moins deux des classes, Dole et Choisey. Mais les effectifs de ces classes étaient très importants et le programme d'apprentissage de la lecture trop conséquent pour les instituteurs. Le fonctionnement de ces classes aurait été perturbé par une expérimentation lourde. Les effectifs étaient plus réduits en classe de CE1, ce qui laissait plus de disponibilité aux instituteurs et aux élèves.

Apprentissage de la musique à l'école

Sur les trois classes composant l'échantillon (voir tableau récapitulatif n°13), il n'y a aucun enseignement musical dans celle de Dole ; nous l'avons donc considérée au départ comme non musicale (NM). A Choisey, un professeur de musique intervient une fois par semaine depuis la grande section de maternelle pour une séance de 45 mn.

Mais, ce taux horaire est trop faible pour faire de cette classe de CE1 une classe musicale. Nous l'avons donc considérée également comme non musicale (NM). En revanche, la classe de CE1 de la Maîtrise de Dijon est une classe musicale (M). L'horaire des enfants est aménagé de manière à ce qu'ils suivent un enseignement traditionnel en parallèle avec une formation quotidienne au chant. Les enfants appartiennent au chœur d'enfants de la Cathédrale Ste Bénigne. Depuis la rentrée 2000, les enfants sont initiés aux techniques vocales et à la pratique d'un instrument à clavier. Nous avons donc considéré cette classe comme la classe expérimentale de notre échantillon.

	Apprentissage de l'anglais	Apprentissage de la musique	Classification
Classe de Dole	1 séance par semaine depuis le C.P. assurée par une anglophone bilingue.	Aucun.	Classe non-musicale (NM) « contrôle »
Classe de Choisey	1 séance de 45 mn par semaine depuis le C.P. assurée par un professeur de collège.	1 séance de 45 mn par semaine assurée par un professeur de musique.	Classe non-musicale (NM)
Classe de Dijon	1 séance par semaine depuis le C.P. assurée par un professeur de collège. 1 séance de 30 mn par semaine assurée par l'instituteur de la classe. L'initiation commence dès la grande section de maternelle.	Horaires aménagés permettant une formation quotidienne au chant. Les enfants font partie du chœur de la cathédrale. Depuis la rentrée 2000, initiation aux techniques vocales et à un instrument à clavier.	Classe musicale (M) « expérimentale »

Tableau n°13 : tableau récapitulatif des enseignements délivrés en anglais et en musique dans les trois classes de CE1 composant notre échantillon (NM pour non musicale, M pour musicale, C.P. pour Cours Préparatoire).

Un échantillon éloigné de ce que nous voulions au départ

Notre échantillon est donc très éloigné de celui que nous voulions constituer au début. En effet, les trois classes proviennent de trois écoles de trois villes différentes. Dans ces conditions, il est bien sûr impossible d'avoir le même professeur d'anglais. Par ailleurs, la population d'enfants participant à notre expérimentation au sein de chaque classe est réduite, car certains parents ont refusé d'y participer. Afin d'avoir une population suffisamment importante, nous avons choisi trois classes au lieu de deux, multipliant ainsi les variables potentielles (enseignement d'anglais, école différentes...). La population se décompose de la manière suivante : 7 enfants dans la classe de Dole, 11 enfants dans la classe de Choisey et 9 enfants dans la classe de Dijon. Notre effectif total est de 27 enfants.

Comment isoler les effets de la variable musique sur les aptitudes en anglais ?

Ces difficultés rencontrées lors de la constitution de l'échantillon nous ont amenée à remettre en cause la méthodologie que nous voulions adopter au départ. Dans un premier temps, parce qu'il est presque impossible de trouver dans la réalité la configuration que nous voulions. Dans un second temps, est-ce qu'un tel verrouillage permet vraiment d'isoler l'effet de la variable musique sur les aptitudes des enfants en anglais ? Est-ce qu'il n'est pas arbitraire de considérer a priori qu'un enfant est musicien et que les autres ne le sont pas ? C'est oublier que certains enfants peuvent suivre une formation musicale en dehors du cadre de l'école (Conservatoire, école de musique, cours particuliers), qu'ils peuvent être non musiciens au niveau d'une pratique du chant ou d'un instrument, mais provenir d'un environnement familial très musical. En outre, si nous avons travaillé avec une classe musicale de conservatoire, nous risquons d'avoir un biais socio-économique important, les élèves fréquentant les conservatoires étant plutôt issus de milieux favorisés. Enfin, il existe des grandes différences d'aptitudes musicales entre les enfants, même au sein de la même classe musicale. Bien sûr, on s'attend à trouver plus d'oreilles expertes dans une classe musicale, mais notre échantillon n'est pas suffisamment important pour gommer les différences au niveau statistique. Un tel verrouillage, c'est finalement l'arbre qui cache la forêt. Il introduit un biais dangereux, du fait même qu'on pense avoir une situation expérimentale maîtrisée. La tâche n'est-elle tout simplement pas impossible ?

Nécessité d'adopter une autre méthodologie : évaluation objective des capacités de l'oreille par les tests auditifs

Finalement, toutes ces difficultés nous ont poussée à envisager le problème autrement. En fait, j'essaie de savoir si une « oreille entraînée » ou « experte » dans le domaine de la musique l'est également dans le domaine linguistique, et spécifiquement d'une langue étrangère, l'anglais, dont les caractéristiques sont très musicales (voir chapitre V). Or il existe une technique pour évaluer les capacités d'une oreille de manière « objective » : les tests de perception.

Pourquoi ne pas avoir réalisé des audiogrammes ?

Avant toute chose, il faudrait commencer par s'assurer que tous les enfants de notre échantillon possèdent une audition normale, d'un point de vue médical. Des déficits auditifs pourraient en effet engendrer des problèmes de perception et fausser nos résultats.

A cet âge, l'audition des enfants n'est jamais testée en milieu scolaire, pourtant des déficits auditifs peuvent se manifester, dus à des maladies du bébé (otites répétées, etc.) ou de la mère enceinte et à des environnements bruyants. A quel point une audition défectueuse peut-elle mettre l'enfant en difficulté scolaire ? Cette question, fondamentale, est souvent négligée, mais elle tombe en dehors de notre expérimentation. Néanmoins, si nous sommes amenée à détecter chez certains enfants des troubles de l'audition, nous les écarterons de notre échantillon. Pour la détection des troubles de l'audition, c'est l'audiométrie tonale liminaire qui est l'examen le plus courant. Grâce à elle, on peut détecter une audition anormale par rapport à la moyenne, et détecter des « trous ou pertes auditives » pour certaines fréquences. Il s'agit d'un bilan médical, qui doit normalement être administré à l'hôpital. Dans le milieu scolaire, il n'est pratiquement jamais réalisé. Nous ne pouvons pas demander aux parents d'envoyer leurs enfants à titre individuel à l'hôpital pour subir cet examen, à cause de l'expérience traumatisante pour les enfants (milieu hospitalier) et également pour des raisons financières. Nous aurions pu faire appel à la médecine scolaire, mais notre expérimentation est déjà lourde, ce qui a fait reculer certains établissements. Afin de limiter nos interventions au maximum, nous avons donc renoncé aux audiogrammes.

1.2 Les tests musicaux

Pourquoi réaliser des tests musicaux ?

Nous devons évaluer les capacités musicales des enfants de manière fine. La perception est une capacité cognitive générale et d'autres facteurs peuvent intervenir dans la formation de l'oreille. L'un des seuls moyens « objectifs » d'établir si les enfants de notre corpus ont une oreille musicale, c'est de leur faire passer à tous les mêmes tests musicaux. Les tests sont définis par Valette (1975 : 3) comme une « *technique d'évaluation objective des connaissances* ». Ils permettent en effet d'évaluer « objectivement » les capacités des sujets. Ce faisant, ils remplacent en partie les audiogrammes que nous voulions réaliser, car si un enfant souffre d'une audition défectueuse, il ne pourra pas réussir les tests musicaux. Toutefois, nous ne pourrions pas savoir si cet échec provient d'une audition défectueuse ou d'une « mauvaise oreille musicale ». Nous ne pourrions pas lever cette ambiguïté sans les audiogrammes. Nous avons choisi la batterie de tests musicaux élaborée par Zenatti (1980). Elle est particulièrement adaptée à notre population puisqu'elle a été conçue pour de jeunes enfants (pour une description de la batterie, se reporter au Volume II, annexe 6.1.).

Une batterie conçue pour les jeunes enfants

Il existe peu de batteries de tests conçues pour des enfants âgés de moins de 10 ans, mis à part celle de Bentley (1966) concernant des enfants âgés de 7 à 8/10 ans et celle de Gordon (1979) concernant des enfants âgés de 5 ans. La batterie de Zenatti concerne des enfants âgés de 4 ans à 7 ans 11 mois. Les autres batteries concernent des enfants plus âgés (Seashore, Kwalasser-Dytema, Wing) avec des comparaisons très fines (hauteur de 1/200^{ème} de ton et durée de 5/100^{ème} de seconde). Or, il est nécessaire de tenir compte du développement psychologique de l'enfant : par exemple, avant 8 ans, les enfants ne situent pas le grave et l'aigu dans l'espace, car ils ne maîtrisent pas le symbolisme spatial (Andrews et Madeira, 1977). Les tâches de perception utilisant dans les consignes des termes relationnels du type « *haut-bas* » et « *plus haut - plus bas* » sont donc inappropriées chez les jeunes enfants. Par exemple, la question « *Est-ce que ce ton est bas ou haut ?* » implique la compréhension des concepts relationnels et de la construction comparative. Elle met en jeu des relations cognitives complexes et des instructions verbales compliquées. Les expérimentations réalisées par Andrews et Madeira montrent que les enfants de moins de 8 ans échouent dans les tâches de discrimination, non pas parce qu'ils sont incapables de détecter les différences de hauteur, mais parce qu'ils sont incapables de traiter le langage relationnel. Dans les tests de Zenatti (1980), les consignes n'utilisent pas de langage relationnel.

Par ailleurs, l'utilisation d'un matériel non verbal permet de lever l'inhibition liée à l'utilisation du langage verbal. Le matériel a été réalisé pour éviter au maximum les réponses verbales et faciliter la compréhension de la tâche. Pendant les tâches, les signaux auditifs sont appariés avec des signaux visuels et tactiles (on demande à l'enfant, non pas une réponse verbale, mais une réponse motrice : il doit placer des images sur une autre image ou indiquer avec le doigt une image). Zenatti a essayé de réduire la part du hasard dans les réponses de l'enfant et les procédures faisant intervenir deux éventualités (pareil ou différent) ont été systématiquement écartées. Enfin, les activités étant présentées sous une forme ludique, elles permettent de limiter le stress inhérent à une situation de test.

Axe global-local

Dans nos hypothèses, nous avons établi une analogie entre l'analyse de la forme musicale et celle de la forme verbale à deux niveaux : traitement au niveau global et traitement au niveau local. En plus de porter sur des activités rythmiques, les tests de Zenatti portent à la fois sur les activités globales et analytiques, ce qui s'inscrit parfaitement dans le cadre de notre travail.

Evaluation « objective » des aptitudes musicales

Ces tests permettant de déceler très tôt les capacités musicales, ils nous aideront à découvrir, à l'intérieur de la classe contrôle, des enfants montrant des aptitudes musicales. Des entretiens viendront compléter ces résultats. Ils permettront peut-être de trouver dans le milieu familial, une explication au développement de ces capacités. En fonction des résultats, nous pourrons ainsi créer des sous-groupes à l'intérieur de la classe contrôle et de la classe expérimentale.

Validité de la batterie

Zenatti a réalisé trente-quatre tests qu'elle a éprouvés sur plus de 8 000 enfants. Elle en a sélectionné dix pour la réalisation de cette batterie. L'étalonnage des tests a été effectué à partir des résultats de 2300 enfants ayant un développement normal, âgés de 4 ans à 7 ans 11 mois, filles et garçons, appartenant à des milieux socio-économiques divers. La répartition des catégories socioprofessionnelles des parents se rapproche de celle répertoriée dans la population française lors du recensement de 1980. Ces épreuves sont indépendantes de la culture (« *culture free* ») et culturellement équilibrées (« *culture fair* »). Aucune classe sociale n'est donc favorisée. De plus, Zenatti a essayé de conserver le pouvoir discriminant des tests aux différents âges, tout en tentant de donner à la batterie une certaine homogénéité. Par ailleurs, la batterie de Zenatti concerne des enfants francophones, partageant donc la même culture musicale et la même langue.

1.3. Nécessité d'un questionnaire

Le développement musical de l'enfant dépend des deux facteurs principaux que sont le développement des fonctions cognitives (maturation du système nerveux central : augmentation des capacités d'enregistrement et de structuration des informations provenant de la sensorialité) et le milieu (environnement plus ou moins riche).

« Un double courant s'instaure entre l'enfant et le milieu, d'une part de réception et d'assimilation des données extérieures, d'autre part, d'adaptation à des situations nouvelles, cette adaptation s'accompagnant d'une transformation des structures mentales de l'enfant avec, pour conséquence, un meilleur niveau de développement » (Zenatti, 1981 : 13).

Un enfant classé comme non musicien peut provenir d'un environnement très musical, ce qui influencera la « qualité » de son oreille musicale. Nous savons que le groupe familial joue un rôle crucial sur le plan de la réussite scolaire, autant que sur le développement des aptitudes musicales de l'enfant. Certains facteurs ont plus d'importance que d'autres, tels que le statut socioprofessionnel des parents, la tradition musicale de la famille et le suivi de l'apprentissage et de la formation musicale (Manturzevska, 1974). Mais, comment évaluer la richesse d'un milieu ?

« Un milieu n'est définissable qu'à une certaine échelle d'observation, actuellement très grossière. A l'intérieur de chacun des milieux définissables, un très grand nombre de facteurs peuvent varier d'un individu à un autre » (Rostand, J, 1975 : 15).

Nous avons réalisé un questionnaire destiné aux parents afin de tenter de maîtriser partiellement toutes ces données et d'élaborer une sorte de « carte d'identité sonore » pour chacun des enfants de notre échantillon. Ce questionnaire est constitué de quatre grandes parties (pour une description détaillée de l'élaboration du questionnaire, se reporter au Volume II, annexe n°5.3.1., pour le questionnaire lui-même, annexe n°5.3.2. et pour un exemplaire rempli, annexe n°5.3.3.).

- les renseignements d'identité, permettant de décrire le milieu familial de l'enfant d'un point de vue sociologique (CSP, nombre d'enfants dans le foyer...) et de définir le statut socio-économique du foyer ;
- les « faits relatifs à l'enfant », permettant de dresser la « personnalité musicale » de l'enfant ;
- les « faits concernant les parents », permettant de déterminer si le milieu familial est « musicien » (environnement sonore à la maison) ;

- des questions sur l'apprentissage précoce de l'anglais destinées à déterminer l'influence éventuelle de l'entourage (personnes anglophones, ou maîtrisant la langue) sur les motivations, les capacités de l'enfant et sa perception du déroulement des séances d'anglais à l'école.

Nous aurions pu faire passer un questionnaire aux enfants comme nous l'avions fait dans une expérimentation précédente (Dodane, 1996), mais leurs réponses auraient été moins précises et surtout, nous voulions limiter au maximum le temps d'intervention dans les classes, notre expérimentation étant déjà assez prenante. Le questionnaire a été expédié aux parents avec un document de présentation de l'expérimentation (Volume II, annexe n°5.1.). Dans ce document, les objectifs ont été modifiés et, au lieu d'une étude sur l'influence de la formation musicale sur l'apprentissage d'une langue étrangère, c'est une étude sur les avantages de la « précocité » dans l'apprentissage précoce d'une langue étrangère et dans celui de la musique qui a été présentée. Cette affabulation permet de placer tous les enfants, qu'ils soient musiciens ou non musiciens, sur un même plan car ils bénéficient tous d'un apprentissage précoce de l'anglais et d'un enseignement musical, même minimal. Si les véritables objectifs avaient été présentés, les enfants non musiciens auraient été dévalorisés, ce qui aurait influencé leurs résultats. Une demande d'autorisation pour faire passer les tests et pour la publication des résultats était également jointe à l'envoi (Volume II, annexe n° 5.2.).

Entretien avec les institutrices

Enfin, pour compléter, il nous fallait obtenir des informations concernant le profil des enfants en classe. Certains enfants sont en effet agités, d'autres distraits, ce qui peut avoir des répercussions importantes sur les résultats aux tests. Nous avons donc demandé aux trois institutrices de décrire le profil de chaque enfant, son attitude en classe et ses résultats dans les différentes disciplines. Ces informations nous permettent de mieux cerner la personnalité de chacun des enfants. Nous verrons dans le chapitre VII à quel point ces informations peuvent être utiles pour expliquer certains résultats.

2) Définition des groupes

2.1. En fonction de la classe, musicale ou non

Au départ, comme nous l'avons déjà précisé, nous pensions répartir les enfants de notre corpus en deux catégories : la catégorie des « non musiciens » et celle des « musiciens ». Cette répartition devait être réalisée en fonction de l'appartenance des enfants à une classe expérimentale « musicale » ou à une classe contrôle « non musicale ». Nous avons vu que les spécificités de notre échantillon nous ont conduite à revoir notre méthodologie, en établissant un classement, non plus à partir de l'appartenance de l'enfant à telle ou telle classe, mais par rapport à la note générale obtenue aux tests musicaux.

Les résultats viennent justifier ce changement de méthodologie. Le tableau n°14 nous montre la note générale obtenue par chacun des sujets de notre échantillon, regroupés selon sa classe d'appartenance. La classe de Dijon est musicale et les deux autres sont non musicales. Dans un premier temps, nous constatons que la moyenne générale de la classe musicale est plus élevée que celle des deux autres : 80,11 à Dijon contre 74 à Dole et 71,45 à Choisey, même si cette différence n'est pas démesurée. Mais, si nous regardons les résultats plus en détail, nous remarquons qu'ils sont très disparates. Dans la classe de Dijon, il y a trois bons éléments qui obtiennent une note supérieure à 90 sur 102, mais un élément qui n'obtient que 51 sur 102, une des notes les plus basses de notre population. Par ailleurs, dans la classe de Dole, on trouve également trois bons éléments qui obtiennent une note supérieure à 90 sur 102. Quant à cette classe de Dole, sa moyenne est plus élevée que celle de Choisey, parce qu'il y a moins de dispersion entre les notes ($\sigma=10$). Les résultats sont trop disparates à Choisey ($\sigma=18,86$) et Dijon ($\sigma=16,11$) pour que le classement par classe soit pertinent. D'ailleurs, une analyse de variance (ANOVA, tests de Fisher) confirme que ce classement n'est pas significatif, puisqu'on n'observe aucun effet de groupe sur les résultats ($F=0,61$; $p=0,5532$).

Dans notre échantillon, l'appartenance d'un enfant à telle ou telle classe n'a aucune influence sur ses résultats, bien qu'on constate une moyenne supérieure dans la classe de Dijon, notre classe musicale. Le fait que ces trois classes soient issues de trois villes différentes et que deux d'entre elles soient privées ne semble pas influencer davantage les résultats.

N° des sujets	Classe de Dole (NM)	Classe de Choisey (NM)	Classe de Dijon (M)
1	66	42	101
2	64	72	72
3	72	49	102
4	72	51	80
5	88	62	77
6	88	52	51
7	68	98	70
8		97	72
9		75	96
10		91	
11		97	
Moyenne par classe	74	71,45	80,11
Ecart-type σ	10	18,86	16,11

Tableau n°14 : note générale obtenue aux tests de musique (sur 102) dans les deux types de classes : musicales (M) et non musicales (NM).

La définition des groupes composant notre expérimentation ne se fera donc pas en fonction des types de classes (musicales ou non), mais en fonction de la note obtenue aux tests de musique.

2.2. En fonction de la note globale obtenue aux tests musicaux

Le total général des notes obtenues aux tests de discrimination de modifications harmoniques, mélodiques et rythmiques et aux tests d'identification harmonique, mélodique, rythmique par apprentissage discriminatif correspond à une note globale de 102. En fonction de cette note, nous avons réparti les 27 enfants composant notre échantillon dans six sous-groupes différents (voir tableau n°15) : les « très bonnes oreilles musicales » (désormais « TBO »), les « bonnes oreilles musicales » (désormais « BO »), les « oreilles musicales moyennes » (désormais « OM »), les « oreilles musicales passables » (désormais « OP »), les « mauvaises oreilles musicales » (désormais « MO ») et enfin, les « très mauvaises oreilles musicales » (désormais « TMO »).

Note globale obtenue aux tests de musique (sur 102)	Nom des sous-groupes	Effectif par sous-groupe (n=27)	Moyenne générale pour chaque sous-groupe
≤ 90	TBO	7	97,42
80-89	BO	3	85,33
70-79	OM	8	72,75
60-69	OP	4	65
50-59	MO	3	51,33
≤ 49	TMO	2	45,5

Tableau n°15 : répartition de notre population en six sous-groupes, définis par rapport à la note globale obtenue aux tests de musique (TBO pour très bonne oreille, BO pour bonne oreille, OM pour oreille moyenne, OP pour oreille passable, MO pour mauvaise oreille, TMO pour très mauvaise oreille, n pour l'effectif).

Le classement général des enfants est expliqué à la lumière de la tradition musicale du foyer, de la pratique instrumentale, de la formation musicale des parents et du soutien offert dans la formation musicale de leur enfant, paramètres déterminants dans la formation des aptitudes musicales.

2.2.1. Description des six catégories

Très bonne oreille musicale (TBO)

Sept enfants ont obtenu une note supérieure à 90 sur 102 (dont 3 enfants de la Maîtrise, « classe musicale »). L'un d'eux a même obtenu la note maximale de 102, un autre 101 sur 102. Sur ces sept enfants, six sont musiciens. Tous les enfants musiciens sont issus d'un milieu musicien et ont en commun de pratiquer un instrument.

Les deux parents de Nicolas, l'enfant qui a obtenu un résultat parfait, sont musiciens et ont été formés en conservatoire. On note aussi beaucoup de musique à la maison. Ils ont placé leur enfant dans une classe « musicale » à cause de son goût pour le chant et de son « *oreille musicale* » (confirmée par les tests). Ils veulent inciter fortement leur enfant à poursuivre sa formation musicale en conservatoire. Actuellement, Nicolas suit des cours de piano au conservatoire.

La mère d'Armand (101 sur 102) fait partie d'une chorale et avoue avoir « *toujours regretté de ne pas avoir eu l'opportunité* » d'intégrer une classe musicale. Elle dit de son fils « *qu'il chantait à tue-tête quand il jouait* ». Elle souhaite qu'il approfondisse sa formation musicale, mais seulement s'il le demande (« *je ne veux pas en faire un virtuose au détriment des matières plus scolaires* »).

Ce sont les deux seuls enfants dont les parents mentionnent explicitement l'existence d'un « don musical » (également les deux meilleurs résultats dans les tests musicaux).

Les deux sœurs d'Hugo (96 sur 102) sont musiciennes (l'une appartient au chœur de la Maîtrise et l'autre est pianiste). Hugo est pianiste débutant. Ses parents souhaitent qu'il approfondisse son éducation musicale « *sauf opposition forte de sa part* » (« *pas de carrière dans la musique envisagée* »). Les trois autres enfants sont issus de la classe de Choisey.

Violette (98 sur 102) fait de la musique depuis l'âge de 5 ans (jardin d'éveil musical). Sa mère est pianiste (8 ans de piano) et son frère joue du basson. Violette pratique le piano en conservatoire (ENM de Dole). Ses parents souhaitent qu'elle continue « *le plus longtemps possible* ».

La mère de Lucie a appris l'orgue et le solfège et Lucie (97 sur 102) pratique le synthétiseur depuis 2 ans. Concernant l'éducation musicale de sa fille, elle « *ne la forcera pas* ». « *Cela ne doit pas devenir une obligation* ».

Enfin, le grand-père d'Agathe est chef de chœur et sa tante est pianiste. Agathe (97 sur 102) fait de la musique depuis l'âge de 4 ans (jardin d'éveil musical) et elle apprend le violoncelle (l'instrument préféré de sa mère) en conservatoire (ENM de Dole).

Le seul enfant qui ne soit pas musicien, Lucas (score le plus faible de ce groupe, 91 sur 102), est issu d'un milieu musicien. Son père est en effet professeur de musique (formation au conservatoire) et son frère, guitariste.

Les sept enfants proviennent tous d'un milieu musicien, fortement impliqué dans leur éducation musicale. Dans cette catégorie, on remarque les réactions les plus fortes de la part des parents : ils veulent en majorité que leurs enfants continuent. On verra plus loin que les autres laissent plutôt libre choix à leurs enfants.

Bonne oreille musicale (BO)

Trois enfants ont obtenu un résultat compris entre 80 et 89 sur 102 et deux de ces enfants sont musiciens.

L'oncle d'Emeline (80 sur 102), qu'elle voit souvent, est musicien (harmonica, guitare et flûte). Celle-ci fait partie de la classe musicale de la Maîtrise.

Le père de Matthias est guitariste (88 sur 102) et Matthias joue du synthétiseur (cours particuliers depuis septembre 2000). Enfin, Mélissa (88 sur 102) est non musicienne, mais sa grande sœur qui est pianiste, a fréquenté l'ENM de Dole pendant 10 ans. Ses parents étaient au conservatoire dans leur jeunesse. Comme précédemment, ils ont en commun de provenir tous d'un milieu musicien.

On constate moins d'implication de la part des parents dans l'éducation musicale de leurs enfants que dans le groupe TBO. Emeline, « *peut-être approfondir si tel est son désir* ». Matthias, a le « *choix d'arrêter* ». Mélissa « *Si elle le souhaite* ». Le libre choix est laissé à l'enfant, les parents n'exercent pas de contrainte.

Les dix enfants des groupes TBO et BO ont obtenus des résultats égal ou supérieur à 80 aux tests musicaux. Parmi ces dix enfants, huit sont musiciens (dont sept instrumentistes). Tous proviennent d'un milieu possédant une tradition musicale marquée (parents ou proches musiciens ou anciens musiciens).

Oreille « moyenne » (OM)

Huit enfants ont obtenu un résultat compris entre 70 et 79 sur 102. Six enfants sont musiciens, dont deux appartiennent à un milieu non musicien. L'un des deux non musiciens provient d'un milieu non musicien. Parmi les enfants musiciens, trois enfants sont instrumentistes, les trois autres font partie de la Maîtrise.

Dans cette catégorie, Matthieu joue d'un instrument (batterie) et fait partie de la Maîtrise. Seuls les parents de Lison veulent qu'elle continue le piano dans une école de musique (sa sœur et son frère sont également pianistes). Pour tous les autres enfants, les parents répondent presque invariablement à la question « *Souhaitez-vous qu'il approfondisse son éducation musicale ?* » par « *Oui, si il ou elle le désire* ». Le père de John (John est guitariste) aimerait que son fils suive une formation musicale, mais pour l'instant il n'en a pas les moyens financiers. Quant à Quentin, non musicien issu d'un milieu non musicien, ses parents ont répondu, que s'il le souhaite, ils l'inscriront au conservatoire.

Oreille « passable » (OP)

Quatre enfants ont obtenu un résultat compris entre 60 et 69 sur 102. Tous les quatre sont non musiciens issus de milieux non musiciens.

Les parents de Laeticia et de Chloé laissent le libre choix à leur enfant pour commencer une éducation musicale. Ceux de Laurène aimeraient qu'elle commence en école de musique (mais libre choix de continuer). Ceux de Laeticia ne l'envisagent pas (« *oui, comme matière enseignée à l'école* »).

« Mauvaise » oreille (MO)

Trois enfants ont obtenu un résultat compris entre 50 et 59 sur 102. Deux d'entre eux sont non musiciens et issus de milieux non musiciens.

Le troisième, Kevin, est musicien (classe Maîtrise) et issu d'un milieu musicien (sa mère a fait deux ans de conservatoire et pratique la flûte, l'alto et la clarinette). Il est étrange de trouver un musicien avec de tels résultats (51 sur 102). Un commentaire de sa mère a attiré notre attention « *On dit que la musique adoucit les mœurs. Kevin en avait besoin pour apprendre à se maîtriser* ». Nous avons constaté en faisant passer les tests de musique, mais également ceux d'anglais, que Kevin n'arrivait pas à se concentrer très longtemps. Il a besoin de bouger, de se lever et au bout d'un moment, il n'écoute plus. Ce comportement, confirmé en classe par son institutrice, peut expliquer les mauvais résultats de Kevin. C'est en effet le seul musicien issu en outre d'un milieu musicien à se trouver dans une des catégories « mauvaise » oreille.

« Très mauvaise » oreille (TMO)

Deux enfants ont obtenu un résultat compris entre 40 et 49. Timothé (40 sur 102) et Medhi (49 sur 102) sont non musiciens et proviennent de milieux non musiciens. Il est à noter que Medhi est un enfant qui rencontre beaucoup de difficultés scolaires. Il sait à peine lire, malgré deux CP et deux CE1.

2.2.2. Remarques sur les résultats généraux aux tests de musique

Ces tests se révèlent très efficaces pour séparer les musiciens des non musiciens. Au-delà de 80 sur 102 (catégories TBO et BO), on trouve huit enfants musiciens, et seulement deux enfants non musiciens, mais qui sont issus de milieux très musiciens. Le milieu semble donc jouer un rôle très important dans la formation de l'oreille, en influençant même l'oreille des non musiciens.

Chez les musiciens, trois enfants sont issus de milieux non musiciens. Le fait que l'enfant pratique un instrument semble constituer un facteur encore plus important, puisque sept enfants sur les huit musiciens sont instrumentistes. Cela vient compenser un milieu non musicien chez les musiciens (se reporter au tableau n°16 ou à la figure n°76).

Dans la catégorie intermédiaire OM, on trouve six musiciens sur huit, deux musiciens provenant d'un milieu non musicien. Les deux non musiciens proviennent d'un milieu non musicien. Dans les catégories inférieures, la proportion de non musiciens chute, huit enfants sur neuf sont non musiciens, tous issus de milieux non musiciens. On relève toutefois une exception : dans la catégorie MO, on trouve un enfant musicien issu d'un milieu musicien (catégorie MO). Mais nous avons vu que cet enfant constitue un cas particulier.

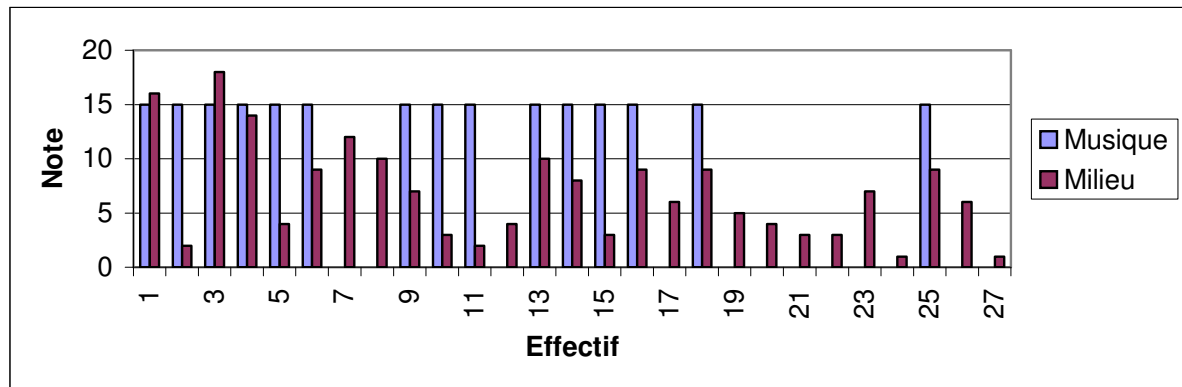


Figure n°76 : pratique de la musique et richesse du milieu familial (sujets classés du meilleur au plus faible).

Groupe	Sujet	Note générale en musique	Pratique musicale	Note richesse musicale du milieu
TBO	1	102	M	16
	2	101	M	2
	3	98	M	18
	4	97	M	14
	5	97	M	4
	6	96	M	9
	7	91	NM	12
BO	8	88	NM	10
	9	88	M	7
	10	80	M	3
OM	11	77	M	2
	12	75	NM	4
	13	72	M	10
	14	72	M	8
	15	72	M	3
	16	72	M	9
	17	72	NM	6
	18	70	M	9
OP	19	68	NM	5
	20	66	NM	4
	21	64	NM	3
	22	62	NM	3
MO	23	52	NM	7
	24	51	NM	1
	25	51	M	9
TMO	26	42	NM	6
	27	49	NM	1

Tableau n°16 : pratique de la musique (M) et richesse du milieu familial en fonction des résultats obtenus aux tests de musique (**TBO** pour très bonne oreille, **BO** pour bonne oreille, **OM** pour oreille moyenne, **OP** pour oreille passable, **MO** pour mauvaise oreille, **TMO** pour très mauvaise oreille, **M** pour enfant bénéficiant d'une formation musicale, **NM** pour enfant ne bénéficiant pas de formation musicale).

Les résultats confirment qu'il ne suffisait pas de classer tous les enfants de la classe musicale comme musiciens, car il n'y a pas d'uniformité dans les résultats obtenus aux tests de musique, les notes s'échelonnant de 51 à 102. On ne constate pas davantage d'uniformité chez les non musiciens, les notes s'échelonnant de 42 à 91. Par ailleurs, certains non musiciens ont obtenu de très bonnes notes, l'un se trouvant même dans la catégorie TBO. Comment aurions-nous pu expliquer ces résultats si nous n'avions pas eu d'information sur le milieu familial ? Ces résultats confirment donc l'importance du questionnaire, qui nous a permis bien sûr d'affiner le profil des enfants, mais surtout d'expliquer pourquoi certains enfants non musiciens obtenaient de si bons résultats aux tests de musique. Grâce à ces informations, nous remarquons l'importance du milieu dans la formation de l'oreille et de manière encore plus marquée, l'influence d'une pratique instrumentale en dehors de la classe pour les meilleurs.

2.2.3. Résultats détaillés des groupes aux six tests de musique de Zenatti (1980)

Pour composer sa batterie de dix tests, Zenatti (1980) a sélectionné les tests les plus fiables d'une batterie initiale composée de trente-quatre tests (1981), dont la validité a été éprouvée sur 4173 enfants âgés de 4 à 10 ans, examinés dans des écoles maternelles et primaires de Paris (7^e, 11^e, 15^e et 20^e arrondissements) et d'Asnières (Hauts de Seine). Parmi ces dix tests, nous avons sélectionné les six tests qui étaient les plus adaptés à l'âge de notre population (c'est-à-dire ceux destinés à des enfants de 7 ans 11 mois) et à la spécificité de notre travail (pour la description de ces tests, se reporter à l'annexe n°X et pour la justification de leur utilisation, se reporter à la partie 1.2.). La batterie que nous avons utilisée est donc composée de trois tests de discrimination portant sur des modifications harmoniques (désormais DH), des modifications mélodiques (désormais DM) et des modifications rythmiques (désormais DR) et de trois tests d'identification harmonique (désormais IH), mélodique (désormais IM) et rythmique (désormais IR), par apprentissage discriminatif (voir tableau n°17). On observe une forte corrélation ($r=0,67$) entre la note générale obtenue aux tests de discrimination et celle obtenue aux tests d'identification (voir figure n°77). Les résultats détaillés par test et par enfant sont présentés dans le Volume II, annexe 6.3.

	Harmonie	Mélodie	Rythme
Discrimination	DH (extrait n°34)	DM (extrait n°35)	DR (extrait n°36)
Identification	IH (extrait n°37)	IM (extrait n°38)	IR (extrait n°39)

Tableau n°17 : les 6 tests de musique sélectionnés dans la batterie de Zenatti, 1980 (avec DH pour test de discrimination harmonique, DM pour test de discrimination mélodique, DR pour test de discrimination rythmique, IH pour test d'identification harmonique, IM pour test d'identification mélodique et IR pour test d'identification rythmique).

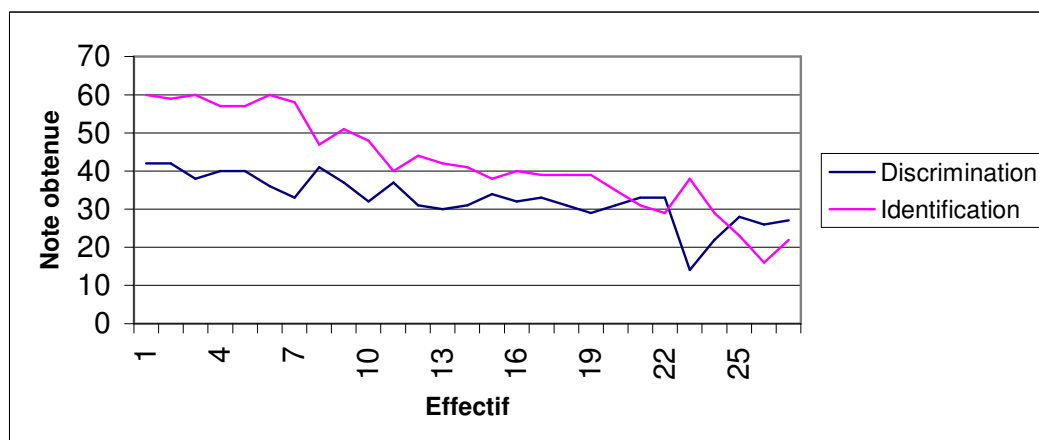


Figure n°77 : évolution simultanée de la note obtenue aux tests de discrimination et la note obtenue aux tests d'identification ($r=0,67$).

Le succès à un test reflète son pouvoir discriminant. Si tous les enfants réussissent un test, c'est qu'il n'est pas assez difficile pour être discriminant. A l'inverse, si tous les enfants échouent à un test, c'est qu'il est trop difficile pour être discriminant. Parmi les six tests de musique que tous les enfants de notre échantillon ont passés, il est possible que certains ne soient pas discriminants. Dans ce cas, il serait inutile de s'encombrer pour la suite de l'analyse de résultats non pertinents, qui ne feraient que noyer, diluer les informations intéressantes. Nous ne garderons donc que les tests dont les résultats sont en adéquation avec notre classement en groupes.

Afin de sélectionner le ou les tests les plus discriminants parmi les six tests retenus de la batterie de Zenatti, nous allons donc déterminer quelles épreuves étaient les plus difficiles, puis, une fois ces épreuves sélectionnées, regarder si elles permettent de faire une différence nette entre les résultats obtenus par les différents groupes.

2.2.3.1. Résultats aux tests de discrimination

Moyennes générales des épreuves de discrimination

Discrimination	TBO	BO	OM	OP	MO	TMO	Total	σ
DH (sur 12)	10,85	11,66	10,5	10,25	6	4,5	9,74	2,65
DM (normalisé sur 12)	10,85	10	11,62	7,75	7,83	4,88	8,33	2,67
DR (sur 12)	11,57	10	10,25	9,5	8	9,5	10,14	1,70

Tableau n°18 : moyennes obtenues par les groupes aux trois tests de discrimination, les zones grisées correspondent aux résultats obtenus par les oreilles les plus musiciennes (TBO pour très bonne oreille, BO pour bonne oreille, OM pour oreille moyenne, OP pour oreille passable, MO pour mauvaise oreille, TMO pour très mauvaise oreille, DH pour test de discrimination harmonique, DM pour test de discrimination mélodique et DR, pour test de discrimination rythmique).

L'épreuve la plus difficile pour les enfants a été le test DM. La moyenne générale ($m=8,33$) est la plus basse des trois tests de discrimination. C'est cette épreuve qui différencie le plus nettement les oreilles musiciennes (zone grisée dans le tableau n°18 : 10,85 pour TBO, 10 pour BO, 11,62 pour OM) des oreilles non musiciennes (7,75 pour OP, 7,83 pour MO et 4,88 pour TMO) parmi les trois tests de discrimination. L'effet de groupe pour ce test est en effet le plus significatif ($F=14,05$; $p=0,0000$).

Le test DH a été mieux réussi que l'épreuve DM ($m=9,74$), mais il se montre discriminant malgré tout. En effet, il permet également de différencier les oreilles musiciennes des oreilles non musiciennes, bien que la « frontière » se soit déplacée sur la droite pour intégrer les OP dans le groupe des meilleurs (voir zone grisée dans le tableau n°18).

Ainsi, d'un côté, on trouve les TBO (10,85), les BO (11,66), les OM (10,5), mais aussi les OP (10,25) et de l'autre, les MO (6) et les TMO (4,5). L'effet de groupe est légèrement moins marqué que pour le test DM, mais reste hautement significatif ($F=8,46$; $p=0,0002$).

Le test DR est l'épreuve la mieux réussie des trois tests de discrimination ($m=10,14$), mais aussi de toute la batterie. C'est donc le test le plus facile. Si nous regardons les moyennes obtenues par les différents groupes, nous ne pouvons pas placer de « frontière » comme c'était le cas dans les deux tests précédents (11,57 pour TBO, 10 pour BO, 10,25 pour MO, 9,5 pour OP, 8 pour MO et 9,5 pour TMO). Il y a pourtant un effet de groupe significatif, mais le plus faible des trois tests ($F=2,88$; $0,0391$). Ce test ne se montre pas assez discriminant, nous ne le conserverons pas.

Au vu de ces résultats, nous conserverons seulement les tests DH et DM, ces tests étant les plus discriminants pour différencier les groupes de notre échantillon, selon la variable « musique ».

2.2.3.2. Résultats aux tests d'identification

Moyennes générales des épreuves d'identification par apprentissage discriminatif

Identification	TBO	BO	OM	OP	MO	TMO	Total	σ
IH (sur 20)	19,28	18,66	1,87	5,5	5	6,5	9,48	8,55
IM (sur 20)	19,42	15,66	19	13,75	12,66	7,5	16,40	4,90
IR (sur 20)	20	14,33	19,5	14,25	12,33	5	16,40	6,09

Tableau n°19 : moyennes obtenues par les groupes aux trois tests d'identification (TBO pour très bonne oreille, BO pour bonne oreille, OM pour oreille moyenne, OP pour oreille passable, MO pour mauvaise oreille, TMO pour très mauvaise oreille, IH pour test d'identification harmonique, IM pour test d'identification mélodique et IR, pour test d'identification rythmique).

La tâche la plus difficile pour l'ensemble de l'échantillon a été la tâche d'identification harmonique (IH). La moyenne générale à cette épreuve est de 9,48 alors qu'elle est de 16,40 pour les deux autres épreuves (IM et IR). C'est également le test le plus difficile de la batterie (si on normalise les notes sur 20, la moyenne des autres tests ne se situe pas en dessous de 14,23 pour DM et n'excède pas 16,91 pour DR).

Ce test est le plus discriminant de la batterie : il a été très bien réussi par le groupe des musiciens avec des résultats proches de la perfection (19,28 pour les TBO et 18,66 pour les BO). En revanche, pour les autres, il a été très difficile, voire infaisable (1,87 pour les OM, 5,55 pour les OP, 5 pour les MO et 6,5 pour les TMO). L'effet de groupe sur ce test est d'ailleurs maximal ($F=17,35$; $p=0,0000$).

La seule « frontière » marquée mise à jour par les deux tests d'identification, IM et IH est celle qui sépare le groupe TMO des autres groupes (zones grisées dans le tableau n°19). En effet, la moyenne du groupe TMO est la plus basse (7,5 pour IM et 5 pour IR), comparativement aux autres groupes. La moyenne générale élevée de ces tests (16,40 pour IM et IR) montre que ces tests ne sont pas discriminants. On observe pourtant un effet de groupe, mais moins important que pour IH, surtout pour IR ($F=4,68$; $p=0,0050$). IM ($F=5,26$; $p=0,0028$). Il est à noter que l'effet de test est maximal sur le groupe des OM ($F=399,21$; $p=0,0000$), car ils ont échoué au test d'identification harmonique, pas aux autres. Il n'est pas significatif pour les autres groupes (TBO et BO ont réussi tous les tests ; OP, MO et TMO, résultats très divers selon les tests, non tranchés par test).

Pour notre étude, nous ne conserverons que le test IH, qui possède d'ailleurs le pouvoir le plus discriminant de toute la batterie. Il permet en effet de séparer nettement les BO et TBO des autres.

2.2.4. Les tests les plus discriminants et nouvelle définition des groupes

Sur les six tests de musique, seuls trois tests permettent de séparer nettement les oreilles musiciennes des oreilles non musiciennes : deux tests de discrimination, DH et DM et un test d'identification, IH. Nous conserverons donc seulement ces trois tests, car ils sont les plus discriminants pour différencier les groupes de notre échantillon, selon la variable « musique » (voir tableau n°20). Cet effet discriminant justifie le fait d'utiliser ces résultats dans les tests d'anglais (étude des corrélations). Les tests concernant le domaine harmonique, DH et IH permettent de tracer une frontière très nette entre les TBO et BO d'un côté et les autres groupes de l'autre. Le test DM a déplacé cette frontière sur la droite pour intégrer le groupe des OM. Pour plus de sécurité, nous conserverons la frontière tracée par les deux premiers tests pour redéfinir nos groupes. Le nouveau classement se composera donc de deux sous-groupes, le groupe 1 (désormais G1) qui intégrera les TBO et BO et le groupe 2 (désormais G2) qui intégrera les OM, OP, MO et TMO. L'effectif du premier groupe est de 10 enfants, celui du second groupe, de 17 enfants. Ce classement sera utilisé pour comparer les résultats aux tests de musique et aux tests d'anglais.

	G1	G2
DH	11,10	8,94
DM	10,59	7,33
IH	19,10	3,82

Tableau n°20 : moyennes obtenues par les deux groupes aux trois tests les plus discriminants de la batterie (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », DH pour test de discrimination harmonique, DM pour test de discrimination mélodique et IH pour test d'identification harmonique).

L'effet de groupe pour DH est très significatif ($F=4,77$; $p=0,0386$), encore accru pour DM ($F=32,83$; $p=0,0000$) et hautement significatif pour IH ($F=85,16$; $p=0,0000$). IH est donc bien le test le plus discriminant de la batterie, suivi de DM puis de DH.

2.2.5. Importance des domaines mélodiques et harmoniques pour différencier les oreilles musiciennes des oreilles non musiciennes

Il est intéressant de remarquer que, pour les tests d'identification et les tests de discrimination, le test concernant le domaine rythmique est celui où l'effet de groupe est le moins important, bien que restant significatif (discrimination : $F=2,88$; $p=0,0391$) (identification : $F=4,68$; $p=0,0050$). De plus, les tests concernant les domaines harmoniques et mélodiques sont les tests dans lesquels on peut remarquer un effet de groupe G.T. maximal (effet discriminant musiciens / non musiciens), en identification et en discrimination. La réussite à ces tests semble dépendre de la formation musicale et de la provenance d'un milieu musicien. Ce qui différencie donc le plus les oreilles musiciennes des non musiciennes, ce ne sont pas les aptitudes rythmiques, mais les aptitudes mélodiques et harmoniques. Ces résultats rejoignent ceux de Zenatti (1981 : 201) qui a montré que les domaines musicaux les plus affectés par les variables profession du père, richesse musicale du milieu, pratique musicale, étaient en tête le domaine harmonique (37 % des résultats, différences significatives entre les groupes) et le domaine mélodique (31 %). En revanche, pour le rythme, seulement 11 % des résultats étaient significatifs.

2.2.5.1. La perception de l'harmonie est étroitement liée au degré de formation de l'oreille

Le domaine de l'harmonie semble le plus influencé par le degré de formation musicale de l'oreille. Dans les trois tests ayant le pouvoir le plus discriminant de la batterie, deux concernent le domaine harmonique, DH et IH, ce dernier étant le plus puissant.

Selon Zenatti (1981 : 153), « *l'épreuve harmonique est la plus difficile, quel que soit l'âge* ». Ces tests sont conçus pour de jeunes enfants et ne devraient pas poser de problèmes à un adulte. Il est intéressant de rapprocher ces résultats de ceux d'adultes souffrant de vrais problèmes d'oreille. Nous avons cherché des personnes qui prétendent « *ne pas avoir d'oreille ou chanter comme des casseroles...* ». L'une d'entre elles, une étudiante âgée de 27 ans, affirme que dans son cas, « *c'est irrémédiable* ». Elle a suivi des cours de chant et une formation musicale en conservatoire, mais n'a jamais pu améliorer son oreille et était sujette aux quolibets de ses professeurs de musique. Le fait qu'un adulte ait conscience de son « déficit d'oreille », qu'il ait suivi une formation musicale et que l'expérience se soit soldée par un échec est caractéristique des personnes « amusiques ». De fait, elle a échoué au test d'IH car elle n'arrivait pas à entendre la différence entre les deux stimuli présentés,

« un groupe de quatre accords consonants (accords parfaits) opposé à un groupe de quatre accords nettement dissonants (voir figure n°78 et extrait sonore n°37). La partie supérieure de ces accords forme un dessin mélodique qui est le même pour les deux fragments musicaux » (Zenatti, 1981 : 97).

Le seul moyen de lui faire prendre conscience de la différence a été de lui faire remarquer que la suite d'accords dissonants provoquait un sentiment d'angoisse, apaisé par la suite d'accords consonants.

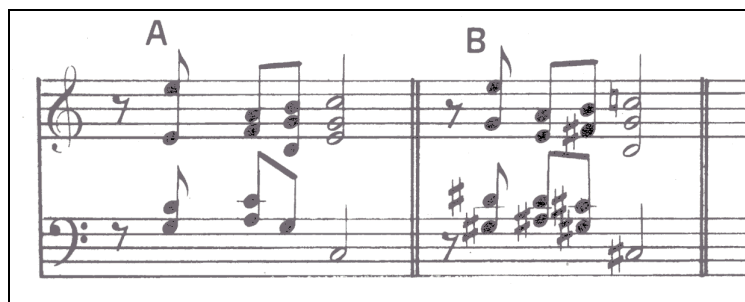


Figure n°78 : transcription musicale d'un item comprenant un groupe d'accords consonants (A) et un groupe d'accords dissonants utilisés dans le test IH (emprunté à Zenatti, 1980 : 37) (IH, test d'identification harmonique).

D'après nos résultats, il semblerait que la réussite à ce test semble dépendre fortement de la qualité « musicale » de l'oreille.

« Dans une épreuve d'identification harmonique, il coexiste deux dimensions : l'une est la consonance, dimension dont l'enfant doit saisir la pertinence pour réussir à différencier les structures, l'autre est le dessin mélodique, dimension non pertinente mais qui, du fait de sa prégnance, peut éventuellement entraver la perception harmonique » (Zenatti, 1981 : 97).

Sans formation musicale, c'est la dimension mélodique qui l'emporte sur la dimension harmonique. Une oreille qui n'entend pas la différence entre les deux suites d'accords, c'est une oreille qui n'est pas capable de se libérer de la prégnance de la mélodie. Une oreille « musicienne », plus analytique est capable de porter son attention sur les accords et d'en analyser la structure (cf. chapitre III, 4.2. « l'oreille musicienne »). Cette capacité se développe progressivement au cours de la formation musicale, mais l'environnement musical de l'enfant peut participer à la formation de son oreille, par imprégnation inconsciente. Selon Zenatti (1981 : 111),

« Il lui fournit l'occasion d'entendre des accords plus ou moins consonants. Des habitudes perceptives contractées alors exercent une influence sur la discrimination d'un changement harmonique qui est très significativement plus aisée dans le cas d'accords consonants que dans celui d'accords dissonants, lorsque les calculs statistiques portent sur l'ensemble des enfants examinés, âgés de 5 ans à 7 ans 11 mois ».

Dans les cas extrêmes, la dimension harmonique n'est pas perçue chez les non musiciens. Il y a de fortes chances pour que cette différence se reporte dans d'autres domaines, notamment dans la perception d'une langue, qui possède des caractéristiques musicales évidentes (se reporter au chapitre V).

2.2.5.2. La perception de la mélodie est également liée au degré de formation de l'oreille

Après l'harmonie, c'est la mélodie qui semble le domaine plus influencé par le degré de formation musicale de l'oreille. Nous avons vu que le test DM avait un fort pouvoir discriminant. Dans ce test, la tâche consiste à repérer s'il y a une différence d'ambitus entre deux intervalles (qui s'étend de la tierce mineure à la neuvième majeure), et s'il y en a une, à la localiser. En relevant les items qui totalisent le plus d'erreurs dans les trois tests de discrimination, nous avons relevé pour 12 items sur 16, que le stimulus déviant était le premier de la deuxième paire. La tâche paraît donc plus difficile lorsque le stimulus déviant est présenté en premier, parce que, peut être, il est plus éloigné dans le temps au moment de la réponse. Il paraît donc plus facile de répondre quand le stimulus déviant est présenté en dernier. Regardons ce qui se passe chez les non musiciens. Dans le test de discrimination mélodique, le stimulus déviant est présenté en premier dans les items 2, 3, 4, 5, 9, 10 et 16. Et c'est là où le pourcentage d'erreur est le plus grand, pouvant aller jusqu'à 55 % de mauvaises réponses.

Quand les enfants ne savent pas, ils répondent que les deux stimuli de la paire sont identiques. Par ailleurs, les résultats obtenus par les groupes TBO et BO sont bien meilleurs que ceux obtenus par les autres groupes. En effet, ils ont commis beaucoup moins d'erreurs, mais lorsque cela arrive, ces erreurs concernent les items où le pourcentage d'erreur est le plus élevé du test. Nous avons constaté le même phénomène dans les deux autres tests de discrimination, mais de manière moins marquée que pour DM. Les musiciens réussissent donc mieux et s'ils échouent, ils échouent sur les items les plus difficiles du test.

Cette tâche pose donc un problème aux non musiciens, comme si leur attention, leur capacité de mémorisation était limitée. Leur mémoire de travail est trop courte et leur attention mal dominée. Si le stimulus est trop éloigné dans le temps, il est oublié. Dans cette tâche, le facteur temporel est important et le rôle de la mémoire immédiate accru. L'adulte dont nous parlions dans la partie précédente éprouve les mêmes difficultés lorsque les items déviants sont placés en premier. Est-ce que nous constaterons le même problème avec les tests d'anglais, où la même procédure de test a été utilisée pour les accents de mots ?

2.3. Existe-t-il d'autres classements plus pertinents ?

Le classement en deux groupes que nous avons mis au point repose uniquement sur les résultats aux tests musicaux. Mais nous savons que d'autres facteurs, notamment les facteurs sociobiographiques (Manturzewska, 1994) peuvent avoir beaucoup d'influence sur le développement des aptitudes musicales. Parmi ces facteurs, le statut socioprofessionnel des parents revêt une importance considérable par le fait qu'un enfant puisse suivre une formation musicale et par le choix de la structure dans laquelle il est formé. Pour la suite des analyses, il nous semble nécessaire de vérifier si un classement fondé sur la catégorie socioprofessionnelle serait également pertinent. Si ce facteur se montre plus déterminant dans la réussite aux tests musicaux, il constituera un biais qui empêchera d'identifier clairement les effets de la variable « musique ».

Par ailleurs, il est clair que le milieu familial joue un rôle crucial dans le développement des aptitudes scolaires de l'enfant, et en particulier celui des aptitudes musicales. Dans notre échantillon, les enfants musiciens classés dans le groupe TBO proviennent presque tous d'un milieu musical « riche » ; quant à l'unique enfant non musicien appartenant à ce groupe, son père est professeur de musique. Un classement fondé sur une évaluation de la richesse musicale du milieu n'est-il donc pas plus pertinent que celui que nous avons adopté ?

Enfin, Zenatti (1981 : 200) a montré l'importance de la pratique instrumentale. La classe musicale de notre échantillon est composée d'enfants chanteurs, mais certains d'entre eux pratiquent en plus un instrument en dehors du cadre scolaire. C'est également le cas de quelques enfants provenant des deux classes non musicales. Dans le groupe TBO, six enfants sur sept pratiquent un instrument. Il est probable que le fait qu'un enfant pratique un instrument influence les résultats aux tests de musique. Il est donc intéressant d'étudier la pertinence d'un classement fondé sur la pratique instrumentale.

2.3.1. Définition des groupes par rapport à la catégorie socioprofessionnelle

Le statut socio-économique joue un rôle très important dans la réussite scolaire. Selon Manturzevska (1994 : 274),

« on a découvert qu'un minimum de prospérité et d'aisance était une condition nécessaire de la réussite scolaire. En deçà de ce seuil minimal, il faut dépenser une énergie considérable pour résoudre divers problèmes de la vie quotidienne au détriment des résultats scolaires. Au-delà d'une certaine norme économique, il y a généralement trop de facteurs de distraction qui réduisent la motivation pour étudier et obtenir de bons résultats scolaires ».

Dans une recherche sur les facteurs prédictifs de réussite dans les écoles de musique, Manturzevska (1974) a étudié, parmi six facteurs environnementaux, le statut socioprofessionnel des parents et l'activité professionnelle des mères. Cette étude concernait trois groupes d'une centaine d'élèves. Elle a montré que 12 % d'entre eux provenaient d'un milieu ouvrier et on n'en retrouvait aucun dans les 10 % qui avaient obtenu les meilleurs scores aux tests d'aptitudes musicales. En revanche, 20 % des élèves avaient des parents qui étaient musiciens professionnels et leur proportion était de 25 % dans le groupe de tête. Dans une étude portant sur soixante-quinze enfants âgés de dix à onze ans, Sergeant et Thatcher (1974) ont montré que, dans les tests mélodiques et rythmiques, le QI et l'environnement musical familial avaient des liens étroits avec le niveau socio-économique. Parmi les variables socioculturelles, Zenatti (1981 : 192) a étudié l'influence de la profession du père sur les résultats obtenus à ses tests musicaux. Les enfants dont le père est issu d'une CSP plus élevée (cadres supérieurs) préfèrent les mélodies tonales et les harmonies consonantes démontrant une meilleure assimilation du langage musical. Ils obtiennent de meilleurs résultats dans les tests d'identification harmonique et mélodique, et surtout dans les tâches de discrimination harmonique.

Mais, sur soixante-dix résultats concernant cette variable, seulement huit se sont montrés significatifs, soit 11 %, contre 35 % pour la variable concernant la richesse musicale du milieu et 27 % pour la variable concernant la pratique instrumentale des enfants.

La rubrique « renseignements d'identité » du questionnaire (se reporter au chapitre VI) a permis de déterminer la catégorie socioprofessionnelle dans laquelle se plaçait chaque foyer. Quatre familles ont été écartées de cette analyse, car elles n'ont pas souhaité remplir cette rubrique par souci d'anonymat. Ici, notre effectif est donc réduit à 23 enfants (voir tableau n°22). Dans son étude, Zenatti (1981 : 190) a regroupé les CSP en deux catégories (1981 : 190) : cadre moyen et supérieur ou profession libérale (catégories 3 et 4 du Code des CSP), employé, ouvrier ou personnel de service (catégories 5, 6 et 7 du Code). Sur le modèle de Zenatti, mais avec quelques aménagements (dans notre échantillon, sept catégories sont représentées : les catégories 2, 3, 4, 5, 6 et 8, mais pas la catégorie 7) dus à la spécificité de notre population, nous regrouperons également les CSP en deux grandes catégories, comprenant :

- Groupe 1 (désormais « CSP1 ») : catégories 3 et 4. Effectif : 8 sujets sur 23.
- Groupe 2 (désormais « CSP2 ») : catégories 2, 5, 6 et 8. Effectif : 15 sujets sur 23.

Descriptif des catégories	Numéro de catégorie
<i>Agriculteurs exploitants.</i>	1
<i>Artisans, commerçants et chefs d'entreprise.</i>	2
<i>Cadres et professions intellectuelles supérieures (cadres de la fonction publique, professions intellectuelles et artistiques).</i>	3
<i>Professions intermédiaires (de l'enseignement, de la santé, de la fonction publique et assimilés), techniciens, contremaîtres et agents de maîtrise.</i>	4
<i>Employés (de la fonction publique, administratifs d'entreprise, de commerce), personnels des services directs aux particuliers.</i>	5
<i>Ouvriers qualifiés, non qualifiés et agricoles.</i>	6
<i>Retraités.</i>	7
<i>Autres personnes sans activité professionnelle (Chômeurs n'ayant jamais travaillé, inactifs divers dont militaires du contingent, élèves étudiants...)</i>	8

Tableau n°21 : nomenclature PCS (INSEE : 1994²).

² Institut National de la Statistique et des Etudes Economiques (1994). Nomenclature des Professions et Catégories Socioprofessionnelles PCS.

Catégorie CSP	Effectif
1	
2	3
3	2
4	6
5	5
6	6
7	
8	1
TOTAL	23

Tableau n° 22: répartition de notre population selon la CSP du père (CSP, catégorie socioprofessionnelle).

Ce classement a été réalisé à partir de la nouvelle nomenclature des professions et catégories socioprofessionnelles (PCS) de 1994 (voir tableau n°21). Zenatti (1981 : 190) a tenu compte de la profession du père pour établir son classement. Mais, depuis les années 80, la société a évolué et la féminisation de l'emploi constitue un des traits majeurs de ces vingt dernières années. Les femmes représentent en 1992 44 % de la population active contre 40 % en 1982. Dans la tranche d'âge où des 25-49 ans, 71 % des femmes sont actives contre 45 % en 1967 et c'est parmi les cadres moyens et supérieurs que le nombre de femmes s'accroît le plus rapidement. Aujourd'hui, il n'est donc plus étonnant qu'une femme appartienne à une CSP supérieure à celle de son conjoint, d'autant que les femmes ont en moyenne un niveau de formation supérieur aux hommes. Dans ce cas, on parle de couple hypogame. Alors, pourquoi continuer à prendre comme référence la profession du père ? Dans notre population, malgré 77,27 % de femmes salariées, cette situation ne concerne qu'un cas. Pour cette raison et sauf pour ce cas, nous avons donc pris en compte la CSP du père.

	CSP1	CSP2	F	p
DH	16,87	16,66	-0,02	-0,9005
DM	15,55	13,40	-2,91	-0,1028
IH	13,00	6,60	-3,10	-0,0926

Tableau n°23 : moyennes obtenues par les deux groupes répartis selon la CSP aux trois tests les plus discriminants (CSP pour catégorie socioprofessionnelle, DH pour test de discrimination harmonique, DM pour test de discrimination mélodique, IH pour test d'identification harmonique).

Une analyse de variance (Anova, test de Fisher) ne fait apparaître aucun effet de groupe sur chacun des trois tests les plus discriminants de la batterie (voir tableau n°23). La moyenne du groupe CSP1 est plus élevée pour chacun des trois tests, mais de manière non significative. Dans le cas du test IH, le test le plus discriminant de la batterie, on peut remarquer qu'on approche de la signifiante (F=-3,10 ; p=-0,0926), sans toutefois l'atteindre. Même chose pour l'effet de groupe sur les trois tests (F=2,96 ; p=0,0998).

Sur les huit enfants qui appartiennent au groupe CSP1, quatre ont été classés dans la catégorie TBO, un dans BO, deux dans OM et un dans TMO. Sur les quinze enfants qui appartiennent au groupe CSP2, deux ont été classés dans la catégorie TBO, un dans BO, six dans OM, quatre dans OP et deux dans MO. Les CSP sont donc très mélangées selon les résultats aux tests. Il est à noter que quatre TBO sur six et un BO sur deux proviennent d'un milieu socioprofessionnel favorisé (CSP1). Il y a donc un léger effet de la CSP sur nos résultats, comme le confirme l'ANOVA sur IH, mais trop faible pour être significatif.

Si cette étude avait été réalisée dans les années 60-70, il est possible que les effets de la CSP auraient été plus marqués. Mais aujourd'hui, l'accès à la musique est beaucoup plus facile : les conservatoires (CNR) et les écoles nationales de musique (ENM) sont moins chers, grâce à des aides accordées aux familles par les municipalités. Au plan national, l'effectif total d'enfants bénéficiant d'une formation musicale en ENM et en CNR suit une évolution ascendante, passant de 88 521 en 1990/91 à 89 010 en 1998/99 pour les ENM et de 46 647 à 49 328 pour les CNR³. Grâce aux ENM et aux nombreuses écoles de musique municipales ou privées, il est plus facile d'accéder à une formation musicale de qualité et surtout de proximité. Mais il existe également beaucoup d'autres structures qui permettent l'accès à une formation musicale à moindre coût et près de chez soi : maisons des jeunes et de la culture, groupes musicaux locaux⁴. La formation musicale a également progressé dans les écoles, grâce à des intervenants extérieurs. C'est le cas par exemple de l'école de Choisey où tous les enfants de l'école, du CP au CM2 bénéficient d'une séance de 45 mn par semaine.

Il n'y a pas que la pratique musicale qui se soit démocratisée, mais également les supports musicaux et le matériel d'écoute. En 2000, 61 % des ménages possédaient une chaîne HiFi⁵, bien qu'il subsistât encore de grandes disparités selon les CSP, avec un taux d'équipement s'échelonnant de 23 % chez les agriculteurs à 83 % chez les cadres et professions intellectuelles supérieures (se reporter à la figure n°79). Ce taux d'équipement a progressé dans le temps, puisque la proportion de français âgés de plus de quinze ans possédant une chaîne HiFi est passée de 8 % en 1973, à 29 % en 1981 et à 56 % en 1988⁶.

³ Source : Ministère de la Culture et de la Communication ; INSEE, 2001 : 299

⁴ En 1997, 10 % des français âgés de plus de quinze ans ont pratiqué le chant ou la musique avec une organisation ou des amis au cours des douze derniers mois (Source : Ministère de la Culture et de la Communication – INSEE, 2001 : 301).

⁵ Source : INSEE, Enquêtes Permanentes Conditions de Vie – mai 97, 98, 99, 2000 ; échantillon de 5407 ménages sur 24 millions de ménages en France.

⁶ Source : Ministère de la Culture et de la Communication – Les Pratiques Culturelles des Français, 88-89.

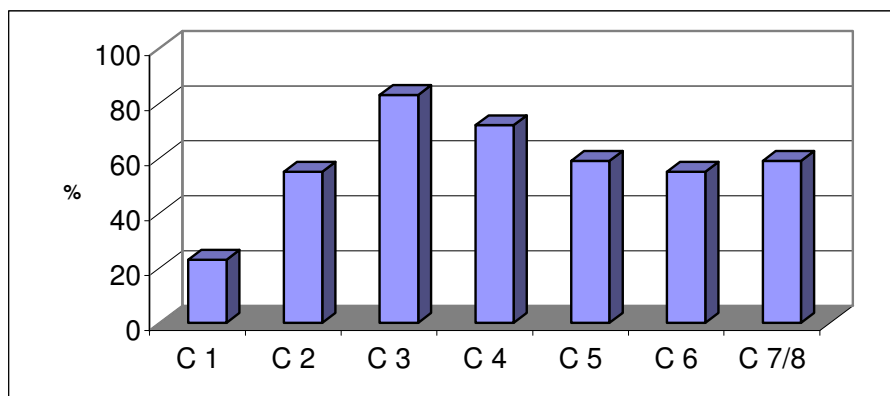


Figure n°79 : taux d'équipement en chaîne HiFi selon la CSP⁵

Source : INSEE, Enquêtes Permanentes Conditions de Vie – mai 97, 98, 99, 2000 (CSP, catégorie socioprofessionnelle).

En ce qui concerne les pratiques audiovisuelles, la proportion de français âgés de plus de quinze ans qui écoutent des disques et des cassettes est passée de 66 % en 1973 à 76 % en 1997 et la proportion de ceux qui en écoutent tous les jours, de 9 % en 1973 à 27 % en 1997⁷ (pourcentage au cours des 12 derniers mois). Parallèlement, la proportion de ceux qui écoutent la radio a légèrement augmenté (de 81 % en 1973 à 87 % en 1997), mais la proportion de ceux qui l'écoutent tous les jours a baissé (passant de 72 en 1973 % à 69 % en 1997). Selon les statistiques fournies par le Syndicat National de l'Édition Phonographique (INSEE, 2001 : 295), la vente de supports sonores (comportant les 45 T, les 33 T, les compacts, les cassettes et autres supports) connaît une progression, passant de 131 511 000 unités en 1990 à 153 305 000 en 1999. La vente de compacts a explosé passant de 56 220 000 unités en 1990 à 138 602 000 unités en 1999 (figure n°80).

⁷ Source : Ministère de la Culture et de la Communication, INSEE, 2001 : 302

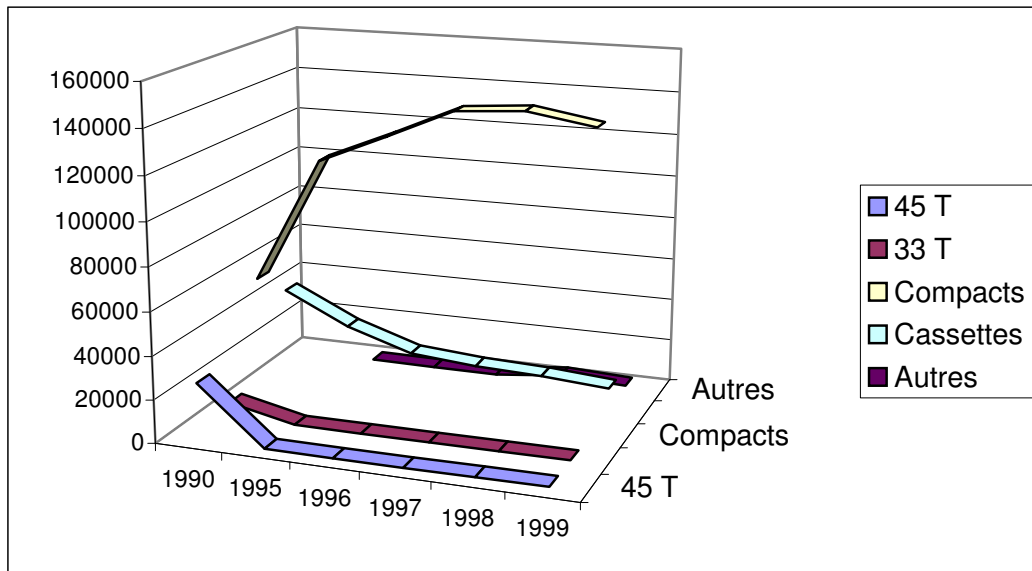


Figure n°80 : évolution des ventes des différents phonogrammes de 1990 à 1999.

La fréquentation des concerts de rock ou de jazz a progressé : la proportion de français âgés de plus de quinze ans ayant fréquenté un concert de rock ou de jazz au cours des douze derniers mois est passée de 6 % en 1973, à 10 % en 1981 et à 13 % en 1988⁸.

2.3.2. Définition des groupes par rapport à la richesse musicale du milieu

Plus que le statut socio-économique, le rôle de l'environnement familial est capital dans le développement des aptitudes musicales. Zenatti (1981 : 199) montre que l'assimilation de la langue musicale est meilleure dans un milieu musicalement riche, et ce, dans tous les domaines, harmonique, mélodique et rythmique. Les enfants élevés dans un contexte familial riche obtiennent de meilleurs résultats dans les tâches de discrimination (Lenz, 1978), dans les tâches liées au rythme et plus encore dans les tâches liées à la hauteur des sons (Moore, 1973). En ce qui concerne le chant, Kirkpatrick (1962) et Apfelstadt (1983) ont montré que la justesse vocale dépendait du niveau d'engagement musical du foyer. Dans une enquête biographique concernant 165 musiciens polonais menée par Manturzevska (1990), 93 % d'entre eux venaient de familles possédant une certaine tradition musicale. Sur douze enfants classés comme non doués par leurs professeurs de musique, Shelton (1965) a montré que onze provenaient de familles « *non-musiciennes* ». La provenance des dix-huit enfants doués, était beaucoup plus disparate.

⁸ Source : Ministère de la Culture et de la Communication – Les Pratiques Culturelles des Français, 88-89

Selon Shuter-Dyson (1994 : 220),

« les chiffres de Shelton laissent à penser que les cas atypiques d'un enfant doué pour la musique originaire d'une famille non-musicienne sont plus nombreux que l'inverse ».

Les résultats de notre expérimentation rejoignent ceux de Shelton : sur les neuf enfants qui ont obtenus les plus mauvaises notes aux tests de musique, huit sont non musiciens et proviennent d'un milieu non musicien. Sur les dix enfants de notre échantillon qui ont obtenus les meilleures notes aux tests musicaux, sept proviennent d'un milieu musicien et les deux enfants non musiciens proviennent tous les deux d'un milieu musicien (voir figure n°X).

La richesse d'un milieu musical peut se déterminer à partir d'une constellation de facteurs tels que le fait que les parents (et/ou les frères et sœurs) sont musiciens et qu'ils pratiquent leur instrument à la maison, qu'ils sont équipés d'un matériel HiFi et écoutent beaucoup de musique à la maison, qu'ils emmènent leurs enfants à des concerts...

Mais les études montrent que le facteur le plus important dans la formation des aptitudes musicales semble être le degré d'implication des parents dans l'éducation de leur enfant. Selon Manturzevska (1994 : 283),

« chaque fois que l'on a observé des rapports chaleureux et que la famille attache une grande importance à l'école et à l'éducation, les enfants progressent beaucoup mieux à l'école, indépendamment de leurs potentiels et de leurs aptitudes intellectuelles ».

Bloom (1985) remarque que les jeunes talents ont des parents très différents de par leurs études, leur travail, leur niveau économique et leurs centres d'intérêts, mais ils ont tous en commun d'être « orientés vers l'enfant », c'est-à-dire qu'ils sont prêts à leur consacrer du temps, des ressources et de l'énergie afin de leur donner les meilleures conditions d'études possibles. 90 % des élèves composant la population étudiée par Manturzevska (1974) ont

« étudié et appris la musique sous le contrôle systématique des parents, au moins jusqu'à neuf ans, 70 % jusqu'à quatorze ans » (1994 : 284).

La **question 2-11** de notre questionnaire est destinée à **déterminer si les parents sont « orientés vers l'enfant »** : « *Souhaitez-vous qu'il approfondisse son éducation musicale et si oui, dans quelle structure ? Souhaitez-vous le pousser dans cette voie ou lui laisser le libre choix de continuer ou pas ?* ».

La fonction de la **question 2-5** est à peu près équivalente : « *Si vous lui chantez des chansons, lesquelles ? Est-ce que vous chantez avec lui ? L'aidez-vous à apprendre ses chansons ?* ». Ce facteur (« *parent qui chante à, avec l'enfant* ») a été utilisé par Brand (1985) pour élaborer son échelle d'évaluation de l'environnement familial.

Sept autres questions ont été posées aux parents pour évaluer la richesse du milieu musical des enfants de notre échantillon. En fonction des réponses, nous avons attribué une note fonctionnant comme un indicateur du degré de richesse musicale du foyer. Ces questions servent à déterminer :

- **si les parents sont au courant des goûts musicaux de leur enfant et s'ils interviennent pour lui faire découvrir d'autre musiques : question 2-2** « *Quels sont les musiques, les interprètes que votre enfant préfère ?* » et **question 2-8** « *Est-ce qu'il choisit ses musiques seul ou est-ce que vous lui en faites découvrir ?* », les réponses à ces deux questions allant de pair quand les parents s'intéressent à l'éducation « de l'oreille » de leur enfant.
- **si les parents font la démarche d'emmener leur enfant aux concerts : question 3-4** « *Assistez-vous avec vos enfants à des concerts et si oui, avec quelle fréquence ?* ». Le « *fait d'aller au concert* » est un facteur qui a été également utilisé par Brand (1985) pour élaborer son échelle d'évaluation de l'environnement familial.
- **si certains membres du foyer et de l'entourage plus large de l'enfant sont musiciens : question 3-2** « *Si vous êtes musicien, quelle est votre formation musicale (en indiquant l'instrument que vous pratiquez)* » et **question 2-6** « *Une autre personne de son entourage a-t-elle des influences musicales sur lui ? Si oui, pour quelles raisons et si elle est musicienne, pouvez-vous décrire sa formation musicale, et éventuellement, l'instrument pratiqué ?* ».
- **si ces musiciens éventuels pratiquent leur instrument à la maison : question 3-3** « *Pratiquez-vous encore la musique, et si oui, dans quelle structure, avec quelle fréquence ? Est-ce que vous jouez de votre instrument à la maison ?* ».
- **si la musique écoutée par les parents est variée : question 3-1** « *Quels sont vos musiques (compositeurs, interprètes) et vos instruments préférés ?* ». Si la réponse à

cette question est détaillée et que les styles de musique écoutés sont variés, cela dévoile une plus grande ouverture musicale et sûrement une écoute quotidienne plus importante.

Nous n'avons pas retenu la question 2-7, « *l'enfant possède ou peut passer des disques ou des cassettes sans autorisation* », comme l'avait fait Brand (1985), car actuellement, beaucoup d'enfants possèdent un lecteur CD ou cassettes personnel, ce qui n'était pas vrai dans les années 80. Pour preuve, dans notre corpus, il n'y a que trois enfants sur vingt-sept qui n'en possèdent pas.

Evaluation des réponses

- question 2-2 (2) : jusqu'à 2 points si la réponse est très détaillée et le type de musique varié. Si aucun titre n'est cité, 0.
- question 2-5 (3) : si l'enfant chante tout seul, 0 ; si la réponse est affirmative, 1 ; si la réponse est plus précise que oui, 2 ; un point supplémentaire est accordé si les parents apprennent des chansons à leur enfant.
- question 2-6 (2) : si un membre de l'entourage, hors foyer exerce une influence musicale sur l'enfant, un point est attribué ; s'il est musicien professionnel ou a été formé au conservatoire, un point supplémentaire
- question 2-8 (4) : s'il choisit ses musiques seul, un point est attribué. Si ses parents lui font découvrir des musiques, 2 points. Autre personne, 1 (camarades par exemple). Si la réponse est : « *les deux* », un point, car la réponse est trop expéditive.
- question 2-11 (2) : si la réponse est libre choix, 0 ; si les parents répondent qu'ils aimeraient, mais que c'est à leur enfant de choisir, 1 point ; si la poursuite de la formation musicale dans une école de musique ou un conservatoire est envisagée, 2 points.
- question 3-1 (3) : jusqu'à 3 points si l'éventail est large et détaillé.

- question 3-2 (6) : si le père est musicien, 1 point ; s'il a été formé en conservatoire, 1 point supplémentaire. Même chose pour la mère. Si l'enfant a des frères et sœurs musiciens, 1 point ; s'ils sont ou ont été formés en conservatoire, 1 point supplémentaire.
- question 3-3 (4) : si le père pratique encore la musique, 1 point et s'il pratique à la maison, 1 point supplémentaire ; même chose pour la mère.
- question 3-4 (2) : si les parents assistent à des concerts, 1 point. Si la fréquence est supérieure ou égale à 3 fois par an, un point supplémentaire.

Si nous faisons le total de ces points, nous arrivons à une note globale de 28. D'après cette notation, nous avons réparti les sujets de notre population en deux groupes : les enfants provenant de milieux « favorisés » musicalement (que l'on appellera milieux « musiciens »), et les enfants provenant de milieux « non favorisés » musicalement, (que l'on appellera « non musiciens »). Les milieux musiciens se caractérisent par une forte orientation vers la musique (note obtenue supérieure ou égale à 12 sur 28) et les milieux non musiciens par une orientation plus faible qu'on peut qualifier de normale ou pas d'orientation du tout. On peut distinguer dans chaque groupe des sous-ensembles permettant d'affiner la description :

Groupe des milieux « musiciens », désormais « M1 » (n=10)

- Groupe « milieu très musicien » ($x \geq 12$), effectif : 4 enfants
- Groupe « milieu musicien » ($9 \leq x \leq 11$), effectif : 6 enfants

Groupe des milieux « non musiciens », désormais « M2 » (n=17)

- Groupe « milieu moyen » ($6 \leq x \leq 8$), effectif : 5 enfants
- Groupe « milieu non musicien » ($3 \leq x \leq 5$), effectif : 9 enfants
- Groupe « milieu franchement non musicien » ($x \leq 2$), effectif : 3 enfants

	M1	M2	F	p
DH	18,00	15,19	2,69	0,1132
DM	15,11	13,72	0,96	0,3370
IH	12,10	7,94	1,52	0,2292

Tableau n°24 : moyennes obtenues par les deux groupes répartis selon la richesse du milieu musical aux trois tests musicaux les plus discriminants (M1 pour le groupe des enfant provenant d'un milieu musicien, M2 pour le groupe des enfants provenant d'un milieu non musicien, DH pour test de discrimination harmonique, DM pour test de discrimination mélodique, IH pour test d'identification).

Une analyse de variance (ANOVA, test de Fisher) ne fait apparaître aucun effet de groupe sur chacun des trois tests les plus discriminants de la batterie. La moyenne du groupe M1 est plus élevée pour chacun des trois tests, mais de manière non significative. On ne constate pas non plus d'effet de groupe sur les trois tests ($F=2,71$; $p=0,1124$). L'effet de groupe devient significatif si les six tests musicaux sont pris en compte ($F=4,56$; $p=0,0428$) et on approche même de la signifiante pour le test IM ($F=3,12$; $p=0,0895$).

Si nous regardons les résultats plus en détail dans le groupe des « milieux musiciens », nous pouvons remarquer que le groupe « milieu très musicien » comprend quatre enfants tous classés dans la catégorie TBO. Dans le groupe « milieu musicien », on trouve un enfant classé dans la catégorie TBO et trois enfants, dans la catégorie OM. Un seul de ces enfants est classé dans la catégorie MO, Kevin. Mais cet enfant a posé des problèmes pendant les tests ; il ne tenait pas en place, et il manquait de concentration. Il échoue car il n'écoute pas. Cette attitude est confirmée par son professeur : Kevin est un enfant qui a peur de l'échec et il a tendance à « démissionner » dès qu'il commence à commettre des erreurs.

Dans le groupe des « milieux non musiciens », on trouve cinq OM, quatre OP, deux MO, deux TMO, deux BO et deux TBO. Lucie et Armand, classés comme TBO, proviennent de milieux totalement non musiciens. Il y a donc des exceptions et il est possible pour un enfant provenant d'un milieu non musicien de développer des aptitudes musicales. Il est à noter quand même que Lucie est instrumentiste et que la mère d'Armand chante dans une chorale.

D'après ces résultats, le milieu semble jouer un rôle sur les résultats aux tests de musique, bien que cet effet n'apparaisse pas de manière flagrante au niveau statistique. L'effet de groupe devient significatif si nous gardons les six tests de la batterie et on peut effectivement remarquer un effet de groupe sur l'un des tests que nous avons écartés, le test IM. Le milieu exerce donc une influence sur la réussite à la tâche d'identification mélodique. Par ailleurs, une corrélation apparaît ($r=0,40$) entre la note obtenue au test IH et la note caractérisant la richesse du milieu (voir figure n°81).

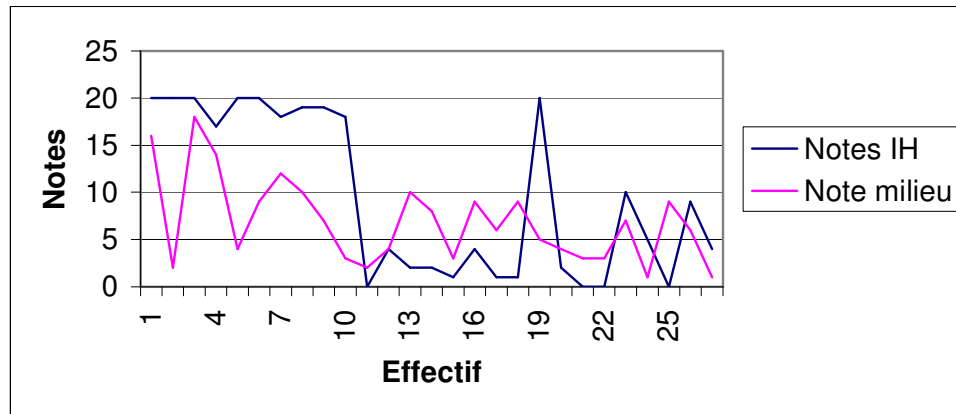


Figure n°81 : évolution simultanée de la note obtenue à IH et la note caractérisant le milieu ($r=0,40$).

2.3.3. Définition des groupes en fonction de la formation musicale et de la pratique instrumentale

Parmi les trois variables que Zenatti (1981) a étudiées pour mettre à jour les influences socioculturelles sur les résultats obtenus à ses tests, c'est la pratique musicale qui s'est révélée être la plus déterminante. En effet, sur trente-sept résultats concernant l'influence de la pratique musicale, dix se sont montrés significatifs, soit 27 %. Par ailleurs,

« seule la pratique musicale fait apparaître des résultats significatifs dans une expérience de discrimination mélodique et une expérience mettant en évidence le phénomène de fusion des sons. Ces résultats s'expliquent facilement. Lorsque l'enfant apprend à jouer un morceau de musique, il doit analyser les sons qu'il produit afin de repérer les fautes qu'il commet. Cet apprentissage affine sa discrimination perceptive, surtout dans le domaine mélodique qui est particulièrement travaillé dans les débuts de l'éducation musicale » (Zenatti, 1981 : 200).

Pour étudier l'influence de cette variable sur notre population, nous avons établi deux types de classement. Dans le premier classement, les enfants sont répartis en trois sous groupes **en fonction de la nature de leur formation musicale** : un groupe de neuf enfants instrumentistes, un groupe de six enfants chanteurs, mais non instrumentistes (scolarisés à la Maîtrise de Dijon) et un groupe de douze enfants non musiciens. Pour ce classement, l'effet de groupe frôle la signifiante ($F=3,36$; $p=0,0515$) et l'atteint pour le test DH ($F=3,81$; $p=0,0366$). Si nous conservons les six tests musicaux, l'effet de groupe devient significatif ($F=6,38$; $p=0,0060$). Il l'est également pour le test IR ($F=3,88$; $p=0,0345$).

Dans un deuxième classement plus drastique, nous avons réparti les enfants **en fonction de leur pratique instrumentale** : un groupe de neuf instrumentistes (appelé « I ») et un groupe de dix-huit enfants rassemblant les enfants chanteurs et les enfants non musiciens (appelé « NI »). Pour ce classement, l'effet de groupe sur l'ensemble des tests est significatif ($F=6,25$; $p=0,0194$) et on constate qu'on approche de la signifiante pour chacun des trois tests les plus discriminants (tableau n°25). Comme pour le classement en trois groupes, si nous conservons les six tests musicaux, l'effet de groupe est encore plus significatif ($F=9,89$; $p=0,0043$). Il l'est également pour le test IR ($F=5,11$; $p=0,0328$).

	I	NI	F	p
DH	18,51	15,09	4,01	0,0561
DM	15,92	13,39	3,32	0,0805
IH	13,66	7,38	3,55	0,0711

Tableau n°25 : moyennes obtenues par les deux groupes répartis selon la pratique instrumentale aux trois tests musicaux les plus discriminants (I pour le groupe des instrumentistes, NI pour le groupe des non instrumentistes, DH pour test de discrimination harmonique, DM pour test de discrimination mélodique, IH pour test d'identification harmonique).

Si la formation musicale influence de manière presque significative les résultats aux trois tests les plus discriminants de la batterie de Zenatti (1980), c'est l'effet de la pratique instrumentale qui se montre le plus déterminant.

2.4. Conclusion

Une répartition des enfants de notre échantillon en fonction de leur appartenance ou non à une classe de type « musicale » semblait au départ le classement le plus approprié pour isoler l'effet de la variable musique. Ce groupement est le plus courant dans les études comparant les performances de sujets musiciens par rapport à celles de sujets non musiciens (cf. études exposées dans le chapitre III, 4.2.1.). Cependant, les spécificités de notre échantillon nous ont amenée à redéfinir la constitution des groupes en répartissant les enfants, non pas en fonction de leur appartenance à une classe de type musical, mais en fonction de leurs résultats aux tests de la batterie de Zenatti (1980) et notamment, aux trois tests les plus discriminants (DH, DM et IH). Nous avons ainsi constitué deux échantillons : les enfants appartenant au groupe G1 ($n=10$) considérés comme ayant une bonne oreille « musicienne » et les enfants appartenant au groupe G2 ($n=17$) considérés comme ayant une « mauvaise » oreille musicienne. La dimension mélodique et la dimension harmonique en particulier (fortement liée au degré de formation musicale) se sont révélées déterminantes pour différencier les oreilles musiciennes des oreilles non musiciennes.

C'est en fonction de ce groupement que les résultats aux tests de perception (études n°1 et 2, chapitre VII) et de productions en anglais (études n°3 et n°4) des enfants de notre échantillon seront analysés.

D'après la littérature, un certain nombre de critères se montrent déterminants dans le développement des aptitudes musicales chez l'enfant. Parmi ces critères, la richesse du milieu musical, la pratique musicale et enfin la pratique instrumentale influencent de manière croissante les résultats des enfants de notre échantillon aux tests de musique. En revanche, la CSP, pourtant désignée d'ordinaire comme un critère déterminant, ne semble pas affecter ces résultats. Il est possible que son influence s'exprime de manière indirecte par l'intermédiaire d'autres critères tels que la richesse musicale du milieu (les milieux favorisés pouvant consacrer plus d'argent pour satisfaire leur « consommation musicale »). Cependant, la démocratisation de l'accès à la musique qui intervient depuis les années 60 pourrait limiter cette influence.

L'incidence de ces facteurs sur la formation de l'oreille peut conduire à trouver des enfants musiciens dans des classes définies a priori comme « non musicales » et à l'inverse, des enfants ayant une mauvaise oreille dans des classes définies a priori comme « musicales ». Les tests, en écartant ces cas de figure, permettent une relative maîtrise de la situation expérimentale et placent les enfants de notre échantillon sur un même plan. Le classement adopté offre donc l'avantage de répartir les enfants selon des critères « objectifs ».

Toutefois, il est impossible d'évacuer complètement les influences de l'effet Hawthorne dans un tel montage expérimental (effet que nous avons présenté dans le chapitre III, 4.2.1. ; pour une description détaillée, voir également Volume II, annexe n°7.1.). En effet, parmi les trois classes que nous avons sélectionnées, l'une se caractérise par une activité supplémentaire par rapport aux deux autres classes : la pratique quotidienne du chant. Dans un tel cas de figure, le chercheur doit impérativement se demander quel est le poids de cette activité particulière sur la perception qu'en a l'entourage, c'est-à-dire les parents, les enseignants et l'institution scolaire elle-même. En effet, il est possible que le climat de valorisation dans lequel baignent les enfants à cause de leur appartenance à une classe « spéciale » influence positivement leurs résultats. Dans cette situation, il est donc difficile d'isoler les effets de la variable étudiée (dans notre étude, la « qualité musicale » de l'oreille).

En outre, il est impossible d'empêcher les transferts positifs que nous pouvions faire sur les enfants musiciens à cause de notre statut personnel de musicienne et de nos hypothèses de travail (ces enfants musiciens devraient mieux « percevoir » et mieux « reproduire » le système prosodique et vocalique de l'anglais). L'interprétation de nos résultats tiendra donc compte de l'importance de ces facteurs, facteurs qu'il ne faudrait jamais sous-estimer dans une étude de ce type.

Bibliographie chapitre VI

- Andrews, Moya L. and Madeira, Shelley S. (1977). The assessment of pitch discrimination ability in young children. In *Journal of Speech and Hearing Disorders*, 42/2, 279-286.
- Apfelstadt, H. E. (1983). *An investigation of the effect of melodic perception instruction on the pitch discrimination and vocal accuracy of kindergarten children*. PhD, Université de Wisconsin-Madison, 1983.
- Bentley, A. (1966). *Musical Ability in Children and its Measurement*. Londres: Harrap.
- Bloom, B. S. (1985). *Developing Talent in Young People*. New-York : Bellentine Books.
- Brand, M. (1985). Development and validation of the home musical environment scale for use at the early elementary level. In *Psychology of Music*, 13, 40-48.
- Dodane, Christelle (1996). *Traitement Comparé du Langage et de la Musique : Contribution de la Musique au Développement Linguistique de l'Enfant*. Mémoire de Maîtrise, Université de Franche-Comté, Besançon.
- Dodane, Christelle (2000). L'apprentissage précoce d'une langue étrangère : une solution pour la maîtrise de l'intonation et de la prononciation ? In Guimbretière, 229-248.
- Garabédian, Michèle (1996). Apprendre les langues étrangères le plus tôt possible à l'école primaire : pourquoi ? Pourquoi faire ? Dans *Les Langues à l'Ecole, un Apprentissage ? Actes du Colloque IUFM*, 13 et 14 mars, Dijon.
- Gordon, Edwin E. (1979). Developmental music aptitude as measures by the primary measures of music audition In *Psychology of Music*, 7/1, 42-49.
- Guberina, Pierre (1991). Rôle de la perception auditive dans l'apprentissage précoce des langues. Dans *Le Français dans le Monde*, août-septembre, 65-70.
- Guimbretière, Elizabeth (2000). *Apprendre, Enseigner, Acquérir : la Prosodie au Cœur du Débat*. Rouen : Publications de l'Université de Rouen, 312 p.
- Hagège, Claude (1996). *L'Enfant aux Deux Langues*. Paris : Odile Jacob, 298 p.
- Kirkpatrick, W. C. (1962). *Relationships between the singing ability of prekindergarten children and their home environment*. PhD, Univ. Southern California.
- Institut National de la Statistique et des Etudes Economiques (1994). *Nomenclature des Professions et Catégories Socioprofessionnelles PCS*. 2^e éd. Paris : Imprimerie Nationale, 420 p.
- Institut National de la Statistique et des Etudes Economiques (2001). *Annuaire Statistique de la France : résultats de 1999*. Paris : Imprimerie Nationale. 104^{ème} Ed.
- Journiac, Claire (2002). *La Place de l'Autre dans la Relation Soignante : Contribution à l'Etude des Processus Relationnels à Travers la Voix*. Université de Bourgogne, 360p.
- Kwalwasser, J. and Dykema, P. W. (1930). *Music Tests. Manual of Direction*. New-York: Fisher.
- Landercy, Albert et Renard, Raymond (1977). *Eléments de Phonétique*. Bruxelles : Didier, 214 p.
- Lenz, S. M. (1978). *A case study of the musical abilities of three-and four-year-old children*. PhD, Univ. Illinois, 1978. Compte rendu de D. R. Tanner in Bulletin of the Council for Research in Music Education, 65, 43-46.
- Luria, Alexander R. (1974). *L'enfant Retardé Mental*. Paris : Privat, 254 p.
- Manturzewska, Maria (1974). Psychologiczne wyznaczniki powodzenia w studiach muzycznych [Facteurs psychologiques permettant un pronostic de réussite dans les études musicales], Varsovie, COPSA.
- Manturzewska, Maria (1990). A biographical study of the life-span development of professional musicians. In *Psychology of Music*, 18, 112-139.
- Manturzewska, Maria (1994). Les facteurs psychologiques dans le développement musical et l'évolution des musiciens professionnels. In Zenatti, 259-290.
- Ministère de l'Education Nationale (2001). *Note d'information* 01-18, avril.
- Moore, D. L. (1973). *A study of pitch and rhythm responses of five-year-old children in relation to their early*

music training. PhD, Florida, State Univ.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. Dans *Travaux du Congrès de Prague*, Den Haag, 4, 31, 79-96.

Rosenthal, Robert, A. and Jacobson, Lenore (1968). *Pygmalion in the Classroom*. New-York : Holt, Rinehart and Winston, Inc. *Pygmalion à l'école* (1971). Casterman, traduit de l'américain par Audebert, Suzanne et Rickards, Yvette, 293 p.

Rostand, Jean (1952). *L'Hérédité Humaine*. Paris : PUF (10ème ed. 1975), 128 p.

Sergeant, D. C. and Thatcher, G. (1974). Intelligence, social status and musical abilities. In *Psychology of Music*, 2, 32-57.

Shelton, J. S. (1965). *The influence of home musical environment upon musical response of first-grade children*. PhD, Nashvill, Peasbody, College for Teachers.

Shuter-Dyson, Rosamund (1994). Le problème des interactions entre hérédité et milieu dans la formation des aptitudes musicales. In Zenatti, 205-231.

Valette, Rebecca, M. (1975). *Le Test en Langue Etrangères : Guide Pratique*. Paris : Hachette, 111 p.

Wing, H. D. (1941). *Musical Ability and Appreciation*. PhD Thesis, London University.

Zenatti, Arlette (1980). *Tests Musicaux pour Jeunes Enfants, avec Applications en Psychopathologie de l'Enfant et de l'Adulte*. Issy Les Moulineaux : Editions Scientifiques et Psychologiques, 53 p.

Zenatti, Arlette (1981). *L'Enfant et son Environnement Musical. Etude Expérimentale des Mécanismes Psychologiques d'Assimilation Musicale*. Issy Les Moulineaux : Editions Scientifiques et Psychologiques, 275 p.

Zenatti, Arlette (1994). *Psychologie de la Musique*. Paris : PUF, 391 p.

Chapitre VII

Effets de la formation musicale sur la perception de l'anglais

« La perception est beaucoup plus qu'une impression des organes des sens : elle est la représentation, par le moyen de cette impression, d'un objet externe en un lieu de l'espace » (Pradines, Traité de psychologie générale, I, p. 400).

Table du septième chapitre

1) Etude n°1 : « Perception de la prosodie »

- 1.1. Plan expérimental
 - 1.1.1. Hypothèse
 - 1.1.2. Variable indépendante « musique »
 - 1.1.3. Variables dépendantes
- 1.2. Passation
 - 1.2.1. Cas particuliers
 - 1.2.2. L'influence de l'expérimentateur sur les résultats aux tests
- 1.3. Résultats
 - 1.3.1. Moyennes générales aux tests : les tests les plus discriminants
 - 1.3.2. Est-ce qu'il y a transfert de compétence de la pratique musicale sur les résultats obtenus aux tests concernant la prosodie ?
 - 1.3.3. Corrélations entre les tests de musique et les tests d'anglais de la batterie « Prosodie »
- 1.4. Conclusion de l'étude n°1 « Perception de la prosodie »

2) Etude n°2 « Perception des contrastes vocaliques »

- 2.1. Plan expérimental
 - 2.1.1. Hypothèse
 - 2.1.2. Variable indépendante « musique »
 - 2.1.3. Variables dépendantes
- 2.2. Passation
- 2.3. Résultats généraux
 - 2.3.1. Moyennes générales aux tests : les tests les plus discriminants
 - 2.3.2. Est-ce qu'il y a transfert de compétence entre la pratique musicale et les résultats obtenus aux tests concernant les contrastes vocaliques ?
 - 2.3.3. Corrélations entre les tests de musique et les tests d'anglais de la batterie « Contrastes vocaliques »
 - 2.3.4. Conclusion de l'étude n°2 « Contrastes vocaliques »
- 2.4. Résultats détaillés des tests de perception
 - 2.4.1. Résultats détaillés par items
 - 2.4.1.1. T6 : différenciation phonétique au sein de l'anglais (« Contrastes NN »)
 - 2.4.1.2. T7 : opposition phonétique entre le français et l'anglais (« Contrastes N/NN »)
 - 2.4.1.3. T8 : opposition phonémique entre le français et l'anglais (« Diphtongues, triptongues »)
 - 2.4.1.4. Contrastes vocaliques les mieux perçus par les musiciens

2.4.2. Résultats détaillés par enfant

2.4.2.1. Très bonne réussite chez les enfants appartenant à la catégorie « TBO »

2.4.2.2. Grande proportion d'enfants appartenant au groupe 1 parmi les meilleurs

2.4.2.3. Grande proportion d'enfants appartenant au groupe 2 parmi les plus mauvais

2.4.2.4. Résultats aux tests « Prosodie » en fonction des résultats aux tests « Contrastes Vocaliques » : interaction du global avec le local

2.5. Test de l'effet des autres classements sur les résultats aux tests de perception en anglais

2.5.1. Catégories socioprofessionnelles

2.5.2. Milieu musical

2.5.3. Pratique instrumentale

2.5.4. Conclusion

3) Conclusion et discussion générale des résultats

Bibliographie du chapitre VII

S'il existe un transfert de compétence qui permette à une oreille entraînée par la pratique musicale de mieux percevoir, ce qui devrait amener une meilleure production, le classement en deux groupes adopté pour les tests musicaux devrait se montrer pertinent pour les tests d'anglais.

Nous posons comme première hypothèse générale que la pratique musicale induit une meilleure perception d'une langue étrangère : « *Une oreille affinée par la pratique musicale pourra mieux percevoir une langue étrangère* ». Cette hypothèse peut être appréhendée en deux hypothèses particulières :

- L'hypothèse particulière « Prosodie » : la pratique musicale induit une meilleure perception de la prosodie d'une langue étrangère.
- L'hypothèse particulière « Contrastes vocaliques » : la pratique musicale induit une meilleure perception des contrastes vocaliques d'une langue étrangère.

1) Etude n°1 : « Perception de la prosodie »

1.1. Plan expérimental

1.1.1. Hypothèse

Le but de cette expérimentation est de déterminer si la pratique musicale induit une meilleure perception de la prosodie d'une langue étrangère.

1.1.2. Variable indépendante « musique »

Dans le chapitre VI, nous avons réparti les enfants composant notre échantillon au sein de deux groupes, en fonction de leurs résultats aux tests musicaux de Zenatti (1980) : le groupe G1 (comprenant les enfants « musiciens ») dont l'effectif se porte à 10 sujets et le groupe G2 (comprenant les enfants non musiciens) dont l'effectif se porte à 15 sujets. La variable indépendante « musique » est donc constituée de deux modalités, la modalité « musique » (G1) et la modalité « non musique » (G2).

1.1.3. Variables dépendantes

A partir de l'analyse des interférences intervenant entre les caractéristiques prosodiques du français et de l'anglais (chapitre V, 2), nous avons réalisé une batterie de cinq tests de perception, la batterie « Prosodie » (pour une description détaillée, voir Volume II, annexe n°4.2.1. et pour leur écoute, CD2). Nous étudierons donc cinq variables dépendantes (voir tableau n°26 pour le plan expérimental de l'étude n°1) :

- la variable dépendante « Schémas intonatifs », correspondant au test T1 ;
- la variable dépendante « Place du pivot », correspondant au test T2 ;
- la variable dépendante « Régularité accentuelle », correspondant au test T3 ;
- la variable dépendante « Tons mélodiques », correspondant au test T4 ;
- la variable dépendante « Accents de mots », correspondant au test T5.

« Prosodie »	T1 « Schémas intonatifs »	T2 « Place du pivot »	T3 « Régularité accentuelle »	T4 « Tons mélodiques »	T5 « Accents de mots »
Groupe G1					
Groupe G2					

Tableau n°26 : plan expérimental de l'étude n°1
(G1 pour « groupe musiciens » et G2 pour « groupe non musiciens »).

1.2. Passation

Les cinq tests ont été administrés aux vingt-sept enfants composant notre échantillon par le même expérimentateur, dans des conditions identiques. Pour éviter qu'un adulte exerce des influences négatives, quelles qu'elles soient, les inspections académiques interdisent qu'il reste seul avec un enfant. La passation des tests doit donc se faire avec deux enfants au minimum. Nous avons donc doublé le matériel, de manière à pouvoir faire passer les tests d'anglais à deux enfants simultanément. Un paravent a été placé entre les deux enfants afin qu'ils ne puissent pas s'influencer dans leurs réponses. Par ailleurs, les enfants répondent en désignant un des animaux qui figurent sur le plateau de jeu (voir description de la procédure de test). Ces réponses motrices évitent le recours aux réponses verbales, qui auraient inévitablement influencé l'enfant répondant en second. Tous ces tests ont été administrés dans une salle de l'école, non bruyante et connue des enfants. Le matériel utilisé a été le même pour tous les enfants de l'échantillon : magnétophone à K7 « Marantz PMD222 », paravent, plateaux de jeux (voir Volume II, annexe n°4.4.).

1.2.1. Cas particuliers

Il est nécessaire de revenir sur le déroulement des épreuves pour expliquer certains cas particuliers. En effet, il nous semble que l'attitude pendant la passation des tests de trois enfants composant notre échantillon a directement influencé leurs résultats.

Violette avait des problèmes de concentration pendant les épreuves d'anglais et son attention était plutôt centrée sur l'expérimentateur, à qui elle posait toutes sortes de questions. Son professeur a confirmé que Violette était une enfant très curieuse, mais qu'à cause de cela, elle se dispersait vite et avait des problèmes de concentration en classe (distractions sur d'autres choses). Malgré de très bons résultats en musique, Violette se classe comme une des plus mauvaises dans les tests de perception en anglais.

Une autre enfant a eu cette attitude, Maude, cette dernière a d'ailleurs obtenu les plus mauvais résultats en anglais (23 sur 62) alors qu'elle est musicienne. Cette enfant était également centrée sur l'expérimentateur. Tout se passait comme si elle profitait de l'attention de l'expérimentateur qui était centrée sur elle pendant les tests. Il semble qu'elle avait besoin d'attention, d'affectivité. Après une discussion avec son professeur, nous avons appris que sa situation familiale était complexe. Cela peut expliquer ses mauvais résultats en anglais, car elle est classée comme OM (72 sur 102 en musique).

Une toute autre attitude était celle de Kevin. Il était dissipé et dès qu'il se trompait, il « démissionnait » et ne tenait plus en place, surtout pendant les tests de musique. Il a obtenu des résultats très mauvais (classé comme MO avec 51 sur 102), bien qu'il appartienne à la Maîtrise de Dijon. Cependant, pour l'anglais, il a obtenu 50 sur 62. Quand on fait passer des tests à des enfants, on remarque ceux qui ont des problèmes d'oreille (ils réfléchissent beaucoup avant de répondre, mais sont très concentrés). Nous n'avons pas eu ce sentiment avec Kevin : il réussit quand il est concentré, mais le problème est qu'il se dissipe très vite, n'écoute plus et répond au hasard. Son professeur confirme que Kevin a des gros problèmes d'attention en classe, qu'il ne tient pas en place et qu'il a très peur de l'échec.

1.2.2. L'influence de l'expérimentateur sur les résultats aux tests

Ces trois enfants sont musiciens. Ceci n'est pas volontaire et nous n'essayons pas de justifier les mauvais résultats obtenus par des musiciens pour aller dans le sens de nos hypothèses. Il se trouve que c'est nous-même qui avons fait passer les tests et que pendant la passation, nous remarquons tout de suite la différence de comportement entre les enfants qui avaient réellement des difficultés à discriminer et les enfants qui n'étaient pas concentrés sur la tâche. La question reste de savoir pourquoi ces trois enfants étaient déconcentrés. Alors que le manque d'attention et la dissipation de Kevin s'expliquent par son comportement habituel en classe, celui de Violette et de Maude semble provenir d'une trop grande centration sur l'expérimentatrice. Nous n'avons pas constaté d'effet semblable pendant la passation des tests de musique. A ce stade de l'expérimentation, les enfants nous rencontraient pour la première fois, ce qui maintenait une certaine distance. Pour les tests d'anglais, nous étions déjà devenue une personne familière, faisant momentanément partie du « paysage scolaire ». Or, nous avons vu avec l'effet Hawthorne à quel point l'influence de l'expérimentateur peut se montrer déterminante sur les résultats des enfants et favoriser leurs réussites (chapitre III, 4.2.1. ; pour une description détaillée de l'effet Hawthorne et de Pygmalion, se reporter au Volume II, annexe n°7.1.). Chez Violette et chez Maude, l'effet est manifestement inverse, puisqu'elles ont échoué aux tests d'anglais. Leur réaction à l'attention de l'expérimentatrice a été de centrer à leur tour leur propre attention sur cette dernière et non sur les tests.

Consciente du poids de l'expérimentateur sur les résultats obtenus, nous avons essayé de le limiter au maximum. Mais, nous savons qu'il est en partie inconscient et se communique notamment par des indices non verbaux. Pour que notre classement reste néanmoins le plus objectif possible, nous avons gardé la totalité des vingt-sept enfants de notre échantillon initial pour les analyses statistiques, sans écarter les trois sujets cités. S'il existe un effet de la musique sur la perception de l'anglais, les résultats de trois sujets ne suffiront pas à modifier la tendance générale sur un échantillon de vingt-sept enfants.

1.3. Résultats

1.3.1. Moyennes générales aux tests : les tests les plus discriminants

« Prosodie »	T1 « Schémas intonatifs »	T2 « Place du pivot »	T3 « Régularité accentuelle »	T4 « Tons mélodiques »	T5 « Accents de mots »	Moyenne générale
Moyenne sur 10	8,20	8,30	7,83	5,87	7,83	7,60
σ	1,14	1,41	1,17	1,36	1,51	

Tableau n°27 : moyennes obtenues par l'ensemble des enfants aux cinq tests de la batterie « Prosodie ».

Les épreuves concernant la batterie de tests « prosodie » sont bien réussies dans l'ensemble (moyenne générale de 7,60/10). La note moyenne générale des enfants de notre échantillon est de 24,25 sur 32. Les épreuves les mieux réussies, donc les plus faciles pour les enfants, sont les épreuves T1 portant sur la discrimination des différents schémas intonatifs de l'anglais (m=8,20) et T2 portant sur la place du pivot au sein de l'énoncé en anglais (m=8,30). Les moyennes de l'épreuve T3 concernant la discrimination de la rythmicité accentuelle de l'anglais par rapport à la rythmicité syllabique du français (m=7,83) et de l'épreuve T5 concernant la discrimination de l'accent de mot (m=7,83) sont également assez élevées.

L'épreuve T4 a posé beaucoup plus de difficultés aux enfants que les autres épreuves. La moyenne générale obtenue à cette épreuve, portant sur la discrimination des différents tons mélodiques de l'anglais est la plus faible de l'ensemble des cinq tests de la batterie (m=5,87) et les notes individuelles les plus dispersées ($\sigma = 1,36$) (voir tableau n°27). Par ailleurs, ce test est également le plus difficile des huit tests composant les deux batteries d'anglais, « prosodie » et « contrastes vocaliques ».

Puisque les résultats ont été très bons pour les quatre tests T1, T2, T3 et T5, il y a de fortes chances pour qu'ils ne se montrent pas discriminants pour différencier les deux groupes, G1 et G2. En revanche, T4 est une épreuve difficile, qui peut se montrer discriminante, à condition qu'elle ne soit pas trop difficile.

1.3.2. Est-ce qu'il y a transfert de compétence de la pratique musicale sur les résultats obtenus au tests concernant la prosodie ? Pertinence du classement

« Prosodie »	T1 Schémas intonatifs	T2 Place du pivot	T3 Régularité accentuelle	T4 Tons mélodiques	T5 Accents de mots
Groupe G1	8,49	8,42	8,50	6,42	8,00
Groupe G2	8,03	8,23	7,45	5,54	7,74
F	0,36	0,06	1,88	1,30	0,06
p	0,5541	0,8152	0,1821	0,2650	0,8048
σ	1,14	1,41	1,17	1,36	1,51

Tableau n°28 : moyennes des notes (sur 10) et signifiante des cinq tests « Prosodie » en fonction du classement retenu pour les tests de musique (G1 pour « groupe musiciens » et G2 pour « groupe non musiciens »).

Malgré un effet de test hautement significatif ($F(4,100)=8,03$; $p=0,0000$), une ANOVA (test de Fisher) ne fait apparaître aucun effet de groupe significatif ($F(1,25)=1,17$; $p=0,2889$). En effet, on n'observe pas de différences significatives entre les deux groupes en ce qui concerne les cinq tests se rapportant à la prosodie, bien que les moyennes générales de G1 soient légèrement plus élevées ($m=7,65$) que celles de G2 ($m=7,39$).

L'analyse de variance ne révèle aucun effet de groupe significatif sur chacun des cinq tests (tableau n°28). Il n'y a donc aucune influence du classement sur les résultats aux tests concernant la prosodie, y compris pour le test le plus difficile de la batterie, T4. Si nous regardons de plus près les moyennes obtenues par chacun des groupes, nous trouvons l'explication de cette absence de signifiante : les moyennes sont trop proches les unes des autres. Par exemple, pour T1, G1 a obtenu une moyenne de 8,49 et G2, une moyenne de 8,03. T1, T2, T3 et T5 sont donc des tests non discriminants car ils sont trop pour notre population faciles (les deux groupes ayant obtenu de bonnes moyennes) et ne permettent pas de faire la différence entre les oreilles musiciennes et les oreilles non musiciennes. A l'inverse, le test T4 est trop difficile pour les deux groupes, G1 ayant obtenu une moyenne de 6,42 et G2, 5,54. La différence entre ces deux moyennes n'est pas significative. Ce test n'est donc pas discriminant parce qu'il est trop difficile. Aucun test de la batterie « prosodie » ne se montre discriminant pour la perception des phénomènes prosodiques de l'anglais, car soit les moyennes sont trop élevées, soit, comme dans le cas de T4, elles sont trop faibles.

1.3.3. Corrélations entre les tests de musique et les tests d'anglais de la batterie « Prosodie »

Tests	T1 Schémas intonatifs	T2 Place du pivot	T3 Régularité accentuelle	T4 Tons mélodiques	T5 Accents de mots
DH	-0,09	0,59	0,17	0,17	0,18
DM	0,12	0,38	0,24	0,17	-0,10
DR	-0,11	0,21	0,10	0,30	-0,14
IH	0,20	0,34	0,25	0,14	0,09
IM	-0,37	0,15	-0,03	0,14	0,04
IR	-0,17	0,16	0,20	0,21	0,00

Tableau n°29 : coefficients de corrélation des tests d'anglais avec les tests de musique (N=27 ; coefficient de liberté de 25 ; quand $r \geq 0,38$; $p = .05$, cellules grisées ; DH pour discrimination harmonique, DM pour discrimination mélodique, DR pour discrimination rythmique, IH pour identification harmonique, IM pour identification mélodique, IR pour identification rythmique).

On observe une forte corrélation entre la note obtenue au test T2 « Place du pivot » et la note obtenue au test musical DH ($r=0,59$, voir tableau n°29). Nous trouvons au contraire une corrélation plus faible entre cette note et la note obtenue au test musical DM ($r=0,38$). Aucune autre corrélation n'est observée entre les autres tests musicaux et les autres tests d'anglais composant la batterie « prosodie ». On observe donc peu de différence de moyenne entre G1 et G2. Rappelons que DH et DM figurent parmi les trois tests les plus discriminants de la batterie « Prosodie ».

1.4. Conclusion de l'étude n°1 « Perception de la prosodie »

Ces résultats révèlent qu'il n'y a pas de transferts de compétence de la pratique musicale à la perception de la prosodie de l'anglais, car l'hypothèse particulière n'est pas vérifiée par nos résultats. La pratique musicale ne semble pas induire une meilleure perception de la prosodie d'une langue étrangère. Cependant, d'autres critères auraient peut-être dû être pris en compte, puisque nos tests sont soit trop faciles, soit trop difficiles, comme T4.

2) Etude n°2 : « Perception des contrastes vocaliques »

2.1. Plan expérimental

2.1.1. Hypothèse

Le but de cette expérimentation est de déterminer si la pratique musicale induit une meilleure perception des contrastes vocaliques d'une langue étrangère.

2.1.2. Variable indépendante « musique »

La variable indépendante « musique » est composée de deux modalités : le groupe d'enfants « musicien » G1 et le groupe d'enfants « non musiciens » G2 (voir 1.1.2.).

2.1.3. Variables dépendantes

En nous fondant sur l'analyse des interférences entre les contrastes vocaliques du français et de l'anglais (chapitre V, 3), nous avons réalisé une batterie de trois tests de perception, la batterie « Contrastes vocaliques » (pour une description détaillée, voir Volume II, annexe n°4.2.2. et pour leur écoute, CD2). Nous étudierons donc trois variable dépendantes (le tableau n°30 expose le plan expérimental de l'étude n°2) :

- la variable dépendante « Voyelles NN », correspondant au test T6 ;
- la variable dépendante « Voyelles N/NN », correspondant au test T7 ;
- la variable dépendante « Diphtongues et triptongues », correspondant au test T8.

« Contrastes vocaliques »	T6 « Voyelles NN »	T7 « Voyelles N/NN »	T8 « Diphtongues et triptongues »
Groupe G1			
Groupe G2			

Tableau n°30 : plan de l'étude n°2

(G1 pour « groupe musiciens » et G2 pour « groupe non musiciens » ; N pour natif et NN pour non natif).

2.2. Passation

Les tests ont été administrés aux vingt-sept enfants composant notre échantillon dans les mêmes conditions que les tests de la batterie « Prosodie » (se reporter à la partie n°1.2.).

2.3. Résultats généraux

2.3.1. Moyennes générales aux tests : les tests les plus discriminants

« Contrastes vocaliques »	T6 Voyelles NN	T7 Voyelles N/NN	T8 Diphtongues et triphthongues
Moyenne	6,85	6,48	6,44
σ	2,14	2,15	2,45

Tableau n°31 : moyennes obtenues par l'ensemble des enfants aux trois tests de la batterie « Contrastes vocaliques » (N pour natif et NN pour non natif).

Les épreuves de la batterie « Contrastes vocaliques » sont nettement plus difficiles (moyenne générale de 6,59 sur 10) que les épreuves de la batterie « Prosodie » (moyenne générale de 7,60 sur 10). La note moyenne générale des enfants concernant les trois tests « Contrastes vocaliques » est de 19,33 sur 30, alors qu'elle était de 24,25 sur 32 (ramenée à 22,73 sur 30 dans un but de comparaison) pour les cinq tests « Prosodie ».

L'épreuve T8 portant sur la discrimination des diphtongues et des triphthongues par rapport aux productions déviantes des francophones est la plus difficile des trois avec la moyenne la plus faible ($m=6,44$) et les notes les plus dispersées ($\sigma=2,45$), non seulement dans cette batterie, mais dans la première également (voir tableau n°31). Dans l'ordre des difficultés, viennent ensuite T7 ($m=6,48$), portant sur la discrimination des voyelles de l'anglais par rapport aux voyelles du français de timbre correspondant et T6 ($m=6,85$), portant sur la discrimination des voyelles de l'anglais. Les tests T6, T7 et T8 sont assez difficiles et peuvent se montrer discriminants.

2.3.2. Est-ce qu'il y a transfert de compétence entre la pratique musicale et les résultats obtenus aux tests concernant les contrastes vocaliques ? Pertinence du classement

« Contrastes vocaliques »	T6 Voyelles NN	T7 Voyelles N/NN	T8 Diphtongues et triphthongues
Groupe G1	7,90	8,00	7,50
Groupe G2	6,23	5,58	5,82
F	4,28	10,88	3,18
p	0,0491	0,0029	0,0865
σ	2,14	2,15	2,45

Tableau n°32 : moyennes et signifiante des trois tests « Contrastes vocaliques » en fonction du classement retenu pour les tests de musique (G1 pour « groupe musiciens » et G2 pour « groupe non musiciens » ; N pour natif et NN pour non natif).

Les trois tests T6, T7 et T8 se sont révélés difficiles pour les deux groupes. Mais, est-ce qu'ils permettent de faire la différence entre G1 et G2 ? L'analyse de variance révèle que oui (tableau n°32). On constate un effet de groupe significatif pour T6 ($F(1,25)=4,28$; $p=0,0491$) et T7 ($F(1,25)=10,88$; $p=0,0029$). Quant à T8, l'effet de groupe approche de la signifiante ($F(1,25)=3,18$; $p=0,0865$). Les deux premiers tests peuvent donc être considérés comme discriminants et le troisième approche la signifiante. Si on fait une analyse de variance sur les trois tests, on relève un effet de groupe très important ($F(1,25)=9,57$; $p=0,0048$).

Les voyelles qui ont été utilisées dans le test T6 n'existent pas en français, elles sont toutes spécifiques à l'anglais et certaines ont des timbres très proches. Le fait d'avoir une oreille entraînée par la pratique musicale paraît aider à mieux discriminer les différences entre les contrastes vocaliques d'une langue étrangère. Le test le plus discriminant (T7) a été réalisé à partir d'une expertise des interférences qui pouvaient intervenir entre les voyelles du français et celles de l'anglais, en fonction de la proximité de leurs timbres (voir chapitre V, 2. et Volume II, annexe 3.5.2.). Le fait d'avoir une oreille entraînée semble combattre la « surdit  phonologique d j   voqu e au chapitre V (1.2.) et am liorer la perception des sons non natifs. Quant au test T8, il est le plus difficile des trois. Il est moins significatif que les deux autres car il a  t  moins bien r ussi par les musiciens. Mais la moyenne de G1 est quand m me nettement plus  lev e ($m=7,50$) que celle de G2 ($m=5,82$). Une oreille entra n e permet donc de mieux faire la diff rence entre ces deux types de sons.

2.3.3. Corrélations entre les tests de musique et les tests d'anglais de la batterie « Contrastes vocaliques »

« Contrastes vocaliques »	T6 Voyelles NN	T7 Voyelles N/NN	T8 Diphtongues et triptongues
DH	0,084	0,14	0,59
DM	0,38	0,36	0,38
DR	0,33	0,056	0,21
IH	0,29	0,46	0,34
IM	0,00	0,01	0,15
IR	0,32	0,35	0,16

Tableau n°33 : coefficients de corrélation des tests d'anglais avec les tests de musique (N=27 ; coefficient de liberté de 25 ; quand $r \geq 0,38$; $p = .05$, cellules grisées ; DH pour discrimination harmonique, DM pour discrimination mélodique, DR pour discrimination rythmique, IH pour identification harmonique, IM pour identification mélodique, IR pour identification rythmique, N pour natif et NN pour non natif).

Le test IH s'était montré le plus discriminant de la batterie de Zenatti pour différencier les oreilles très musiciennes des autres. Il permettait de tracer une frontière nette entre les TBO et BO d'un côté et les OM, OP, MO et TMO de l'autre. Or, on observe une corrélation entre la note obtenue à ce test et la note globale obtenue aux trois tests de la batterie « Contrastes vocaliques » ($r=0,45$), ce qui n'est pas le cas pour la note globale obtenue aux cinq tests de la batterie « Prosodie » où l'on n'observe aucune autre corrélation avec les autres tests de musique. De manière plus fine, une corrélation apparaît entre IH et T7 « voyelles natives et non natives » ($r=0,46$, voir zone grisée dans le tableau n°33).

Le test DH confirme le rôle du domaine harmonique pour différencier une oreille musicienne d'une oreille non musicienne. Il permet de faire la différence entre TBO, BO, OM et OP d'un côté et MO et TMO de l'autre, mais il est moins discriminant que le test IH. Or, on observe une forte corrélation entre la note obtenue à ce test et T8 « Diphtongues et triptongues » ($r=0,59$, voir zone grisée dans le tableau n°33).

On remarque également de fortes corrélations concernant le test DM. C'était l'épreuve la plus difficile parmi les trois épreuves de discrimination musicale. Elle différenciait nettement les bonnes oreilles musiciennes (TBO, BO et OM) des autres (OP, MO et TMO). On observe une corrélation de DM avec T6 « voyelles non natives » ($r=0,38$, voir zone grisée dans le tableau n°33) et avec T8 « Diphtongues et triptongues » ($r=0,38$, voir zone grisée dans le tableau n°33). Quant à T7 « Voyelles natives et non natives », il y a presque corrélation ($r=0,36$).

Les résultats concernant les tests musicaux avaient montré que, si on devait choisir trois tests pour différencier oreilles musiciennes des non musiciennes, on choisirait dans l'ordre IH, DM, puis DH. Or, on observe des corrélations avec certains tests d'anglais seulement pour ces trois tests. Les domaines harmonique et mélodique différencient le plus les oreilles musiciennes des oreilles non musiciennes. Et, c'est dans ces domaines qu'on observe des corrélations avec certains tests d'anglais, plus particulièrement les tests « contrastes vocaliques ». On relève donc une influence de l'oreille musicienne sur le niveau segmental. Est-ce que ces résultats révèlent une perception plus analytique chez les musiciens ? Cette question sera traitée en détail dans la discussion venant clôturer ce chapitre (3.).

2.3.4. Conclusion de l'étude n°2 « Contrastes vocaliques »

Les tests concernant la prosodie ne se sont pas révélés discriminants pour faire la différence entre les oreilles musiciennes (G1) et les oreilles non musiciennes (G2) (cf. partie 1.5.). Rappelons que ce classement est fondé sur la note générale obtenue aux six tests musicaux et en effet, on ne constate aucune corrélation entre cette note et la note générale obtenue aux cinq tests concernant la « prosodie » ($r=0,17$). En revanche, on constate une corrélation de cette note avec la note générale obtenue aux tests de la batterie « Contrastes vocaliques » ($r=0,42$) et un effet de groupe significatif pour chacun des trois tests. L'hypothèse particulière est donc vérifiée. La pratique musicale induit une meilleure perception des contrastes vocaliques d'une langue étrangère. Est-ce que ces capacités de discrimination plus grandes pour les musiciens au niveau local vont se refléter dans la production des voyelles et des diphtongues ? Ce point sera traité dans le chapitre VIII (3.).

2.4. Résultats détaillés des tests de perception

2.4.1. Résultats détaillés par items

Puisque la différence entre les musiciens (G1) et les non musiciens (G2) est significative pour les résultats aux tests de perception de la batterie « Contrastes vocaliques », il est intéressant d'approfondir ces résultats pour déterminer quels items sont les mieux discriminés par les enfants musiciens.

2.4.1.1. T6 « Différenciation phonétique au sein de l'anglais (« Contrastes NN »)

La différence entre G1 et G2 pour ce test est significatif ($F=4,28$; $p=0,0491$). La moyenne de G1 est de 7,90 sur 10 et celle de G2, 6,23 sur 10. L'écart type est de 2,14, le plus faible des trois tests de la batterie. Les trois items les plus difficiles de T6 sont l'item n°5 avec 62,96 % d'erreurs (**ɔ:** → **ʊ**), l'item n°3 avec 51,85 % d'erreurs (**i** → **i:**) et l'item n° 4, avec 37,03 % (**ʊ** → **əʊ**).

Comparaison détaillée entre les musiciens et les non musiciens

Dans deux items, l'item n°5 (le plus difficile) et l'item n°6 (le 7^{ème} plus difficile), la proportion des musiciens (G1) ayant échoué est plus forte que celle des non musiciens (70 % d'échec pour G1 contre 52,94 % pour G2 dans l'item n°5 ; 30 % d'échec pour G1 contre 17,64 % pour G2 dans l'item n°6, voir les résultats dans le tableau n°34).

Tous les autres items sont mieux réussis par les musiciens. Dans certains items, la différence est flagrante en faveur de G1 : c'est le cas dans l'item n°4 (**extrait sonore n°40**) où l'on constate un taux d'erreur de 10 % pour G1 (une seule erreur), contre 52,94 % d'erreurs pour G2. Les musiciens semblent donc moins sensibles à l'assimilation du contraste [ʊ] par les contrastes [əʊ] et [ɔ:]. Dans l'item n°8 (**extrait sonore n°41**), les sujets de G1 ne produisent aucune erreur, contrairement aux sujets de G2 (35,28 % d'erreurs). Les musiciens se montrent donc moins sensibles à l'assimilation du contraste [ɒ] par les contrastes [ɔ:] ou [ʊ]. Dans l'item n°3 (**extrait sonore n°42**), on constate un taux d'erreur de 30 % pour G1 contre un taux de 58,82 % d'erreurs pour G2. Les sujets de G1 sont donc moins sensibles à l'assimilation du contraste [ɪ] par [i:] ou [e]. Enfin, dans l'item n°2 (**extrait sonore n°43**), on constate un taux d'erreur de 10 % pour G1, alors que ce taux monte à 35,29 % pour G2. Les sujets de G1 sont moins sensibles à l'assimilation de [æ] par [ɑ:] ou [ʌ]. Le tableau n°35 dresse un récapitulatif des principales assimilations relevées dans le test T6. On peut également relever que, lorsqu'il y a une erreur chez les musiciens, celle-ci affecte toujours le contraste possédant le pourcentage d'erreur le plus élevé et cela est vrai pour les autres items. Cette erreur se produit toujours en faveur du son le plus difficile à percevoir, le son « crible ». Par ailleurs, les sujets de G1 ne répondent jamais au hasard, comme certains sujets de G2.

Numéro d'item	Nombre d'erreurs (sur 27)	CIBLE	A	B	C	G1 (M)	G2 (NM)
Item n°5	17 (62,96 %)	[sɔ:t]	[sut] 16 (59,25)	[sɔ:t] (0 %)	[saut] 1 (3,7 %)	10 %	17 %
Item n°3	14 (51,85 %)	[sɪt]	[si:t] 9 (33,33 %)	[sɪt] (0 %)	[set] 5 (18,51 %)	70 %	58,82 %
Item n°4	10 (37,03 %)	[sʊt]	[su:t] 4 (14,81 %)	[səʊt] 6 (22,22 %)	[sut] (0 %)	10 %	52,94 %
Item n°7	8 (29,62 %)	[səʊt]	[sɔ:t] 3 (11,11 %)	[sut] 5 (18,51 %)	[səʊt] (0 %)	20 %	35,29 %
Item n°10	8 (29,62 %)	[saɪə]	[saɪ] 3 (11,11 %)	[saɪə] (0 %)	[saɪ] 5 (18,51 %)	20 %	29,41 %
Item n°2	7 (25,92 %)	[saet]	[sæt] 2 (7,40 %)	[saet] (0 %)	[sa:t] 5 (18,51 %)	10 %	35,29 %
Item n°8	7 (25,92 %)	[sɒt]	[sɒt] (0 %)	[sɔ:t] 5 (18,51 %)	[sut] 2 (7,40 %)	0 %	35,18 %
Item n°6	6 (22,22 %)	[sa:t]	[sæt] 5 (18,51 %)	[sa:t] (0 %)	[saet] 1 (3,70 %)	30 %	17,64 %
Item n°1	5 (18,51 %)	[seɪ]	[seɪ] (0 %)	[saɪ] 2 (7,40 %)	[sɔɪ] 3 (11,11 %)	10 %	23,52 %
Item n°9	5 (18,51 %)	[seə]	[seɪə] 3 (11,11 %)	[sɪə] 1 (3,70 %)	[seə] (0 %)	10 %	23,52 %

Tableau n°34 : Pourcentage moyen d'erreurs pour G1 et G2 et résultats détaillés pour chacune des options de différents items de T6 (A pour la première option présentée au cours du test, B, pour la deuxième option et C, pour la troisième option ; dans la colonne de gauche, les items sont classés par ordre décroissant d'erreurs, G1 pour « groupe musiciens » et G2 pour « groupe non musiciens »).

ɔ: → u
ɪ → i:
u → əʊ
æ → ɑ: ou ʌ
ɒ → ɔ:

Tableau n°35 : tableau récapitulatif des principales assimilations mises à jour par T6 (classées par ordre d'importance).

2.5.1.2. T7 : opposition phonétique entre le français et l'anglais (« Contrastes N/NN »)

La différence entre G1 et G2 pour ce test est hautement significative ($F=10,88$; $p=0,0029$). La moyenne de G1 est de 8 sur 10 et celle de G2, 5,58 sur 10. L'écart type est de 2,15, ce qui situe ce test à une place intermédiaire entre T6 ($\sigma=2,14$) et T8 ($\sigma=2,45$).

Les quatre items les plus difficiles de T7 sont l'item n°1 avec 51,85 % d'erreurs (**a → ɑ:**), l'item n°10 avec 48,14 % d'erreurs (**œ → ɜ:**), l'item n°3 avec 44,44 % d'erreurs (**i → i: et ɪ**) et l'item n°7 avec 44,44 % d'erreurs (**ʌ → a**).

Comparaison détaillée des musiciens et des non musiciens pour T7

Dans deux items, l'item n°1 le plus difficile et l'item n°7 (4^{ème} plus difficile), la proportion des musiciens (G1) ayant échoué est approximativement la même que celle des non musiciens (50 % d'échec pour G1 contre 58,82 % pour G2 dans l'item n°1 ; 50 % d'échec pour G1 contre 47,05 % pour G2 dans l'item n°7, voir les résultats dans le tableau n°36). Ces deux items posent problème aux deux groupes et une oreille musicienne n'aide pas à mieux discriminer. Il y a donc assimilation du contraste [a] par [ɑ:] pour l'item n°1 et assimilation du contraste [ʌ] par [a] pour l'item n°7. En revanche, comme dans le test précédent, lorsqu'il y a une erreur chez les musiciens, celle-ci affecte toujours le contraste possédant le pourcentage d'erreur le plus élevé et cela est vrai des autres items, sauf pour l'item n°10 (48,14 % d'erreurs ; 2^{ème} plus difficile) où les réponses sont variées.

Tous les autres items sont mieux réussis par les musiciens. Dans certains items, la différence est flagrante en faveur de G1 : dans l'item n°3 (**extrait sonore n°44**), on constate 10 % d'erreurs pour G1 contre 64,70 % pour G2. Les sujets de G1 sont donc moins sensibles à l'assimilation du contraste [i] par [i:] ou [ɪ]. Dans l'item n°4 (**extrait sonore n°45**), on relève 10 % d'erreurs pour G1 contre 58,82 % pour G2. Les sujets de G1 sont moins sensibles à l'assimilation du contraste [e] par [ɛ] ou le contraste français [e]. Dans l'item n°5 (**extrait sonore n°46**), les sujets de G1 ne font pas d'erreurs, contrairement aux sujets de G2 qui en font 47,05 %, se montrant sensibles à l'assimilation du contraste [ɪ] par le contraste [ɛ]. Dans l'item n°8 (**extrait sonore n°47**), les sujets de G1 font 10 % d'erreurs contre 47,05 % pour les sujets de G2 ; ils sont donc moins sensibles à l'assimilation du contraste [ɔ] par [o] ou [ɒ]. Enfin, dans l'item n°2 (**extrait sonore n°48**), on ne constate pas d'erreurs chez les sujets de G1 contre 35,29 % d'erreurs pour G2. Les sujets de G2 se montrent donc sensibles à l'assimilation du contraste [o] par [ʌ]. Le tableau n° 37 dresse un récapitulatif des principales assimilations mises à jour par le test T7.

Numéro d'item	Nombre d'erreurs (sur 27)	CIBLE	A	B	C	G1 (M)	G2 (NM)
Item n°1	14 (51,85 %)	[sat]	[saet] 1 (3,70 %)	[sɑ:t] 13 (48,14 %)	[sat] 1 (3,7 %)	50 %	58,82 %
Item n°10	13 (48,14 %)	[sœt]	[sɑt] 4 (14,81 %)	[sɜ:t] 9 (33,33 %)	[sœt] (0 %)	40 %	52,94 %
Item n°3	12 (44,44 %)	[sit]	[si:t] 6 (22,22 %)	[sɪt] 6 (22,22 %)	[sit] (0 %)	10 %	64,70 %
Item n°7	12 (44,44 %)	[sɑt]	[sɔt] 2 (7,40 %)	[sɑt] (0 %)	[sat] 10 (37,03 %)	50 %	47,05 %
Item n°4	11 (40,74 %)	[set]	[sɛt] 7 (25,92 %)	[set] (0 %)	[set] (Fs) 4 (14,81 %)	10 %	58,82 %
Item n°6	10 (37,03 %)	[sat]	[saet] 2 (7,40 %)	[sɑt] 8 (29,62 %)	[sat] (0 %)	20 %	47,05 %
Item n°8	10 (37,03 %)	[sɔt]	[sɔt] 4 (14,81 %)	[sɔt] 6 (22,22 %)	[sɔt] (0 %)	10 %	47,05 %
Item n°5	7 (25,92 %)	[sɪt]	[sɛt] 5 (18,51 %)	[sɪt] (0 %)	[set] 2 (7,40 %)	0 %	41,17 %
Item n°2	5 (18,51 %)	[sɔt]	[sɔt] 2 (7,40 %)	[sɑt] 3 (11,11 %)	[sɔt] (0 %)	0 %	35,29 %
Item n°9	4 (14,81 %)	[sɔ:t]	[sɔ:t] (0 %)	[sɔt] (0 %)	[sɔt] 4 (14,81 %)	10 %	23,52 %

Tableau n°36 : Pourcentage moyen d'erreur pour G1 et G2 et résultats détaillés pour chacune des options des différents items de T7 (A pour la première option présentée au cours du test, B, pour la deuxième option et C, pour la troisième option ; dans la colonne de gauche, les items sont classés par ordre décroissant d'erreur G1 pour « groupe musiciens » et G2 pour « groupe non musiciens »).

a → ɑ:
œ → ɜ:
i → i: et ɪ
ʌ → a → ʌ
e → ɛ
ɔ: → o

Tableau n°37 : tableau récapitulatif des principales assimilations mises à jour par T7 (classées par ordre d'importance).

2.5.1.3. T8 : opposition phonétique entre le français et l'anglais (« Diphtongues, triphthongues »)

La différence entre G1 et G2 pour ce test est presque significative ($F=3,18$; $p=0,0865$). La moyenne de G1 est de 7,50 sur 10 et celle de G2, 5,82 sur 10. Mais l'écart type est trop important (le plus grand pour les trois tests « Contrastes Vocaliques »). Cette tâche est la plus difficile de la batterie, car la moyenne des notes obtenues par les enfants est la plus faible ($m=6,44$).

Comparaison détaillée des musiciens et des non musiciens pour T8

Dans deux items, l'item n°10 (le deuxième plus difficile) et l'item n°2 (le plus facile), la proportion des musiciens (G1) ayant échoué est plus importante que celle des non musiciens (70 % d'échec pour G1 contre 35,29 % pour G2 dans l'item n°10 ; 20 % d'échec pour G1 contre 11,76 % pour G2 dans l'item n°2, voir les résultats exposés dans le tableau n°38). D'habitude, les erreurs des musiciens se produisent toujours sur le contraste ayant le plus gros pourcentage d'erreurs en général (cf. T6 et T7), mais ce n'est pas le cas dans ces deux items, où les réponses sont partagées entre les deux solutions non correctes.

Tous les autres items sont mieux réussis par les sujets de G1 que par ceux de G2. Dans certains items, la différence est flagrante en faveur de G1 : dans l'item n°5 (**extrait sonore n°49**), on constate 30 % d'erreurs pour G1 contre 76,47 % pour G2. Les sujets de G1 sont donc moins sensibles à l'assimilation du contraste [ʊə] par [ɔ:]. Dans l'item n°4 (**extrait sonore n°50**), on relève seulement 10 % d'erreurs pour G1 contre 64,70 % d'erreurs pour G2. Les sujets de G1 sont moins sensibles à l'assimilation du contraste [eə] par [e:]. Dans l'item n°7 (**extrait sonore n°51**), on relève 20 % d'erreurs pour G1 contre 47,05 % d'erreurs pour G2. Les sujets de G1 sont moins sensibles à l'assimilation du contraste [eɪə] par [ejə]. Dans l'item n°1 (**extrait sonore n°52**), on relève 20 % d'erreurs pour G1 contre 41,17 % d'erreurs pour G2. Les sujets de G1 sont moins sensibles à l'assimilation du contraste [ɔɪ] par [ɔj]. Enfin, on constate un taux d'erreur moins élevé pour G1 (10 %) que pour G2 (41,17 %) dans l'item n°6 (**extrait sonore n°53**). Les sujets de G1 sont donc moins sensibles à l'assimilation du contraste [aɪə] par [ajə]. Le tableau n°39 dresse un récapitulatif des principales assimilations relevées dans le test T8.

Numéro d'item	Nombre d'erreurs (sur 27)	CIBLE	A	B	C	G1 (M)	G2 (NM)
Item n°5	17 (62,96 %)	[suə]	[sɔ:] 13 (48,14 %)	[suə] (0 %)	[sɔ:] 4 (14,81 %)	30 %	76,47 %
Item n°10	13 (48,14 %)	[seuə]	[sœuə] 5 (18,51 %)	[seuə] (0 %)	[seuə] 8 (29,62 %)	70 %	35,29 %
Item n°4	12 (44,44 %)	[seə]	[se:] 2 (7,40 %)	[se:] 10(37,03 %)	[seə] (0 %)	10 %	64,70 %
Item n°7	10 (37,03 %)	[seɪə]	[seɪə] (0 %)	[sejə] 6 (22,22 %)	[sejə] 4 (14,81 %)	20 %	47,05 %
Item n°3	10 (37,03 %)	[səʊt]	[sɔ:t] 4 (14,81 %)	[sɒt] 6 (22,22 %)	[səʊt] (0 %)	30 %	41,17 %
Item n°9	10 (37,03 %)	[sauə]	[sauə] 4 (14,81 %)	[sauə] 6 (22,22 %)	[sauə] (0 %)	30 %	41,17 %
Item n°1	9 (33,33 %)	[sɔɪ]	[sɔɪ] (0 %)	[sɔj] 5 (18,51 %)	[sɔj] 4 (14,81 %)	20 %	41,17 %
Item n°6	8 (29,62 %)	[saɪə]	[sajə] 5 (18,51 %)	[saɪə] (0 %)	[sajə] 3 (11,11 %)	10 %	41,17 %
Item n°8	5 (18,51 %)	[sɔɪə]	[sɔjə] 3 (11,11 %)	[sɔjə] 2 (7,40 %)	[sɔɪə] (0 %)	10 %	23,52 %
Item n°2	4 (14,81 %)	[saut]	[saot] 2 (7,40 %)	[saut] (0 %)	[saut] 2 (7,40 %)	20 %	11,76 %

Tableau n°38 : Pourcentage moyen d'erreur pour G1 et G2 et résultats détaillés pour chacune des options des différents items de T8 (A pour la première option présentée au cours du test, B, pour la deuxième option et C, pour la troisième option ; dans la colonne de gauche, les items sont classés par ordre décroissant d'erreurs ; G1 pour « groupe musiciens » et G2 pour « groupe non musiciens »).

uə → ɔ:
əuə → œuə
ee → e:
eɪə → ejə
əʊ → ø
ɔɪ → ɔj
aɪə → ajə
ɔɪə → ɔjə

Tableau n°39 : tableau récapitulatif des principales assimilations mises à jour par T8 (classées par ordre d'importance).

2.4.1.4. Contrastes vocaliques les mieux perçus par les musiciens

En fonction de ces résultats, nous pouvons maintenant dresser une liste des contrastes vocaliques que les enfants musiciens discriminent plus facilement que les enfant non musiciens (tableau n°40).

Voyelles non natives	
	ɔ: → ʊ
	ɪ → i: ou e
	ʊ → əʊ
	ɒ → ɔ:
	ae → a: ou ʌ
Voyelles natives	
	i → i: ou ɪ
	e → ɛ et e
	ɪ → ɛ
	ɔ → ɒ ou ɔ
Diphthongues et triphthongues	
	eə → e:
	eɪə → eɪə
	ɔɪ → ɔj
	aɪə → aɪə

Tableau n°40 : tableau récapitulatif des principales assimilations mises à jour par T8
(en gris, contrastes vocaliques français ou prononcés à la française pour les diphthongues et les triphthongues).

Dans cette liste, on relève des contrastes vocaliques de timbre très proches comme les deux voyelles françaises [i] et [e] et la voyelle anglaise [e] (elles se transcrivent par le même symbole phonétique, mais n'ont pas tout à fait le même timbre), les voyelles anglaises [ae], [a:] et [ʌ]. Quant aux différences entre [eə] et [e:], entre [eɪə] et [eɪə], entre [ɔɪ] et [ɔj] et entre [aɪə] et [aɪə], elles sont extrêmement fines. Les oreilles musiciennes sont donc capables de distinguer des sons de timbre extrêmement proches et lorsqu'elles font des erreurs, c'est très souvent sur l'option dont le pourcentage d'erreurs est le plus élevé. En revanche, comme les oreilles non musiciennes, elles échouent à distinguer la voyelle [ɔ:] de la voyelle [ʊ], la voyelle française [a] de la voyelle anglaise [a:], la voyelle anglaise [ʌ] des voyelles françaises [ɔ] ou [a] et enfin, la diphthongue [ʊə] de la voyelle [ɔ:]. Enfin, il convient d'ajouter que dans certains items, la courbe de Fo est légèrement différente d'une option à l'autre, bien qu'elle ait été modifiée par resynthèse (c'est le cas pour certains items dans T8, comme les items n°7 (**extrait sonore n°51**), n°1 (**extrait sonore n°52**) et n°6 (**extrait sonore n°53**) ; cette différence a pu aider les oreilles musiciennes à identifier l'option correspondant au modèle.

Mais le fait même qu'elles aient été capables de repérer cette différence prouvent qu'elles ont des capacités de discrimination supérieures aux oreilles non musiciennes (même si dans de tels cas, on ne peut pas déterminer si elles ont repéré la différence de contraste ou la différence de Fo).

2.4.2. Résultats détaillés par enfant

2.4.2.1. Très bonne réussite chez les enfants appartenant à la catégorie TBO

Parmi les six meilleurs sujets aux tests d'anglais, cinq appartiennent à la catégorie des TBO pour les tests de musique. Le meilleur sujet de l'échantillon, Nicolas, qui avait obtenu la meilleure note dans les tests de musique (102 sur 102) a également obtenu la meilleure note dans les tests d'anglais (58 sur 62). Nicolas a obtenu un total de 29 sur 30 pour les trois tests concernant les contrastes vocaliques, T6, T7 et T8, alors qu'ils figurent parmi les quatre tests les plus difficiles de la batterie. Ses résultats sont parfaits pour les tests T6 et T7. L'erreur qu'il a commise dans le test T8 concerne un des items les plus difficiles (item n°10) où la majorité des enfants a échoué (48,14 % d'erreurs). Dans le test T4, le plus difficile de la batterie, il en va de même. Il a échoué à l'item le plus difficile du test (77,77 % d'erreurs). Néanmoins, il a obtenu la meilleure note aux cinq épreuves « prosodie » (29 sur 32), avec Armand et Hugo, tous deux classés également en TBO. Les résultats de Nicolas sont donc assez impressionnants. Il est en effet très difficile de ne pas faire d'erreur à cause des interférences entre la LM et la langue cible (crible phonologique).

Agathe a également obtenu une excellente note aux tests concernant les contrastes vocaliques (28 sur 30) et, tout comme Nicolas, elle a été classée dans la catégorie TBO en raison de son excellente note aux tests de musique (97 sur 102). Ses erreurs concernent le test T6 (item n°5, le plus difficile du test avec 62,95 % d'erreurs) et le test T7 (item n°10, le second plus difficile du test avec 48,14 % d'erreurs). Agathe est la seule enfant de notre population à réussir parfaitement le test T8, le plus difficile de la batterie. En revanche, elle est moins bonne concernant les cinq tests « prosodie » (23 sur 32).

Armand, Hugo et Lucie ont également obtenu une très bonne note concernant les tests « prosodie » (respectivement 29, 29 et 28 sur 32) et une bonne note concernant les tests « contrastes vocaliques », mais plus faible que celles de Nicolas et Agathe (respectivement 23, 23 et 25 sur 30).

Les deux autres enfants appartenant à la catégorie TBO sont Violette et Lucas. Ce dernier se classe comme le 13^{ème} de l'échantillon (46 sur 62), avec des résultats assez moyens. Il est à noter que Lucas est le seul enfant de la catégorie TBO qui ne soit pas musicien (et qui ne pratique aucun instrument), bien que provenant d'un milieu très riche musicalement, or nous avons vu dans le chapitre n°X que la pratique musicale et plus spécifiquement la pratique instrumentale jouaient un rôle important dans la formation de l'oreille. Le cas de Violette est déroutant. Avec une note de 36 sur 62, elle se classe comme la 22^{ème} sur une population de 27 enfants. C'est la seule TBO qui obtienne des résultats aussi mauvais en anglais. Les tests de musique s'étaient pourtant très bien passés, mais pendant la passation des tests d'anglais, Violette s'est montrée plus intéressée par l'expérimentation et l'expérimentatrice que par les tests (cf. chapitre VI, 1.3.2.).

2.4.2.2. Grande proportion d'enfants appartenant au groupe G1 parmi les meilleurs

Dans la catégorie des BO, on trouve Mélissa en 9^{ème} position (50 sur 62), qui a obtenu une très bonne note aux tests « local » (26 sur 30), Matthias en 11^{ème} position (48 sur 62), et Emeline en 18^{ème} position (41 sur 62) ; ses résultats sont moins bons, que ce soit en global (20 sur 32) ou en local (21 sur 30). Parmi les quatorze enfants qui ont obtenu une note égale ou supérieure à 46 sur 62, on trouve six enfants appartenant à la catégorie TBO, deux enfants appartenant à la catégorie BO, trois enfants à la catégorie OM, deux à la catégorie OP et un à la catégorie MO, soit une proportion de 57,14 % (huit enfants sur quatorze appartenant au groupe G1).

2.4.2.3. Grande proportion d'enfants appartenant au groupe 2 parmi les plus mauvais

Parmi les neuf enfants ayant obtenu les plus mauvaises notes en anglais (inférieures à 38 sur 62), aucun n'appartient à la catégorie BO. En revanche, nous trouvons un TBO, Violette, dont nous avons déjà parlé plus haut (voir cas particuliers). La proportion d'enfants appartenant au groupe 2 est donc de neuf enfants sur dix, soit 90 %. Il semble que les résultats soient très tranchés en faveur de la catégorie TBO, ce qui explique l'effet de groupe relevé plus haut pour le groupe 1, composé en majorité d'enfants appartenant à cette catégorie TBO (sept sur neuf).

Un haut niveau musical correspondant à une note élevée aux tests de musique semble donc avoir une forte répercussion sur les résultats aux tests de discrimination en anglais. Lorsque la note est plus faible en musique, il n'y a plus de différence marquée entre les musiciens et les non musiciens. La formation musicale a donc une influence sur la perception de l'anglais, uniquement pour les enfants qui ont une très bonne oreille.

2.4.2.4. Résultats aux tests « Prosodie » en fonction des résultats aux tests « Contrastes vocaliques » : interaction du global avec le local

Jamais de déséquilibre aux dépens du global

Les épreuves concernant la prosodie ont été mieux réussies par l'ensemble de notre échantillon. D'ailleurs, on ne constate jamais de déséquilibre important aux dépens du global. Un enfant peut avoir un mauvais résultat en local, mais rarement l'inverse. On constate un déséquilibre pour Kevin (28 sur 32 en global et 15 sur 30 en local), Medhi (25 sur 32 en global et 12 sur 30 en local), Laura (26 sur 32 en global et 10 sur 30 en local), Christopher (21 sur 32 en global et 14 sur 30 en local), Marianne (28 sur 32 en global et 18 sur 30 en local), Philippine (21 sur 32 en global et 15 sur 30 en local). Ces six enfants ont en commun d'avoir échoué au test d'identification harmonique IH (en obtenant respectivement la note de 0, 4, 4, 5, 0, 2 sur 20) et de faire partie des onze à avoir obtenu les plus mauvais résultats en local. Pour Laura, c'est la seule épreuve en musique à laquelle elle a complètement échoué.

Importance de IH dans les résultats en « local »

Les résultats sont mauvais, à la fois en global et local, pour Maude (12 sur 32 en global et 11 sur 30 en local), Matthieu (17 sur 32 en global et 13 sur 30 en local) et John (18 sur 32 en global et 19 sur 30 en local). Au cours des tests musicaux, ces trois enfants ont en commun de n'avoir échoué qu'à une seule épreuve, IH (en obtenant respectivement la note de 1, 1 et 4 sur 20). Les quatorze enfants qui ont obtenu les plus mauvais résultats en « local » ont en commun d'avoir échoué au test de musique IH et pour cinq d'entre eux, c'est la seule épreuve musicale qu'ils aient ratée.

Seulement deux exceptions sont à noter : Violette, la seule TBO parmi les quinze enfants à avoir obtenu les plus mauvais résultats en local (14 sur 30) a réussi toutes les épreuves musicales (98 sur 102). Laurène (15 sur 30 en local) a parfaitement réussi l'épreuve IH (20 sur 20), mais elle a échoué à l'épreuve d'identification rythmique IR (1 sur 20).

Or, il n'y a pas de corrélation entre les résultats obtenus en local et à IR ($r=0,32$). Les trois autres enfants à avoir échoué à ce test ont eu des difficultés avec les autres tests musicaux également. Mais le cas de Laurène est intéressant car elle est la seule à n'avoir échoué qu'à l'épreuve IR. Par ailleurs, ses résultats en DR sont moyens. Elle semble donc avoir des difficultés dans le domaine rythmique.

Parmi les 15 derniers sujets en local : Violette (22 et 14), qui avait réussi toutes les épreuves musicales (98/102). Laurène (21 et 15) a échoué à l'épreuve d'identification rythmique (1/20) et a obtenu 20/20 à l'épreuve d'identification harmonique.

Parmi les 12 meilleurs sujets en local, seuls trois enfants ont échoué au test IH : Quentin (1 sur 20), Anaïs (2 sur 20) et Lison (2 sur 20). Tous les enfants appartenant aux catégories TBO et BO l'ont réussi. En musique, nous avons vu que ce test était le plus discriminant de la batterie, car il permet de différencier nettement les très bons musiciens des autres. Le pouvoir discriminant d'IH semble se reporter sur les résultats en anglais et, en effet, on observe une corrélation entre la note obtenue à ce test et la note globale obtenue aux épreuves « Contrastes vocaliques » ($r=0,45$) (voir figure n°82). En revanche, il n'y a pas de corrélation entre IH et les épreuves « Prosodie » ($r=0,16$).

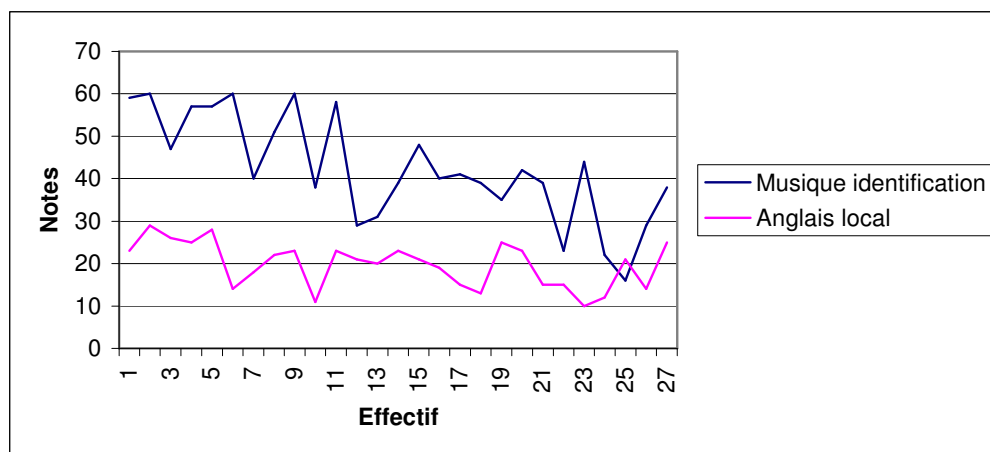


Figure n°82 : évolution simultanée de la note obtenue en IH et de la note globale obtenue aux épreuves « Contrastes vocaliques » en anglais ($r=0,42$).

2.5. Test de l'effet des autres classements sur les résultats aux tests de perception en anglais

2.5.1. Catégories socioprofessionnelles

Le facteur socio-économique joue un rôle important dans la réussite scolaire, mais nous avons vu qu'il ne se montrait heureusement pas déterminant. Les résultats obtenus aux tests de musique ne dépendent pas de la CSP, mais d'autres facteurs plus importants, tels que la musicalité du milieu ou, le plus important, la pratique musicale et en particulier instrumentale de l'enfant. Nous avons relevé un transfert de compétence de la musique à la perception des contrastes vocaliques de l'anglais. Il paraît cependant intéressant de savoir si les résultats aux tests d'anglais peuvent être déterminés par la CSP. Nous avons repris le classement CSP adopté pour les tests musicaux (pour le détail de ce classement, se reporter au chapitre VI, 2.3.1.). Il divise l'échantillon en deux groupes, « CSP1 » (n=8) et « CSP2 » (n=15). CSP1 regroupe les enfants dont les parents appartiennent aux catégories 3 et 4 définies par l'INSEE (cadres supérieurs et professions intermédiaires) et CSP2 regroupe les enfants dont les parents appartiennent aux catégories 2, 5, 6 et 8 (agriculteurs, artisans, employés et ouvriers).

	T1 Schémas intonatifs	T2 Place du pivot	T3 Régularité accentuelle	T4 Tons mélodiques	T5 Accent de mots
CSP1	8,33	9,10	7,91	5,53	6,66
CSP2	7,77	7,90	7,66	5,99	8,10
F	0,42	1,70	0,07	0,28	1,65
p	0,5252	0,2063	0,7895	0,6033	0,2132

*Tableau n°41 : moyennes et signifiante aux tests « prosodie »
en fonction du classement selon la CSP (CSP1 pour le groupe des catégories sociales « supérieures et intermédiaires » et CSP2 pour le groupe des catégories sociales « agriculteurs, artisans, employés, ouvriers »).*

Une analyse de variance (anova) ne révèle aucun effet de groupe sur la note générale obtenue aux trois tests de la batterie « Prosodie » ($F(1,21)=0,00$; $p=0,9739$), ni sur la note obtenus à chacun des cinq tests (voir tableau n°41).

	T6 Voyelles NN	T7 Voyelles N/NN	T8 Diphtongues et triphthongues
CSP1	6,75	6,50	7,00
CSP2	6,46	6,13	6,06
F	0,09	0,16	0,89
p	0,7711	0,6927	0,3555

Tableau n°42 : moyennes et signifiante aux tests « contrastes vocaliques » en fonction du classement selon la CSP (CSP1 pour le groupe des catégories sociales « supérieures et intermédiaires » et CSP2 pour le groupe des catégories sociales « agriculteurs, artisans, employés, ouvriers », N pour natif et NN pour non natif).

De la même manière, le classement selon la CSP ne se montre pas pertinent en ce qui concerne la batterie « Contrastes vocaliques ». Il n'y a pas d'effet de groupe sur la note générale obtenue aux trois test ($F(1,21)=0,53$; $p=0,4757$), ni sur la note obtenue pour chaque test (voir tableau n°42).

Le classement par CSP n'est donc pas significatif. Les résultats aux tests d'anglais ne dépendent pas du statut socio-économique du foyer dont provient l'enfant.

2.5.2. Milieu familial « musicien »

La richesse du milieu familial est un facteur plus important que la CSP dans la formation des aptitudes musicales de l'enfant. Le classement adopté pour les tests musicaux ne s'est pas révélé significatif pour les trois tests les plus discriminants, c'est-à-dire IH, DH et DM, mais significatif pour l'ensemble des six tests de la batterie. Par ailleurs, on remarque que les enfants qui ont obtenu les notes les plus élevées proviennent de milieux musiciens. Les aptitudes musicales de l'enfant sont donc influencées par la richesse musicale du milieu. Mais, est-ce que cette richesse va également influencer les aptitudes concernant la perception de l'anglais ? Pour le savoir, nous avons divisé les enfants de notre échantillon en deux groupes, le groupe des enfants provenant d'un milieu musicien « M1 » ($n=10$) et le groupe des enfants provenant d'un milieu non musicien « M2 » ($n=10$) et avons testé la signifiante de ce classement sur les résultats obtenus aux deux batteries de tests d'anglais (pour le détail de ce classement, se reporter au chapitre VI, 2.3.2.).

	T1 Schémas intonatifs	T2 Place du pivot	T3 Régularité accentuelle	T4 Tons mélodiques	T5 Accent de mots
M1	8,16	7,71	7,83	6,14	7,00
M2	8,23	8,06	7,84	5,71	8,33
F	0,01	0,63	0,00	0,30	1,81
p	0,9298	0,4332	0,9901	0,5909	0,1907

*Tableau n°43 : moyennes et signifiante aux tests « prosodie »
en fonction du classement selon le milieu familial
(M1 pour les milieux « musiciens » et M2 pour les milieux non « musiciens »).*

L'analyse de variance (anova) ne révèle aucun effet de groupe sur la note générale obtenue aux cinq tests de la batterie « Prosodie » ($F(1,25)=0,02$; $p=0,9017$), ni sur la note obtenue à chacun de ces tests (voir tableau n°43).

	T6 Voyelles NN	T7 Voyelles N/NN	T8 Diphtongues et triphthongues
M1	7,70	6,52	7,10
M2	6,35	6,47	6,05
F	2,65	0,00	1,14
p	0,1164	0,9735	0,2961

*Tableau n°44 : moyennes et signifiante aux tests « contrastes vocaliques »
en fonction du classement selon le milieu familial (M1 pour les milieux « musiciens » et M2 pour les milieux
« non musiciens » ; N pour natif et NN pour non natif).*

On ne constate pas non plus d'effet de groupe sur la note générale obtenus aux trois tests de la batterie « Contrastes vocaliques » ($F(1,25)=1,29$; $p=0,2677$), ni sur la note obtenue à chacun de ces tests (voir tableau n°44).

Ce classement n'est donc pas significatif et la richesse musicale du milieu familial n'influence pas les résultats obtenus aux tests de perception en anglais. La CSP ne paraît pas influencer les résultats obtenus aux tests de musique, mais on pourrait penser que les enfants issus de milieux musicalement favorisés proviennent de milieux favorisés, au moins sur le plan culturel. Dans ce sens, ce n'est peut être pas un contact accru avec la musique qui joue un rôle, mais un contexte culturel privilégié. Si cette hypothèse est vraie, l'influence du milieu (précisons bien, en tant que milieu culturellement favorisé) devrait se reporter sur les résultats aux tests d'anglais. Mais ce n'est pas le cas. C'est donc bien un bain musical plus important qui paraît influencer les capacités musicales des enfants.

2.5.3. Pratique instrumentale

La pratique instrumentale s'est révélée être le facteur déterminant dans les résultats aux tests de musique, notamment aux trois tests les plus discriminants de la batterie. La qualité de l'oreille dépend non pas de l'appartenance à une classe musicale, mais du fait qu'un enfant pratique un instrument ou pas, et avec quelle fréquence. Par ailleurs, nous avons relevé un transfert de compétence de la musique à la perception des contrastes vocaliques de l'anglais. Si la pratique instrumentale influence autant le mode de perception de l'oreille, un classement en fonction de la pratique instrumentale devrait se montrer pertinent s'il est appliqué aux résultats aux tests d'anglais. Nous avons repris le classement utilisé pour les tests musicaux (pour le détail, se reporter au chapitre VI, 2.3.3.) ; celui-ci divise notre échantillon en deux groupes, les instrumentistes « I » (n=9) et les non instrumentistes « NI » (n=18).

	T1 Schémas intonatifs	T2 Place du pivot	T3 Régularité accentuelle	T4 Tons mélodiques	T5 Accent de mots
I	8,51	8,88	8,33	6,03	6,48
NI	8,05	8,01	7,59	5,79	8,51
F	0,35	1,12	0,86	0,09	4,42
p	0,5614	0,2994	0,3633	0,7720	0,0457

Tableau n°45 : moyennes et signifiante aux tests « prosodie » en fonction du classement selon la pratique instrumentale (I pour le groupe des « instrumentistes » et NI pour le groupe des « non instrumentistes »).

Une analyse de variance ne révèle aucun effet de groupe sur la note générale obtenue aux cinq tests de la batterie « Prosodie » ($F(1,25)=0,01$; $p=0,9208$). En ce qui concerne la note obtenue dans chacun des tests (voir tableau n°45), on relève un effet de groupe sur T5 « accents de mots » ($F(1,25)=4,42$; $p=0,0457$), mais de manière étrange, c'est ici NI qui se montre meilleur que I.

	T6 Voyelles NN	T7 Voyelles N/NN	T8 Diphtongues et triphthongues
I	8,00	6,55	7,22
NI	6,27	6,44	6,05
F	4,38	0,02	1,37
p	0,0467	0,9024	0,2520

Tableau n°46 : moyennes et signifiante aux tests « contrastes vocaliques » en fonction du classement selon la pratique instrumentale (I pour le groupe des « instrumentistes » et NI pour le groupe des « non instrumentistes » ; N pour natif et NN pour non natif).

On ne constate pas non plus d'effet de groupe sur la note générale obtenus aux trois tests de la batterie « Contrastes vocaliques » ($F(1,25)=1,93$; $p=0,1768$). En revanche, en ce qui concerne la note obtenue dans chacun de ces tests (voir tableau n°46), on observe un effet de groupe sur le test T6 « voyelles anglaises » ($F(1,25)=4,38$; $p=0,0467$). Ici, c'est le groupe I qui obtient la moyenne la plus élevée ($m=8$) par rapport au groupe NI ($m=6,27$).

2.5.4. Conclusion

Aucun des trois classements testés (CSP, « musicalité » du milieu et pratique instrumentale) ne se montre significatif pour expliquer les résultats généraux obtenus par les enfants de notre échantillon aux tests de la batterie « Prosodie » et à ceux de la batterie « Contrastes vocaliques ». On observe cependant un effet significatif de la pratique instrumentale sur les résultats obtenus au test T6 (contrastes vocaliques natifs/non natifs). Dans le chapitre VI, ce classement s'était montré presque significatif pour les résultats obtenus par les enfants aux tests de musique, ce résultat n'a donc rien d'étonnant. En revanche, il est difficile de comprendre pourquoi on relève un effet significatif de la pratique instrumentale sur les résultats obtenus au test T5 (accent de mot)... en faveur des non instrumentistes. Il est possible que l'accent de mot, très marqué en anglais, soit facilement repérable et que la tâche de discrimination du test T5 ne nécessite donc pas d'aptitudes musicales particulières de la part des enfants (en effet, la moyenne générale obtenue à ce test par la totalité des enfants de l'échantillon est élevée ; $m=4,70$ sur 6). Mis à part ce dernier résultat, ces trois classements se montrent moins appropriés que le classement fondé sur la note obtenue aux tests musicaux pour expliquer les résultats obtenus par les enfants aux tests d'anglais des deux batteries.

3) Conclusion générale et discussion des résultats

Les tests de la batterie « Prosodie » sont dans l'ensemble bien réussis par les enfants de notre échantillon sans distinction significative entre les groupes. Et dans le test le plus difficile de la batterie (T4, « tons mélodiques »), les résultats sont mauvais pour les deux groupes. L'hypothèse posée dans l'étude n°1 ne se vérifie donc pas et les musiciens ne se montrent pas meilleurs que les non musiciens dans la discrimination des contrastes prosodiques.

En revanche, les tests de la batterie « Contrastes vocaliques » font apparaître une différence significative entre les résultats des musiciens et des non musiciens et ce, pour chacun des trois tests. Ces tests sont globalement plus difficiles que ceux de la batterie « Prosodie » et ils sont mieux réussis par les musiciens.

L'hypothèse posée dans l'étude n°2 est donc vérifiée et on relève un transfert de compétence de la formation musicale vers la perception des contrastes vocaliques. Les résultats détaillés par items montrent que des contrastes vocaliques de timbre très proches sont très bien discriminés par les enfants musiciens ; ils sembleraient donc moins sensibles que les enfants non musiciens au phénomène de crible phonologique. L'analyse détaillée des résultats par enfant révèle qu'un niveau musical élevé (appartenance à la catégorie TBO) entraîne une forte réussite aux tests d'anglais. En revanche, si le niveau musical est moins élevé (appartenance à la catégorie BO), la réussite aux tests d'anglais est moins flagrante et les résultats sont peu différenciés entre enfants musiciens et enfants non musiciens.

Par ailleurs, on relève plusieurs corrélations entre les notes obtenues par l'ensemble des enfants aux tests d'anglais et à certains tests musicaux. Ces tests concernent le domaine mélodique (DM) et spécialement le domaine harmonique (IH et DH). Or, nous avons vu dans le chapitre VI (2.2.5.), que ce sont essentiellement les capacités harmoniques et mélodiques qui permettent de différencier une oreille musicienne d'une oreille non musicienne.

Une oreille musicienne se caractérise par un traitement analytique plus poussé qu'une oreille non musicienne

Au niveau harmonique...

Les tests de musique montrent qu'une très bonne oreille musicienne se caractérise par des aptitudes plus élevées dans le domaine harmonique, que ce soit en discrimination (DH) ou en identification (IH). Un musicien apprend à décomposer un accord, c'est-à-dire à en percevoir les différentes composantes. Ce type de perception passe par un traitement analytique. Au contraire, une oreille non entraînée va percevoir l'accord comme une forme globale, non décomposable. Il est d'ailleurs probable que dans une suite d'accords, la partie supérieure soit trop prégnante pour qu'une oreille non musicienne puisse percevoir les autres notes composant les accords. La dimension de la mélodie, temporelle, fait donc « disparaître » la dimension harmonique. Nous avons observé cet effet de « cache » avec l'adulte amusique présenté dans la partie précédente. Dans le test d'identification harmonique (IH), il ne faisait pas la différence entre des suites d'accords consonnants et des suites d'accords dissonnants, car la partie supérieure était identique pour les deux suites. Une oreille musicienne semble donc se caractériser par de meilleures aptitudes au niveau harmonique, dues à un traitement analytique des accords musicaux.

Cela impliquerait un fonctionnement perceptif radicalement différent chez les musiciens qui pourrait se transposer à l'écoute des éléments suscitant une perception de type analytique dans une langue. La discrimination des contrastes vocaliques fait précisément appel à des capacités analytiques. En anglais, les tests ont révélé qu'une oreille musicienne se caractérisait par une meilleure aptitude à les discriminer. Il semblerait donc qu'il existe une similitude de traitement entre les accords musicaux et les contrastes vocaliques, puisque les aptitudes à percevoir les premiers se transfèrent à la perception des seconds.

Et au niveau mélodique...

Une oreille musicienne se caractérise également par des capacités accrues dans le domaine mélodique (épreuve de discrimination, DM). Les adultes et les nourrissons (Trehub et al., 1984) commencent par traiter l'information verbale et musicale par contours de hauteur. Ce traitement de type global semble constituer un important mécanisme d'organisation perceptif qui dirige ensuite sur la segmentation de patterns plus complexes. Dans une tâche de reconnaissance de mélodies peu familières, l'information provient des contours. En revanche, si la mélodie est familière, la stratégie devient plus analytique en se concentrant, non plus sur la configuration globale de hauteur, mais sur la « distance » perceptive entre les différents intervalles composant la mélodie. Si l'auditeur est capable de traiter une mélodie par intervalles, il doit pouvoir la transposer à partir de n'importe quelle note, « *Au Clair de la lune* » à la tierce supérieure par exemple. En résumé, il co-existe chez n'importe quel auditeur deux stratégies, une stratégie « globale » pour traiter des mélodies que l'on ne connaît pas et une stratégie « locale » qui permet de traiter les mélodies déjà connues. Zenatti (1981 : 49) souligne d'ailleurs que

l'« acculturation musicale intervient pour faciliter la perception analytique d'un détail lorsque le contexte musical est familier ».

Il est probable qu'il se passe la même chose pendant l'acquisition de la LM ou l'apprentissage d'une langue étrangère : face à un flux langagier inconnu, l'auditeur se concentre d'abord sur une information de type global (prosodie : contours de hauteur et rythme), pour ensuite adopter une stratégie de plus en plus analytique (voir chapitre II, 1.2.3.). Une fois les informations de type global familières, l'oreille peut augmenter sa rapidité de traitement et se concentrer sur les informations de type analytique.

Avantage de l'information globale sur l'information locale

Dans les tests d'anglais, on n'observe jamais de déséquilibre aux dépens du global. En effet, tous les musiciens qui ont obtenu une bonne note en local ont également obtenu une bonne note en global. La prosodie paraît plus facile à percevoir pour notre population. Est-ce que cela vient des tests eux-mêmes, où la tâche serait trop facile ? Mis à part le test T5, tous les tests des deux batteries utilisent la même procédure de passation. Si cette procédure était mauvaise, les résultats aux tests seraient globalement bons ou globalement mauvais. Or, ce n'est pas le cas, car les résultats sont dans l'ensemble bons pour les cinq tests concernant le global et mauvais pour les trois tests concernant le local. Le test T4 est le plus difficile des deux batteries, peut être parce qu'il n'y a pas de tons mélodiques en français et que les différences affectant les contours mélodiques de l'anglais sont beaucoup plus subtiles que celles affectant les contours intonatifs. En revanche, pour les autres tests de la batterie « prosodie », il est probable que les phénomènes appréhendés soient plus faciles à détecter (différence de schémas intonatifs, différence de pivot, de type de rythmicité et d'accent de mot). Ces phénomènes, en raison de leur saillance perceptive sont peut être aussi accessibles aux oreilles musiciennes qu'aux oreilles non musiciennes. Si tel est le cas, il est normal que nous n'ayons pas relevé de différence significative entre les résultats des musiciens et des non musiciens. Les phénomènes prosodiques utilisés dans les tests seraient tellement prégnants qu'il n'y aurait pas besoin d'avoir une oreille exercée pour les détecter, contrairement aux contrastes vocaliques. La prosodie pourrait donc bien continuer à jouer le rôle de structure d'accueil (chapitre II et chapitre III) pour les éléments plus complexes que sont les contrastes vocaliques. Le fait qu'il n'y ait pas de déséquilibre aux dépens du global va dans le sens de cette hypothèse.

Un musicien entend-il mieux ou entend-il différemment ?

Si nous pouvons maintenant affirmer qu'une oreille musicienne experte se caractérise par de meilleures aptitudes analytiques, à la fois dans le domaine harmonique et mélodique, nous ne pouvons pas déterminer si ces capacités proviennent d'un meilleur entraînement global et d'une attention mieux maîtrisée aux phénomènes sonores ou si elles proviennent d'une réorganisation perceptive. Autrement dit, la question est de savoir si un musicien entend mieux ou s'il entend différemment. Il paraît très difficile d'y répondre car, en matière de perception, nous ne pouvons observer que des comportements. Sur le traitement lui-même, il est difficile d'en savoir plus sans avoir recours aux techniques d'imagerie cérébrale.

Or, certains travaux récents en neurosciences et en sciences cognitives apportent un éclairage singulier aux problématiques « perception musicienne vs perception non musicienne », « perception de la musique vs perception de la langue » et « traitement global vs traitement local ». Bien que non spécialiste, nous avons décidé de ne pas en faire l'économie, car ils apportent des éléments de réponse troublants qui ne peuvent pas être négligés.

L'hémisphère droit est-il un « hémisphère musical » (Tramo, 2001 : 54) ?

Depuis les expériences d'écoute dichotique réalisées par Kimura (1961a et b), l'hémisphère droit est considéré comme l'« hémisphère musical » : l'avantage de l'oreille gauche suggère une prédominance de cet hémisphère dans les tâches de reconnaissance mélodique. Shankweiler (1966a) a obtenu les mêmes résultats avec des sujets cérébro-lésés, bien qu'il souligne que cette prédominance soit moins marquée que pour le langage. Mais, d'autres expériences d'écoute dichotique ont montré un avantage de l'oreille droite, notamment chez les musiciens (Bever et Chiarello, 1974), suggérant la participation de l'hémisphère gauche dans le traitement expert de la musique. Ces conclusions sont confirmées par les résultats obtenus grâce à l'imagerie cérébrale. Lorsque une zone cérébrale s'active pendant la perception d'un son, le débit sanguin s'accélère. La tomographie à émission de positons (désormais T.E.P.) permet de visualiser ces zones activées et de les cartographier. Avec cette technique, Nakamura et al. (1999) ont montré une augmentation du débit sanguin dans les deux régions temporales à droite et à gauche (gyrus temporel supérieur incluant le cortex auditif primaire et secondaire), à l'écoute de la musique (8 droitiers, 3 T.E.P. chacun, à 6 mois d'intervalle). Il n'existe donc pas un centre de la musique localisé dans un hémisphère bien spécifique. Ce sont et bien les deux hémisphères qui sont impliqués dans le traitement de la musique et toutes les structures qui participent à ce traitement contribuent également à d'autres formes de cognition.

Par ailleurs, une asymétrie fonctionnelle du planum temporale à gauche a été observée dans la plupart des descriptions post-mortem (Geschwind et Levitsky, 1968) et cette asymétrie a été interprétée comme une dominance du langage dans l'hémisphère gauche. Pourtant, en 1959, dans une étude post-mortem sur un musicien atteint de surdité mélodique, Kleist avait remarqué une lésion circonscrite dans le planum temporale gauche.

Grâce au T.E.P., Schlaug et al. (1995) ont pu travailler sur 30 sujets musiciens et 30 sujets contrôles « in vivo » (tous droitiers). Ils ont mis en évidence que le planum temporale des musiciens est plus étendu à gauche et que cette latéralisation dépend de la qualité de l'oreille des musiciens : l'asymétrie est plus forte chez les musiciens possédant une oreille absolue, alors qu'elle ne diffère pas chez les musiciens possédant une oreille relative des sujets contrôles. Ces résultats suggèrent une différence de latéralisation du traitement musical chez les musiciens et les non musiciens, avec une représentation latéralisée plus à gauche chez les musiciens. Mais cette asymétrie gauche du cortex ne s'observe que chez des musiciens possédant une oreille absolue.

Réorganisation perceptive liée à la précocité de l'apprentissage

L'expérience musicale pendant l'enfance peut effectivement influencer le développement structural du cortex auditif. Dans une étude T.E.P. concernant 9 étudiants en musique possédant l'oreille absolue, 11 étudiants en musique possédant l'oreille relative et 13 sujets contrôles non musiciens (tous droitiers), Pantev et al. (1998) ont montré une activation cérébrale significativement plus forte chez les musiciens (de 25 %) pour des sons de piano (sons périodiques complexes) que pour des sons purs (sons périodiques simples). Le degré d'activation se corrèle avec l'âge auquel les musiciens ont commencé leur apprentissage de la musique. Si cet apprentissage a commencé avant 9 ans, la réorganisation corticale est plus large concernant la reconnaissance des sons de piano. Ces résultats rejoignent ceux d'une étude précédente (Elbert, Pantev et al., 1996) concernant cette fois-ci les représentations somato-sensorielles des positions de violon chez des enfants âgés de moins de 10 ans. Ces études soulignent l'importance de la précocité dans l'apprentissage de la musique. Avant 10 ans, un entraînement intensif à la musique peut conduire à une véritable réorganisation corticale du cerveau, provoquant une grande différence de traitement chez les musiciens et les non musiciens. Après 10 ans, une telle réorganisation ne semble plus possible. Ces résultats rejoignent l'hypothèse d'un « âge critique » en matière d'apprentissage (Lenneberg, 1967, voir chapitre IV).

Envisager le problème de la latéralisation musicale sous un autre angle

Gates et Bradshaw (1977), Bradshaw et Nettleton (1981) proposent d'envisager le problème de la latéralisation musicale sous un autre angle. La divergence de résultat observée lors des tâches d'écoute dichotique proviendrait du fait que l'hémisphère droit serait impliqué dans le traitement global de l'information auditive, alors que l'hémisphère gauche serait impliqué dans son traitement local.

« L'hémisphère droit se spécialise dans la combinaison des éléments pour créer un ensemble. Il s'occupe de synthétiser. Il recherche et construit des structures et reconnaît les relations entre éléments séparés. L'hémisphère droit ne traite pas l'information linéairement, mais simultanément, en parallèle » (Williams, 1986 : 17).

L'hémisphère droit se caractérise par sa capacité à former une « Gestalt » afin de percevoir une forme dans son ensemble, c'est-à-dire par un traitement holistique ou global de l'information.

« L'hémisphère gauche est parfois décrit comme analytique parce qu'il reconnaît les éléments qui constituent un tout [...]. Le traitement de l'information par l'hémisphère gauche est également linéaire et séquentiel. Il va d'un point à un autre et suit une démarche pas à pas » (Williams, 1986 : 16).

Cet hémisphère se caractérise donc par sa capacité à isoler les détails, c'est-à-dire par un traitement analytique ou local de l'information (voir chapitre II, 1.2.).

Cet axe se montre-t-il fructueux pour expliquer le traitement de la langue et de la musique ?

Dans le domaine de la musique, l'étude T.E.P. de Mazziotta et al. (1982) révèle que les sujets qui estiment utiliser une approche analytique, montrent une activation plus grande de la zone pariéto-temporale à gauche (incluant le planum temporale) dans une tâche de mémoire tonale. La situation inverse est observée pour les individus qui n'utilisent pas de stratégie spécifique. Par ailleurs, Peretz et Morais (1988) ont montré que, si une mélodie est transposée mais que son contour est préservé, elle est traitée dans l'hémisphère droit de l'auditeur. En revanche, si le contour de la mélodie est altéré, elle est traitée dans l'hémisphère gauche, probablement parce que le traitement est différent.

En effet, dans le premier cas, c'est une stratégie de traitement par contour mélodique (de type global) qui est mise en jeu et dans le second cas, c'est une stratégie de traitement par intervalle (de type local) (pour une explication de ce type de stratégie, se reporter au chapitre II, 1.2.3.). Ces résultats ont été vérifiés chez des patients cérébro-lésés (Ayotte et al., 2000 ; Peretz, 1990) et chez des patients épileptiques (Liégeois-Chauvel et al., 1998). D'autres études expérimentales réalisées par Peretz et Morais (1987) et par Peretz et al. (1987) montrent que, lorsque le matériel expérimental est manipulé et que les instructions données aux sujets sont modifiées, il est possible de provoquer soit un traitement de type global, soit un traitement de type local. Par ailleurs, l'avantage de l'oreille droite dans le traitement local des mélodies suggère une prédominance de l'hémisphère gauche pour les tâches de type analytique. Bever et Chiarello (1974) suggèrent que les musiciens utilisent un traitement plus analytique que les non musiciens lors de la perception d'une mélodie non familière. Selon Dowling (1982), le degré d'expertise musicale affecte la manière de percevoir la mélodie et influence la latéralité auditive. Ces travaux semblent donc apporter une réponse à la question que nous nous étions posée plus haut, à savoir qu'un musicien entend différemment d'un non musicien et que la formation musicale favorise une réorganisation au niveau cérébral.

Le traitement des voyelles

Cependant, les résultats exposés concernent uniquement le traitement des contours mélodiques, alors que dans l'étude n°2, nous avons relevé un transfert de compétence de la pratique musicale à la discrimination des contrastes vocaliques. Il existe quelques travaux réalisés en imagerie cérébrale sur la perception des phonèmes, dont l'étude menée par Naatanen et al. (1997), mais celle-ci ne s'intéresse pas à la différence de traitement en fonction du degré d'expertise musicale de l'oreille. Cependant, elle fournit des informations intéressantes sur la perception des voyelles. En effet, les auteurs ont montré que la perception des voyelles implique une activation des zones se trouvant dans le cortex auditif gauche. Leur expérimentation a été menée sur 13 sujets finnois et 11 sujets estoniens, tous droitiers (T.E.P. et E.E.G.). En outre, ils ont relevé dans le cortex auditif gauche une plus grande activation à l'écoute de voyelles déviantes de la LM qu'à l'écoute de voyelles déviantes dans une LE. Selon les auteurs, des traces mémorielles corticales des sons de la LM serviraient de patterns de reconnaissance à l'écoute de ces sons. Ces traces se développeraient graduellement au contact de la LM au cours de la première année de la vie. La présence de telles « traces » dans le cerveau s'accorderait donc avec le phénomène de crible phonologique (chapitre V, 1.2.).

Cependant, la conclusion la plus importante pour notre travail, c'est que la discrimination des voyelles intervient dans l'hémisphère gauche, hémisphère qui, nous l'avons vu, semble dédié au traitement de type analytique.

Dans le domaine de la psychologie cognitive, il existe une autre étude intéressante, réalisée par Bigand et al. (2001). Ces chercheurs ont étudié les effets du contexte harmonique sur la discrimination des voyelles anglaises chantées [ɪ] et [u:] par des étudiants français (19 non musiciens et 21 musiciens). Ces voyelles sont situées sur le dernier accord de la mélodie. Les résultats font apparaître qu'elles sont reconnues plus rapidement lors d'une cadence parfaite (V-I) que lors d'une suite d'accord qui n'apporte aucune résolution (I-IV). Bien que les musiciens tendent à répondre plus rapidement que les non musiciens, les auteurs ne relèvent pas de différence significative entre les deux groupes. Cependant, le principal problème de cette étude tient au choix de la paire de voyelles non natives. En effet, dans le modèle d'assimilation de Best, elles correspondent à la catégorie TC (« Two category ») (voir chapitre V, 1.2.) : chacune des deux voyelles est assimilée à une voyelle native différente. Or, dans ce cas de figure, la discrimination est excellente. Il aurait fallu choisir des voyelles difficiles à discriminer l'une de l'autre, en raison de leur proximité de timbre, comme [ɪ] et [i:] par exemple, dont la discrimination a posé beaucoup de difficultés aux enfants de notre échantillon (test T6).

En conclusion

Nos résultats sur la discrimination des contrastes vocaliques, ainsi que la revue de littérature que nous venons d'exposer vont dans le même sens, c'est-à-dire que les musiciens manifestent un traitement plus analytique que les non musiciens. Un transfert de compétence serait donc tout à fait plausible, les capacités analytiques des musiciens se reportant à la perception des phonèmes, cette dernière nécessitant également un traitement de type analytique. Cependant, en imagerie cérébrale, les études sur la perception des phonèmes sont, à notre connaissance, encore peu nombreuses et une étude similaire à celle de Nataanen et al. (1997), réalisée avec un groupe de sujets musiciens et un groupe de sujets non musiciens pourrait permettre de se prononcer plus précisément.

Bibliographie chapitre VII

- Ayotte, J., Peretz, Isabelle, Rousseau, I., Bard, C. and Bojanowski, M. (2000). Pattern of music agnosia associated with middle cerebral artery artefact. In *Brain*, 123, 1926-1938.
- Bever, Thomas G. and Chiarello, R. J. (1974). Cerebral dominance in musicians and non musicians. In *Science*, 185/537, 137-139.
- Bigand, Emmanuel, Tillmann, B., Poulin, B., D'Adamo, D. A. and Madurell (2001). The effect oh harmonic context on phoneme monitoring in vocal music. In *Cognition*, 81, B11-B20.
- Bradshaw, J. and Nettleton, N. (1981). The nature of hemispheric specialization in man. In *The Behavioral and Brain Sciences*, 4, 51-91.
- Cohen, A. J. and Baird, R. (1990). Acquisition of absolute pitch: the question of critical periods. In *Psychomusicology*, 9, 31-37.
- Delcloque, Philippe (1999). *Speech Technology Applications in CALL*. Dundee: University of Albertay Dundee, 87 p.
- Deutsch (1982). *The Psychology of Music*. New-York: Academic Press.
- Dowling, W. (1982). Melodic information processing and its development. In Deutsch.
- Elbert, T. Pantev, C., Wienbruch, C., Rockstroh, B. and Taub, E. (1996). Increased cortical representation of the fingers of the left hand in string player. In *Science*, 270, 305-307.
- Gates, A. and et Bradshaw, J. (1977). Music perception and cerebral asymmetries. In *Cortex*, 13, 390-401.
- Geschwind, N. and Levitsky, W. (1968). In *Science*, 161 (186).
- Hugdahl (1988). *Handbook of dichotic listening: Theory, methgods and research*. New-York: Wiley.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. In *Psychological Bulletin*, 112, 24-38.
- Kimura, D. (1961a). Some effects of temporal-lobe damage on auditory perception. In *Canadian Journal of Psychology*, 15, 156-165.
- Kimura, D. (1961b). Cerebral dominance and the perception of verbal stimuli. In *Canadian Journal of Psychology*, 15, 166-171.
- Kleist, K. (1959). *Sensorische aphasien und amusien auf myeloarchitektonischer grundlage*. Thieme Verlag, Stuttgart.
- Konopczynski, Gabrielle (1999). L'acquisition du système prosodique de la langue maternelle et ses implications pour l'apprentissage d'une L2. Dans Delcloque Philippe, 62-70.
- Lenneberg, Eric Heinz (1967). *Biological Foundations of Language*. New-York: Wiley & Sons, 489 p.
- Liégeois-Chauvel, C., Peretz, Isabelle, Babai, M and al (1998). Contribution of different cortical areas in the temporal lobes to music processing. In *Brain*, 121, 1853-1867.
- Lhote, Elizabeth (1987). *A la Découverte des Paysages Sonores des Langues*. Besançon : Annales Littéraires de l'Université, 215 p.
- Mazziotta et al (1982).
- Mugny, Gabriel, Oberlé, Dominique et Beauvois, Jean-Léon (1995). *La Psychologie Sociale : Relations Humaines, Groupes et Influence Sociale*. Tome I. Grenoble : P.U.G., 373 p.
- Naatanen, Risto, Lehtokoski, Anne, Lennes, Mietta, Cheour, Marie, Huotilainen, Minna, Iivonen, Antti, Vainio, Martti, Alku, Paavo, Ilmoniemi, Risto, Luuk, Aavo, Allik, Juri, Sinkkhonen, Janne and Alho, Kimmo (1997). Language-specific representations revealed by electric and magnetic brain responses. In *Nature*, 385 (6615), 432-434.
- Nakamura, Satoshi, Sadato, Norihiro, Oohashi, Tsutomu, Nishina, Emi, Fuwamoto, Yoshitaka, Yonekura Yoshiharu (1999). Analysis of music brain interaction with simultaneous measurement of regional cerebral blood flow and electroencephalogram beta rhythm in human subjects. In *Neuroscience Letters*, 275, 222-226.

- Pantev, Christo, Oostenveld, Robert, Engelien, Almut, Ross, Bernhard, Roberts, Larry, E. and Hoke, Manfred (1998). Increased auditory cortical representation in musicians. In *Nature*, 392/6678, 811-814.
- Peretz, M. (1990). Processing of local and global musical information by unilaterally brain damaged patients. In *Brain*, 113, 1185-1205.
- Peretz, Isabelle et Morais, José (1988). Determinants of laterality for music: Towards an information processing account. In Hugdahl,
- Peretz, Isabelle, Morais, José and Bertelson, P. (1987). Shifting ear differences in melody recognition through strategy inducement. In *Brain and Cognition*, 6, 202-215.
- Schlaug, Gottfried, Jaencke, Lutz, Huang, Yanxiong and Steinmetz, Helmuth (1995). In vivo evidence of structural brain asymmetry in musicians. In *Science*, 267 (5198), 699-701.
- Shankweiler (1966). Effects of temporal-lobe damage on perception of dichotically presented melodies. In *Journal of Comparative and Physiological Psychology*, 62, 115-119.
- Scott, Thomas (1987). Le français face au paysage anglais. Dans Lhote, 131-142.
- Tramo, Mark Jude (2001). Music of the hemispheres. In *Science*, 291 (5501), 54-56.
- Trehub, Sandra E., Bull, Dale and Thorpe, Leigh A. (1984). Infants' perception of melodies: The role of melodic contour. In *Child Development*, 55, 821-830.
- Williams, Linda V. (1986). *Deux Cerveaux pour Apprendre : le Gauche et le Droit*. Paris : les Editions d'Organisation, 204 p.
- Zenatti, Arlette (1980). *Tests Musicaux pour Jeunes Enfants, avec Applications en Psychopathologie de l'Enfant et de l'Adulte*. Issy Les Moulineaux : Editions Scientifiques et Psychologiques, 53 p.

Chapitre VIII

*Effets de la formation musicale
sur la production
de l'anglais*

Table du huitième chapitre

1) Recueil du matériel de parole

- 1.1. Type de discours
- 1.2. Passation
 - 1.2.1. Entretien avec un locuteur natif
 - 1.2.2. Conditions d'enregistrement
 - 1.2.3. Déroulement de l'entretien
 - 1.2.4. Matériel de parole

2) Etude n°3 « Production de la prosodie »

- 2.1. Plan expérimental
 - 2.1.1. Hypothèse
 - 2.1.2. Variables indépendantes
 - 2.1.2.1. La musique
 - 2.1.2.2. Le discours
 - 2.1.3. Variables dépendantes
 - 2.1.3.1. Forme de la courbe
 - 2.1.3.1.1. Caractéristiques des productions du locuteur natif
 - 2.1.3.1.2. Classement des contours en fonction de l'initiale du mot
 - 2.1.3.1.3. Les cinq modalités
 - 2.1.3.1.4. Classement des courbes
 - 2.1.3.2. Exagération de la courbe
 - 2.1.3.3. Localisation de l'accent
 - 2.1.4. Analyses acoustiques
- 2.2. Résultats
 - 2.2.1. Forme de la courbe
 - 2.2.1.1. Mots dont l'initiale est non voisée
 - 2.2.1.2. Mots dont l'initiale est voisée
 - 2.2.2. Amplitude de la courbe
 - 2.2.3. Localisation de l'accent

3) Etude n°4 « Production des noyaux vocaliques »

- 3.1. Plan expérimental
 - 3.1.1. Hypothèse
 - 3.1.2. Variables indépendantes
 - 3.1.2.1. La musique
 - 3.1.2.2. Le discours
 - 3.1.3. Variables dépendantes
 - 3.1.3.1. Timbre du noyau vocalique
 - 3.1.3.2. Durée du noyau vocalique
 - 3.1.4. Mesures acoustiques
 - 3.1.4.1. Mesures de durée
 - 3.1.4.2. Mesures de timbre
- 3.2. Résultats
 - 3.2.1. Timbre du noyau vocalique
 - 3.2.1.1. Voyelles

- 3.2.1.2. Diphtongues
- 3.2.1.3. Résumé
- 3.2.2. Durée du noyau vocalique
 - 3.2.2.1. Voyelles brèves
 - 3.2.2.2. Voyelles longues
 - 3.2.2.3. Diphtongues

4) Etude n°5 « Stabilité de timbre »

- 4.1. Plan expérimental
 - 4.1.1. Hypothèse
 - 4.1.2. Variables indépendantes
 - 4.1.2.1. La musique
 - 4.1.2.2. Le discours
 - 4.1.3. Variable dépendante : durée de la zone de stabilité de timbre de la voyelle
- 4.2. Résultats
 - 4.2.1. Résultats généraux
 - 4.2.2. Résultats détaillés
 - 4.2.2.1. Voyelles longues
 - 4.2.2.2. Voyelles brèves

5) Résumé des résultats et discussion générale

- 5.1. Production de la prosodie
 - 5.1.1. Forme de la courbe
 - 5.1.2. Localisation de l'accent
 - 5.1.3. Exagération de la courbe
- 5.2. Production des contrastes vocaliques
 - 5.2.1. Timbre du noyau vocalique
 - 5.2.1.1. Voyelles brèves
 - 5.2.1.2. Voyelles longues
 - 5.2.1.3. Diphtongues
 - 5.2.2. Durée du noyau vocalique
 - 5.2.3. Stabilité de timbre

Bibliographie du chapitre VIII

Dans l'apprentissage d'une langue étrangère, l'objectif consiste à se rapprocher le plus possible de la langue cible et, pour ce qui nous intéresse, de la prononciation native. Analyser, puis juger les productions des apprenants d'une langue donnée, c'est donc déterminer où ils se situent par rapport au « modèle » à acquérir et quelles stratégies sont mises en œuvre pour y parvenir ; c'est déterminer également si la LM produit des interférences avec la langue cible. Dans le chapitre V, nous avons décrit le système prosodique et vocalique de l'anglais « standard » (dans le sens du modèle à acquérir) par rapport à celui du français « standard », de façon à pouvoir prévoir ces interférences. Dans ce contexte, est-ce que la pratique musicale d'un enfant va lui permettre d'« être » mieux armé pour se confronter à une nouvelle langue et la reproduire ? Nous avons vu que les capacités musicales ont une influence sur la perception des contrastes vocaliques. Est-ce que le fait d'avoir une oreille musicienne va permettre de mieux reproduire ce que l'on entend mieux ? Et est-ce que les capacités à percevoir vont susciter une meilleure production des éléments musicaux de la langue ? Afin de le savoir, nous avons réalisé une étude sur la qualité des productions en anglais des enfants de notre corpus, en fonction de la « qualité » de leur oreille et de leur pratique musicale.

Nous posons donc comme hypothèse générale que la pratique musicale induit une meilleure production d'une langue étrangère : *« une oreille affinée par la pratique musicale induira une meilleure production d'une langue étrangère »*. Cette hypothèse peut être appréhendée en deux hypothèses particulières : une oreille affinée par la pratique musicale induira une meilleure production de la prosodie d'une langue étrangère ; une oreille affinée par la pratique musicale induira une meilleure production des contrastes vocaliques d'une langue étrangère.

1) Recueil du matériel de parole

1.1. Type de discours

Les enfants sont débutants en LV, on ne peut donc pas s'attendre à ce qu'ils s'expriment spontanément en anglais. Nous avons donc exclu de travailler sur du discours spontané. En revanche, les enfants apprennent un certain nombre de mots en classe et ce répertoire de mots est commun à de nombreux débutants. Ainsi, ils commencent par apprendre à compter, à nommer les animaux familiers, les couleurs et les objets de leur entourage. Ce répertoire constitue le premier vocabulaire qu'un apprenant enfant mémorise dans une langue étrangère, et pour ce qui nous intéresse, en anglais.

Nous avons ainsi constitué une liste de 43 mots, tirés du vocabulaire des couleurs, des nombres, des animaux et des objets familiers (voir matériel de parole, 1.2.4.). Nous avons également interrogé les professeurs d'anglais des trois écoles pour connaître le vocabulaire que les enfants avaient appris depuis le début de l'année scolaire (débutée il y a 6 mois).

Dans des études précédentes (Dodane, 1997, 2000), nous avons observé des comportements différents chez les enfants lorsqu'ils reproduisaient des mots pour la première fois, par rapport au moment où ils répétaient des mots qu'ils avaient déjà assimilés. En « répétition immédiate », l'apprenant essaie de reproduire fidèlement le mot qu'il entend de la bouche du locuteur, car il ne connaît pas ce mot. Ce type d'imitation est nécessaire pour lui permettre de réussir à reproduire et à mémoriser le mot cible. En « répétition différée », il a déjà mémorisé le mot et il peut « jouer » avec lui, de manière à se l'approprier complètement, quitte à exagérer les caractéristiques sonores qui lui paraissent « nouvelles » par rapport à celles de sa LM. Ce comportement est positif, puisqu'il démontre que l'apprenant est en pleine phase d'apprentissage et d'appropriation. On peut rapprocher la notion de « répétition immédiate » de la notion d'« *imitation-copie* » et celle de « répétition différée », d'« *imitation-jeu* ». Selon Piaget (1937), ces deux types d'imitation sont indispensables à l'enfant car ils lui permettent de s'adapter à son environnement et ici, de s'approprier progressivement une langue qui lui est étrangère.

Cependant, il est impossible de déterminer a priori quels sont les mots que les enfants connaissent et ceux qu'ils ne connaissent pas. Leur entourage a pu en effet leur apprendre des mots supplémentaires par rapport au vocabulaire appris en classe et ils ont pu oublier des mots qu'ils avaient appris en début d'année. Les enfants des trois classes sont donc censés connaître une partie des mots de cette liste, mais pas tous les mots. Nous avons donc établi notre classement de la manière suivante : si les enfants connaissent les mots de la liste, leurs productions sont classées dans la catégorie de discours « répétition différée » (RD). S'ils ne les connaissent pas et qu'ils doivent les répéter à la suite du locuteur natif, leurs productions sont classées dans la catégorie de discours « répétition immédiate » (RI).

1.2 Passation

1.2.1. Entretien avec un locuteur natif

Afin d'obtenir ces deux types de discours, nous avons choisi de placer les enfants dans une situation d'entretien avec un locuteur anglophone natif qui n'est pas l'enseignant habituel de la classe et que les enfants n'ont jamais rencontré auparavant. Cette situation d'entretien nous permet de conserver les mêmes conditions d'expérimentation pour les 3 classes (personne inconnue, la même pour les 3 classes). La situation d'entretien avec un anglophone est au départ une situation un peu angoissante pour l'enfant, qui a l'habitude de s'exprimer en classe et avec son professeur habituel. Les entretiens ont donc eu lieu juste après les tests de musique (chapitre VI) et les tests de perception en anglais (chapitre VII). Ainsi, les enfants, déjà habitués à l'expérimentatrice principale, ne se sont pas montrés trop déroutés par l'arrivée d'un nouvel inconnu, qui plus est, parlait une langue étrangère.

Le locuteur anglophone natif avait déjà fait l'expérience de ce genre d'entretien puisqu'il avait participé à l'élaboration de notre corpus de maîtrise, réalisé dans des conditions similaires (Dodane, 1996). Nous connaissions donc son comportement avec les enfants et c'est pour cette raison que nous avons de nouveau fait appel à lui. Son accent est non marqué, bien qu'il soit natif américain. Cependant, la question de l'accent de ce locuteur a relativement peu d'importance étant donné que notre étude ne porte pas sur cet accent, mais sur la capacité des enfants à reproduire. Le locuteur natif ne devait jamais s'exprimer en français pendant toute la durée de l'entretien. En cas de chevauchement ou d'inintelligibilité, il avait pour consigne de faire répéter les mots aux enfants, de manière à ce que ces mots soient exploitables pour l'analyse acoustique. Si leur voix était trop faible, il leur demandait de parler plus fort et si les enfants répondaient en même temps ou s'il y avait un bruit (la sonnerie au moment de la récréation par exemple), il leur faisait répéter le mot.

1.2.2. Conditions d'enregistrement

Dans les trois écoles, les enregistrements ont eu lieu dans une salle familière des enfants, la plus calme possible (salle de classe de chant pour les enfants de l'école de Dijon, salle de classe d'anglais pour les enfants de l'école de Choisey et salle de langue pour les enfants de l'école de Dole). En raison de la législation concernant la protection des mineurs, un enfant ne peut rester seul avec une ou plusieurs personnes adultes étrangères à l'établissement scolaire.

Comme pour les tests de perception, les enfants passent donc par deux, ce qui porte à quatre le nombre de personnes présentes dans la pièce : l'expérimentateur principal, le locuteur natif et deux enfants. Les enregistrements ont eu lieu au cours des mois de janvier et février 2001 alors que les enfants avaient commencé l'anglais depuis 6 mois. Ils avaient donc un niveau suffisant pour participer à un entretien avec un locuteur natif. Pour enregistrer les productions des enfants, nous avons utilisé un enregistreur mini-disque (de marque Sony, modèle « Portable Minidisc Recorder, M2-R30 ») avec un microphone (de marque Sony).

1.2.3. Déroulement de l'entretien

L'entretien se déroule en cinq étapes. La première commence par la présentation entre les enfants et le locuteur natif. Les enfants ont déjà appris à se saluer, à se présenter, à dire leur âge et comment ils vont. Le locuteur commence par se présenter, par exemple « *Hello, my name is Ronald* » puis il pose des questions aux enfants telles que « *What is your name ?* », « *How are you ? Are you fine ?* » ou encore « *How old are you ?* ». Toutefois, tous les enfants n'étaient pas capables de répondre à ces questions et ce matériel ne peut donc être étudié systématiquement dans notre corpus.

Après les présentations, viennent les couleurs. Le locuteur montre aux enfants des couleurs tirées d'un livre illustré (Beaumont et Desmoineaux, 2000). Il fait la même chose pour les animaux, tirés d'un autre livre illustré destiné à des enfants anglophones (Butterfield et O'Neill, 1997), dans lequel un système de pliage permet de découvrir les animaux, cachés derrière un tas de foin ou une porte d'écurie par exemple. Viennent ensuite les nombres, pour lesquels nous avons réalisé des supports cartonnés les représentant, de 1 à 13. Enfin, le locuteur termine avec des images tirées de l'Imagier du Père Castor en anglais (1997). Dans ce livre, les mots qui sont indiqués sous les images ont été occultés par des caches. Nous avons diversifié les supports de manière à introduire de la nouveauté dans un entretien qui pouvait s'avérer un peu long pour les enfants (les entretiens pouvaient durer jusqu'à 20 minutes environ).

Tout au long de l'entretien, le locuteur natif devait susciter la production en répétition différée des mots en montrant une image et en posant simultanément une question aux enfants du type « *What colour is it ?* », « *What is this animal ?* », « *Who is it ?* » ou encore « *What's this ?* », « *What is it ?* » ou simplement « *And this ?* » (voir exemple d'entretien, Volume II, annexe n°8.1. et **extrait sonore n°54**).

Ces questions étaient comprises par les enfants puisqu'elles étaient posées régulièrement par leur professeur d'anglais pendant les cours. Pour susciter la répétition immédiate, le locuteur pose les mêmes questions, mais après le temps de silence lui indiquant que l'enfant ne peut pas répondre, il dit le mot et demande de répéter après lui. Par exemple, « *And this ? What's this ?* (il attend, tout en désignant l'image) *No ? It's a duck. Can you repeat, a duck* » ou « *This... this animal. What is this animal ?* (il attend) *This animal is a parrot* » ou encore « *And this ?* (il attend) *Can you say what is this ?* (il attend) *No ?* (il attend) *It's a car* » et l'enfant doit répéter.

1.2.4. Matériel de parole

La liste est composée de 43 mots, dont 10 mots désignent des couleurs (« *green, pink, red, black, orange, blue, purple, white, yellow, brown* »), 7 mots, des animaux (« *dog, cow, sheep, bird, cat, rabbit, horse* »), 13 mots, des nombres (suite de 1 à 13) et 13 mots, des objets ou des êtres vivants familiers (« *boy, baby, doll, teddy, book, cake, fish, apple, parrot, duck, umbrella, car, banana* »). Parmi ces mots, 30 sont monosyllabiques, 10 sont bisyllabiques et 3 trisyllabiques (voir présentation du corpus, Volume II, annexe n°8.2.).

2) Etude n°3 « Production de la prosodie »

2.1. Plan expérimental

2.1.1 Hypothèse

Les locuteurs dont l'oreille a été entraînée par la pratique musicale rencontreront plus de facilités à reproduire les éléments musicaux de l'anglais.

2.1.2. Variables indépendantes

2.1.2.1. La musique : deux modalités

La variable indépendante « musique » est constituée de deux modalités, la modalité « musique » (G1) et la modalité « non musique » (G2) (voir chapitre VII, 1.1.2.) Cependant, par rapport aux études n°1 et n°2 sur la perception de l'anglais, nous avons dû réduire l'effectif du groupe G2 à 25 sujets, car l'un des entretiens n'était pas exploitable pour une analyse acoustique (mauvais rapport signal/bruit).

2.1.2.2. Le discours : deux modalités

Au cours de l'entretien avec le locuteur natif, les enfants sont amenés à produire deux types d'énoncés, des énoncés en répétition immédiate (RI) à la suite du locuteur ou des énoncés en répétition différée (RD) du vocabulaire appris en classe (voir 1.1.).

2.1.3. Variables dépendantes

2.1.3.1. Forme de la courbe

2.1.3.1.1. Caractéristiques des productions du locuteur natif

A cause de l'entraînement de leur oreille, les enfants musiciens devraient avoir plus de facilité à repérer les contours caractéristiques de l'anglais et donc, si nous supposons qu'une meilleure perception entraîne une meilleure production, ils réussiraient mieux à en reproduire la forme. Le contour typique qui affecte les mots monosyllabiques commençant par une initiale voisée en anglais est un contour qui prend la forme d'une cloche (montée, pic de F_0 sur le noyau vocalique, puis descente). La figure n°83 nous en montre un exemple pour le mot « green »¹ (extrait sonore n°55).

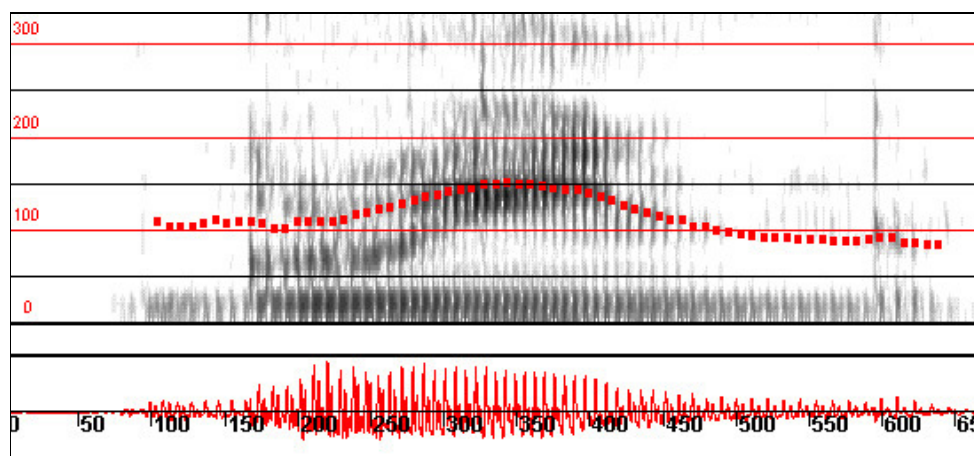


Figure n°83 : courbe mélodique en cloche sur le mot « green » par le locuteur natif
(représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

Sur chacun des 17 mots monosyllabiques de notre corpus qui commencent par une initiale voisée, (voir la liste des mots dans l'annexe n°X et les mesures réalisées, volume II, annexe 9.1.1.), le locuteur natif produit un contour en cloche.

¹ Durée de l'énoncé, 546 ms – points d'inflexion de la courbe « en cloche » : 105 Hz / 150 Hz / 85 Hz.

Ces mots peuvent commencer soit par une consonne voisée comme [g] dans « *green* », soit par une voyelle comme [ʌ] dans « *one* ».

Mais le locuteur natif ne produit pas que des « cloches » sur les mots monosyllabiques. Il produit également des contours descendants dans les mots commençant par une consonne non voisée. C'est le cas pour les 11 autres mots du corpus. La figure n°84 nous donne un exemple de ce type de contour, qui ici, affecte le mot « *ten* »² (**extrait sonore n°56**).

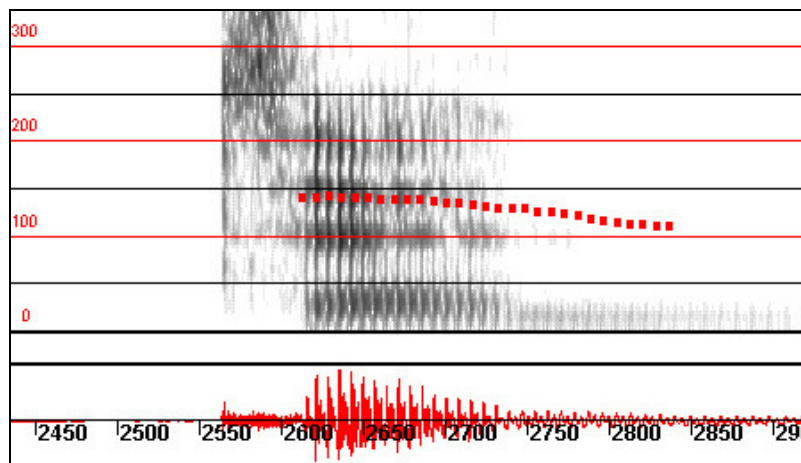


Figure n°84 : courbe mélodique descendant sur le mot « *ten* » par le locuteur natif (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

Il y a cependant deux exceptions à ce classement, le mot « *twelve* » et le mot « *three* » qui commencent tous les deux par une consonne non voisée, mais sont affectés d'un contour en cloche. La présence de ce contour s'explique par le fait qu'un son voisé est intercalé entre la consonne et la voyelle, dans ces deux mots « *w* » ou « *r* ».

2.1.3.1.2. Classement des contours en fonction de l'initiale du mot

En fonction de ces données, nous classerons les productions des enfants en deux catégories :

- la première catégorie « initiale non voisée » comprend les mots commençant par une consonne non voisée ; elle englobe 11 mots sur les 30 mots du corpus ;
- la deuxième catégorie « initiale voisée » comprend les mots commençant par une consonne voisée, par une voyelle ou par une consonne sourde suivie d'un son voisé ; elle englobe 19 mots sur les 30 mots du corpus.

² Durée de l'énoncé, 437 ms – points d'inflexion de la courbe « en descente » : 142 Hz / 110 Hz.

2.1.3.1.3. Les cinq modalités

Si les enfants restituent parfaitement les contours des 30 mots de notre corpus, ils devraient donc produire un contour en cloche pour les mots commençant par un son voisé et un contour descendant pour les mots commençant par une consonne non voisée. Mais, le contour en cloche est caractéristique de l'anglais et il est probable que certains enfants, à cause du crible prosodique, produisent des contours descendants comme ils le feraient dans leur langue maternelle. Par ailleurs, les enfants peuvent également produire des contours montants, des contours plats ou des contours en forme de « cloche inversée » correspondant à un contour de type haut-bas-haut (désormais HBH).

La variable dépendante « forme » de la courbe est donc une variable à 5 modalités : contour en forme de « cloche », de « cloche inversée », de « descente », de « montée » ou « recto-ono » (voir figure n°85 et figure n°86, correspondant à l'**extrait sonore n°57**).

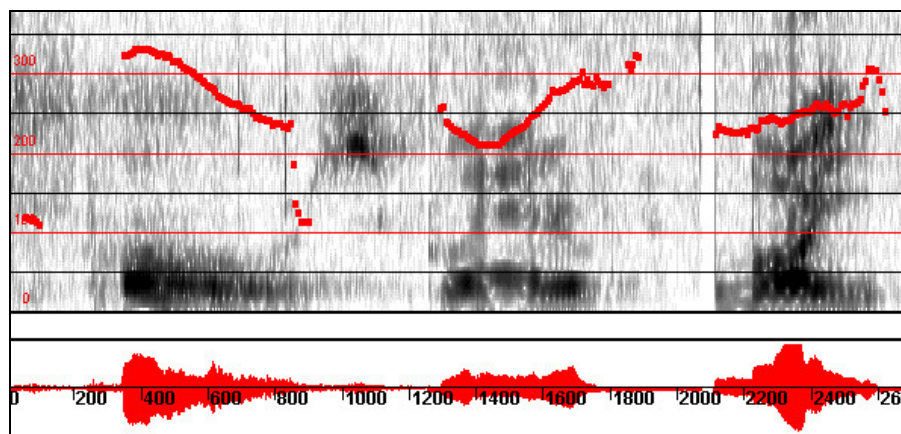


Figure n°85 : courbes mélodiques affectant les mots « horse », « twelve » et « boy » représentant successivement les modalités « cloche », « cloche inversée », « montée » (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

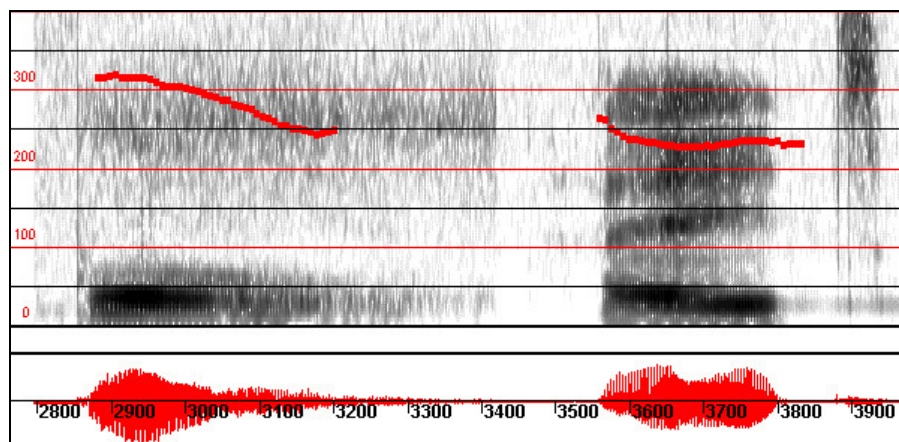


Figure n°86 : courbes mélodiques affectant les mots « cow » et « eight » représentant successivement les modalités « descente » et « recto-tono » (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

Pour cette variable, nous posons l'hypothèse opérationnelle que la proportion de contours en « cloche » sera plus importante dans les mots monosyllabiques à l'initiale voisée chez les enfants appartenant au « groupe musiciens » (G1) que chez les enfants du « groupe non musiciens » (G2). Par ailleurs, le contour en cloche est caractéristique de l'anglais et il est possible que certains enfants généralisent son utilisation pour donner une couleur plus « anglaise » à leurs énoncés, et ce, même dans les occurrences où il ne devrait pas apparaître. Nous posons donc comme hypothèse que les enfants du groupe G1 produiront plus de contours en cloche dans les mots monosyllabiques à l'initiale non voisée que les enfants du groupe G2. Le plan expérimental est exposé dans le tableau n°47.

Initiale de mot	Type de contour	G1		G2	
		RD	RI	RD	RI
Non voisée	« Cloche »				
	« Cloche inversée »				
	« Montée »				
	« Descente »				
	« Recto-tono »				
Voisée	« Cloche »				
	« Cloche inversée »				
	« Montée »				
	« Descente »				
	« Recto-tono »				

Tableau n°47 : plan expérimental pour la variable « forme (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate et « RD » pour répétition différée »).

2.1.3.1.4. Classement des courbes

Nous avons volontairement limité le nombre de catégories à cinq, notre but n'étant pas de réaliser une analyse prosodique fine et détaillée, mais de faire apparaître des tendances générales dans la production des contours. Ces contours ont été répartis dans les différentes catégories en fonction de la forme de leur courbe mélodique apparaissant sur la représentation mélodique du logiciel de mesure et du jugement auditif de l'expérimentatrice. Pour le « groupe musiciens » (G1), 300 énoncés ont ainsi été collectés, parmi lesquels 26 énoncés n'ont pu être analysés (soit 8,66 % de la totalité des énoncés). Pour le « groupe non musiciens » (G2), 450 énoncés ont été collectés, dont 27 n'ont pas pu être analysés (soit 6 % de la totalité des énoncés). Le total des énoncés analysés se porte donc à 697, dont 274 pour G1 et 423 pour G2.

2.1.3.2. Exagération de la courbe

Des études précédentes (Dodane, 1997, 2000) ont montré que certains enfants musiciens exagéraient les paramètres prosodiques et spécialement la courbe mélodique lorsqu'ils restituaient des mots qu'ils avaient appris en classe (répétition différée, RD). Cette exagération pouvait être franchement caricaturale par rapport aux productions d'un locuteur natif. Est-ce que cette exagération est caractéristique d'un locuteur qui possède une oreille entraînée par la pratique musicale ? En effet, un enfant musicien, habitué à manipuler les paramètres musicaux au cours de sa formation musicale, pourrait reporter cette stratégie dans la restitution des contours mélodiques d'une langue étrangère, en particulier lorsque les sons de cette langue sont difficiles à reproduire.

Il serait donc intéressant d'examiner si les contours des enfants musiciens sont plus amples que ceux du locuteur natif dans notre corpus. Afin de simplifier la comparaison et de restreindre la complexité des contours, nous nous limiterons aux seuls mots monosyllabiques du corpus, la forme de la courbe étant nettement influencée par la localisation de l'accent dans les mots multisyllabiques. Nous avons vu précédemment que dans ce type de mots, le locuteur natif produisait deux types de contours, des contours descendants et des contours en cloche. Pour connaître l'amplitude des contours produits par le locuteur natif, une analyse a été réalisée sur 37 contours en cloche et 17 contours descendants (les résultats détaillés sont exposés dans le volume II, annexe n°9.2.1.) ; ces énoncés ont été extraits de deux entretiens différents entre le locuteur natif et quatre enfants prélevés dans le corpus.

Le nombre d'énoncés analysés est limité, le but n'étant pas de faire une analyse exhaustive des productions du locuteur natif, mais de celles des enfants (les contours exposés pour la variable précédente sont compris dans cette analyse). Néanmoins, elle nous permettra d'avoir une idée du portrait prosodique du locuteur natif.

Pour les contours en cloche, le contour « moyen » (pour 37 productions) est composé d'une inflexion montante de $8,21 \frac{1}{4}$ de tons, soit approximativement d'une sixte mineure et d'une inflexion descendante de $13,67 \frac{1}{4}$ de tons, soit approximativement d'une neuvième mineure (figure n°87).

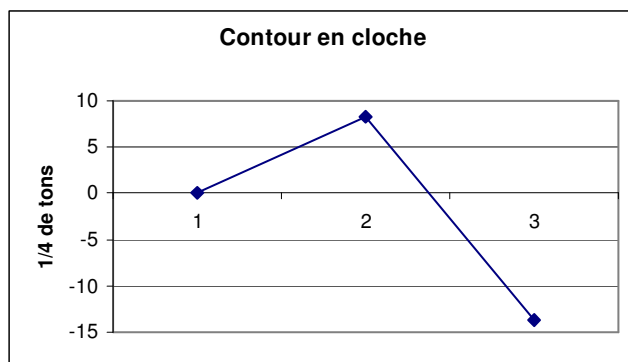


Figure n°87 : « contour moyen » construit à partir de 37 contours en cloche produit par le locuteur natif.

Le contour le plus ample a été relevé pour le mot « blue »³. La première inflexion du contour s'élève d'un intervalle de $21 \frac{1}{4}$ de tons, soit approximativement d'une septième mineure, et la deuxième inflexion s'abaisse d'un intervalle de $24 \frac{1}{4}$ de tons, soit approximativement une octave (figure n°88 et **extrait sonore n°58**).

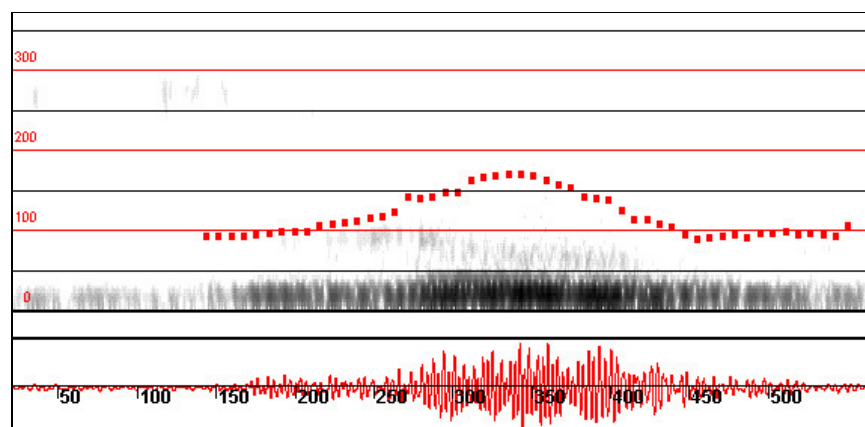


Figure n°88 : courbe mélodique sur le mot « blue » par le locuteur natif (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

³ Durée de l'énoncé, 538 ms – points d'inflexion de la courbe en cloche : 92 Hz / 170 Hz / 93 Hz.

Pour les contours descendants, l'intervalle moyen (pour 17 productions) est de $10,40 \frac{1}{4}$ de tons, soit approximativement d'une septième mineure (figure n°89). Le contour le plus ample a été relevé pour le mot « *pink* »⁴, qui s'abaisse d'un intervalle de $20 \frac{1}{4}$ de tons, soit approximativement une septième mineure (voir figure n°90 et **extrait sonore n°59**).

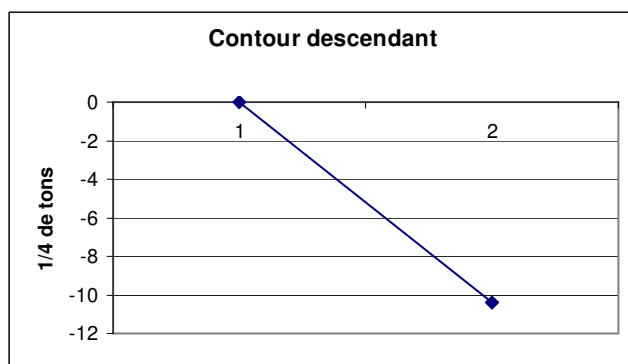


Figure n°89 : « contour moyen » construit à partir de 17 contours descendants produit par le locuteur natif.

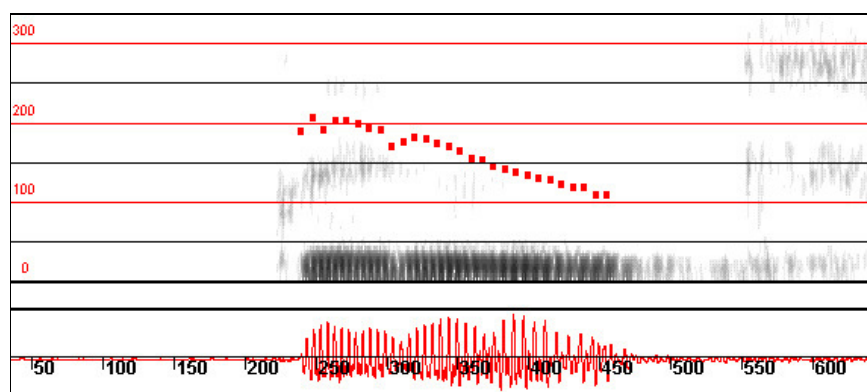


Figure n°90 : courbe mélodique sur le mot « *pink* » par le locuteur natif (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de F_0 indiquée sur l'axe des abscisses, à l'extrême gauche).

Les contours du locuteur natif sont donc très amples et nous pensons que les enfants musiciens vont reproduire plus facilement cette amplitude de la courbe mélodique, voire l'exagérer.

En fonction des différents types de contours produits par les enfants dans notre corpus, la variable « exagération de la courbe » comprend quatre modalités : « cloche », « cloche inversée », « montée » et « descente » (nous avons retiré de l'analyse les contours « recto-ono »). Nous posons l'hypothèse opérationnelle qu'en situation de répétition différée (RD), l'amplitude de la courbe sera exagérée chez les enfants musiciens.

⁴ Durée de l'énoncé, 476 ms – points d'inflexion de la courbe en descente : 206 Hz / 110 Hz.

Le plan expérimental est exposé dans le tableau n°48.

	G1				G2			
	RD		RI		RD		RI	
	A	B	A	B	A	B	A	B
Cloche								
Cloche Inversée								
Montée								
Descente								

Tableau n°48 : plan expérimental pour la variable « exagération de la courbe » (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et « RD » pour « répétition différée » ; A désigne l'intervalle correspondant à la première inflexion de la courbe et B, l'intervalle correspondant à la deuxième inflexion de la courbe pour les contours en cloche et en cloche inversée).

Afin d'étudier l'amplitude de la courbe, nous avons relevé les intervalles (exprimés en ¼ de tons) séparant :

- Le point de départ et le point d'arrivée de la courbe dans les contours montants et descendants ;
- Le point de départ et le point d'arrivée de la première inflexion de la courbe dans les contours en cloche et en cloche inversée (intervalle A) ;
- Le point de départ et le point d'arrivée de la deuxième inflexion de la courbe dans les contours en cloche et en cloche inversée (intervalle B).

Le nombre d'énoncés analysés dans le corpus figure dans le tableau n°49, ci-dessous.

	G1		G2	
	RD	RI	RD	RI
Nombre d'énoncés	122	109	243	172
Total	231		415	
Total général	646			

Tableau n°49 : nombre d'énoncés analysés dans l'étude de la variable « amplitude de la courbe » (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

2.1.3.3. Localisation de l'accent

Dans le chapitre V (2.1.3.) et dans le catalogue d'interférences (Volume II, annexe n°3.2.1., « erreur de localisation de l'accent tonique »), nous avons vu que les francophones ont tendance à produire l'accent en finale de groupe rythmique, alors qu'en anglais, la tendance est à l'accentuation de la syllabe initiale. Est-ce que les enfants musiciens (G1) restitueront mieux la place de l'accent ?

Par ailleurs, il est probable qu'en répétition immédiate (RI), les accents soient mieux localisés qu'en répétition différée puisque les enfants répètent les mots tout de suite après le locuteur natif. En revanche, la répétition différée met en jeu la mémoire à long terme de l'enfant et il est possible que, dans cette situation, les interférences avec la LM soient plus puissantes. Le corpus est composé de 10 mots bisyllabiques et de 3 mots trisyllabiques (avoir annexe n°X). Dans les mots bisyllabiques, l'accent tombe sur la première syllabe comme dans le mot « *parrot* »⁵ par exemple (figure n°91 et **extrait sonore n°60**) ; dans les mots trisyllabiques, l'accent tombe sur la deuxième syllabe dans les mots trisyllabiques comme dans le mot « *eleven* »⁶ (figure n°92 et **extrait sonore n°61**).

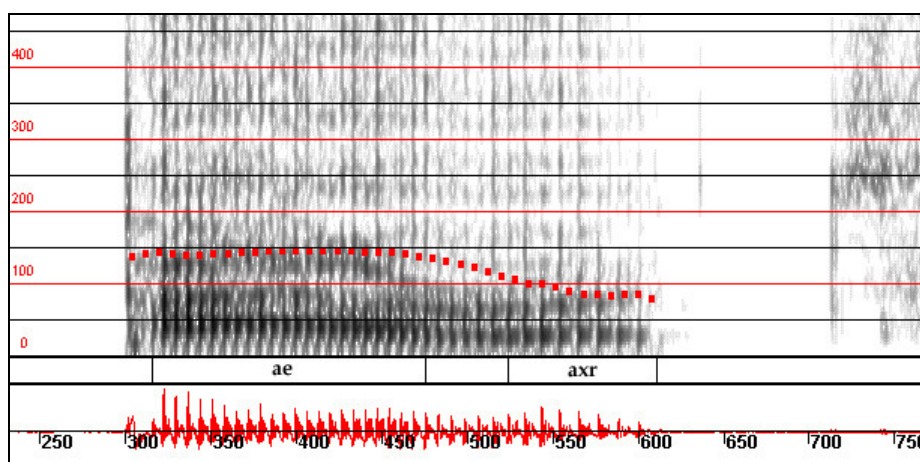


Figure n°91 : courbe mélodique affectant les mots « *parrot* », (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

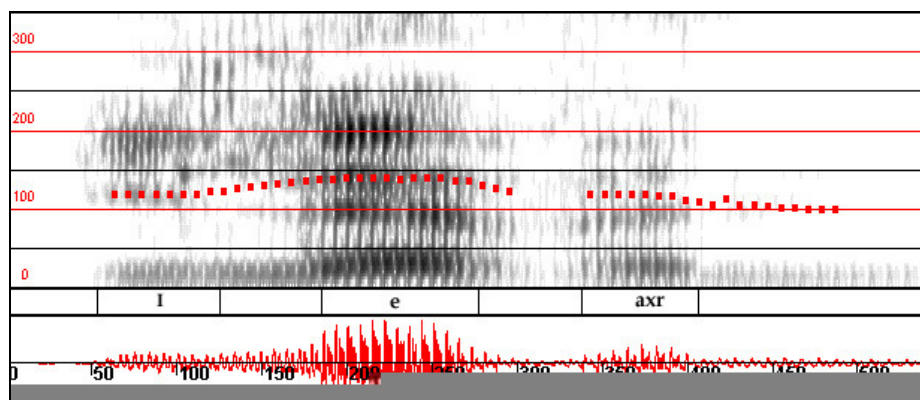


Figure n°92 : courbe mélodique affectant les mots « *eleven* », (représentation de la courbe mélodique superposée à une représentation spectrographique à bandes larges – échelle de Fo indiquée sur l'axe des abscisses, à l'extrême gauche).

⁵ Pic de Fo sur la voyelle [æ], d'une durée de 161 ms (la voyelle [ə] mesurant 74 ms) – Contour de Fo : 146 – 80 Hz.

⁶ Pic de Fo sur la voyelle [e], d'une durée de 139 ms (la voyelle [ɪ] mesurant 73,7 ms et la voyelle [ə], 67 ms) – Contour de Fo : 118-139-99 Hz.

Influencés par l'accentuation finale du français, les enfants pourront produire l'accent sur la deuxième syllabe dans les mots bisyllabiques. L'**extrait sonore n°62** nous fournit un exemple de ce type d'erreur, le mot « seven » étant accentué par l'enfant sur la deuxième syllabe. Pour la même raison, l'accent pourra être localisé sur la dernière syllabe des mots trisyllabiques, mais également sur la première. L'**extrait sonore n°63** nous fournit trois exemples de reproductions du mot « eleven », la première étant accentuée sur la première syllabe, la seconde, correcte, sur la deuxième syllabe et la troisième sur la dernière syllabe.

La variable « localisation de l'accent » comprend donc deux modalités pour les mots bisyllabiques : la modalité « première syllabe » (A) et la modalité deuxième syllabe (B) ; une troisième modalité se rajoute pour les mots trisyllabiques « troisième syllabe » (C). Nous posons comme hypothèse opérationnelle que la place de l'accent sera mieux localisée dans les productions des enfants musiciens. Le plan expérimental est exposé dans le tableau n°50.

Type de mot	Syllabe	G1		G2	
		RD	RI	RD	RI
Mots bisyllabiques	A				
	B				
Mots trisyllabiques	A				
	B				
	C				

Tableau n°50 : plan expérimental pour la variable « localisation de l'accent » (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, A pour première syllabe accentuée, B pour deuxième syllabe accentuée et C pour troisième syllabe accentuée).

Pour le groupe G1, la place de l'accent a été localisée dans 83 mots bisyllabiques, 17 n'ayant pu être analysés (soit 17 % de la totalité des énoncés) et pour le groupe G2, dans 122 mots bisyllabiques, 28 n'ayant pu être analysés (soit 22,95 % de la totalité des énoncés). Pour les mots trisyllabiques, la place de l'accent a été localisée dans 24 mots ont été analysés pour le groupe G1, 6 n'ayant pu être analysés (soit 25 % de la totalité des énoncés) et 40 mots pour le groupe G2, 5 n'ayant pu être analysés (soit 12,5%). Le nombre de mots analysés figure de manière détaillée dans le tableau n°51.

	G1		G2	
	RD	RI	RD	RI
Bisyllabiques	41	42	42	80
Trisyllabiques	4	20	18	22
Total	45	62	60	102

Tableau n°51 : nombre d'énoncés multisyllabiques analysés (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate).

2.1.4. Analyses acoustiques

Les analyses acoustiques ont été effectuées avec le logiciel Winsnoori⁷, la détection de la mélodie avec ce logiciel s'étant avérée particulièrement robuste pour l'analyse des voix d'enfants (certains enfants s'exprimant avec une voix très faible, d'autres avec des voix très aiguës) et dans les passages bruités (les enregistrements ayant eu lieu dans les écoles et non pas en chambre sourde).

Pour chaque mot du corpus, nous avons relevé :

- la fréquence correspondant à chaque point d'inflexion de la courbe mélodique (en Hz) ; à partir de ces mesures, les intervalles en ¼ de tons (Konopczynski et Mac Carthy, 1977) séparant deux points consécutifs de la courbe⁸ ont été déterminés ;
- la place de l'accent en combinant notre propre jugement auditif (notre oreille étant entraînée par la pratique musicale), à la localisation du pic de Fo et à la durée syllabique.

2.2. Résultats

2.2.1. Variable « Forme de la courbe »

Une fois analysés, tous les mots du corpus ont été répartis en fonction de leur appartenance à une catégorie spécifique de contours « cloche, cloche inversée, montée, descente, recto-ono », en fonction de leur initiale (« voisée ou non voisée »), en fonction du groupe (« G1, G2 ») et en fonction du type de discours (« RD, RI »). Une fois cette répartition achevée, la proportion des contours a été calculée en fonction du nombre total d'énoncés produits par les enfants dans chaque type de discours et dans chaque groupe. Cette proportion est exprimée en pourcentage (se reporter au Volume II, annexe n°9.1.2. pour les résultats exprimés en nombre d'énoncés). Les résultats sont exposés dans le tableau n°52 pour les énoncés dont l'initiale est non voisée (figure n°93 et n°94) et dans le tableau n°53 pour les énoncés dont l'initiale est voisée (figure n°95 et n°96).

⁷ Logiciel distribué par Babel Technologies (<http://www.babeltech.com>), mis au point par Yves Laprie, LORIA, Nancy. Version utilisée v.1.2. (1998-1999).

⁸ Les fréquences ont été converties en quart de tons selon la formule : $S = (\log(X) - \log(220)) / (\log(2)/24)$, X étant la valeur à convertir et 220, la fréquence de référence.

2.2.1.1. Mots dont l'initiale est non voisée

Initiale non voisée	G1		G2	
	RD	RI	RD	RI
« Cloche »	10	14	5,43	14,10
« Cloche inversée »	10	8	5,43	7,69
« Montée »	6,66	2	7,60	6
« Descente »	68,33	66	71,73	64,10
« Recto-tono »	5	10	9,78	10,25

Tableau n°52 : proportion de mots monosyllabiques commençant par un son non voisé répartis selon la forme de leur courbe mélodique, exprimée en pourcentage (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

Les résultats (tableau n°52) font apparaître une répartition semblable des différents types de contours pour G1 et G2 (figures n°93 et 94) en fonction des deux types de discours, RD et RI. La majorité des contours produits sont de type descendants (à plus de 60 % pour chacun des groupes dans chaque discours), ce qui correspond aux contours produits par le locuteur natif dans ce type de mots. Aucun effet notable de groupe, ni de discours n'a été relevé et l'ensemble des enfants de notre corpus restitue correctement le contour des mots monosyllabiques commençant par un son non voisé. En français, sur les mots monosyllabiques, les locuteurs ont également tendance à produire un contour descendant. Il est donc probable que, même s'il existait un effet de filtre prosodique du français vers l'anglais, il ne pourrait pas apparaître dans la restitution des contours descendants.

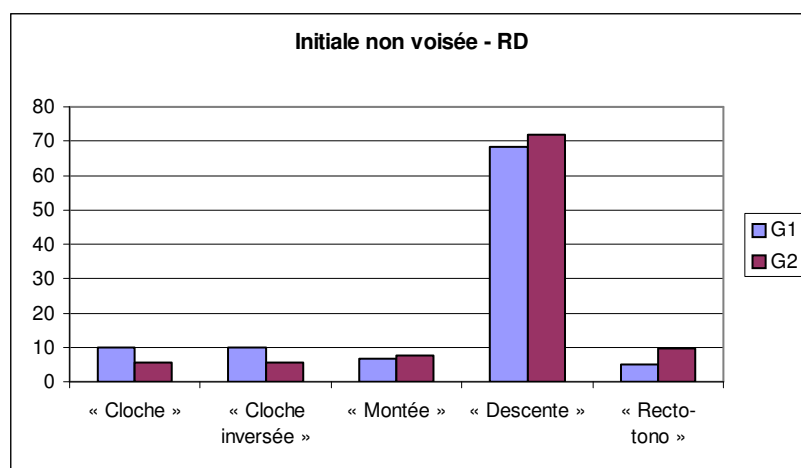


Figure n°93 : répartition en % des mots à l'initiale voisée selon la forme de leur contour mélodique en situation de répétition différée (RD) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

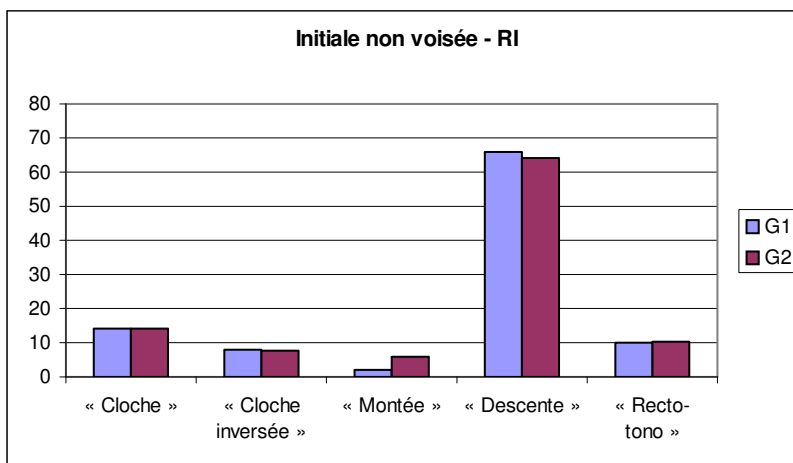


Figure n°94 : répartition en % des mots à l'initiale non voisée selon la forme de leur contour mélodique en situation de répétition immédiate (RI) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

Malgré les similarités, de légères différences entre les groupes peuvent être relevées. En RD, les enfants du groupe G2 produisent plus de contours recto-tono (9,78 %) que les enfants du groupe G1 (5 %). Par ailleurs, les enfants du groupe G2 produisent plus de contours descendants (71,73 %) et moins de contours en cloche (5,43 %) et en cloche inversée (5,43 %) que les enfants du groupe G1 (68,33 % de contours descendants ; 10 % de cloches et 10 % de cloches inversées). On peut se poser la question de savoir pourquoi les enfants musiciens produisent plus de contours en cloche sur les mots commençant par une initiale non voisée, alors qu'un contour descendant serait plus facile à réaliser et surtout, qu'il correspond mieux au modèle. Une réponse possible pourrait être que les enfants, ayant repéré les caractéristiques distinctives des contours en cloche de l'anglais, les appliquent plus souvent qu'il ne faudrait. De cette façon, ils doivent estimer que leurs énoncés sonnent « plus anglais ». Mais les différences sont cependant trop minimes pour que cette explication puisse être sûre.

En RI, les différences sont encore moins apparentes et les proportions de chaque catégorie sont remarquablement proches pour les deux groupes, à part le fait que les sujets de G2 (6 %) produisent plus de contours montants que les sujets du groupe G1 (2 %).

Si on ne relève pas de différences entre les productions des enfants des deux groupes en RI, une légère différence apparaît en RD allant en faveur de notre hypothèse initiale. Les enfants de G1 ont tendance à produire plus de contours en cloche que les enfants de G2 dans les mots à l'initiale non voisée, mais la différence n'est pas assez marquée pour vérifier notre hypothèse.

2.2.1.2. Mots dont l'initiale est voisée

<i>Initiale voisée</i>	G1		G2	
	RD	RI	RD	RI
« Cloche »	10	14	5,43	14,10
« Cloche inversée »	10	8	5,43	7,69
« Montée »	6,66	2	7,60	6
« Descente »	68,33	66	71,73	64,10
« Recto-tono »	5	10	9,78	10,25

Tableau n°53 : proportion d'énoncés de mots monosyllabiques commençant par un son voisé répartis selon la forme de leur courbe mélodique, exprimée en pourcentage (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

Comme dans les mots à initiale non voisée, la répartition des différents contours affectant les mots à initiale voisée est très proche pour les deux groupes en situation de discours différé (figure n°95 et tableau n°53). Comme précédemment, on relève une proportion de contours en cloche plus importante pour G1 (10 % de cloches et 10 % de cloches inversées) que pour G2 (5,43 % de cloches et 5,43 % de cloches inversées), ainsi qu'une proportion de contours recto-tono plus grande pour G2 (9,78 %) que pour G1 (5 %). Ces différences ne sont pas très importantes en terme de proportion, mais on les retrouve en RD dans les deux types de mots analysés.

Cependant, les résultats divergent entre les deux groupes en situation de répétition immédiate (figure n°96). Les résultats montrent que les courbes des enfants musiciens se répartissent de manière tranchée soit dans la catégorie « cloche » (46,05 %), soit dans la catégorie « descente » (42,10 %), à proportion relativement égale. Ils réussissent relativement bien à reproduire les contours en cloche caractéristiques de l'anglais, mais continuent à produire des contours descendants dans une forte proportion. Ces deux types de contours représentent 88,15 % de la totalité des énoncés produits par les enfants du groupe G1.

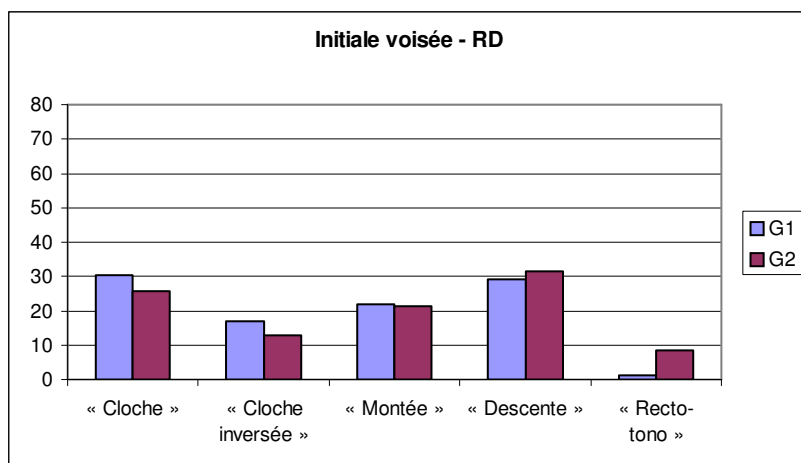


Figure n°95 : répartition en % des mots à l'initiale voisée selon la forme de leur contour mélodique en situation de répétition différée (RD) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

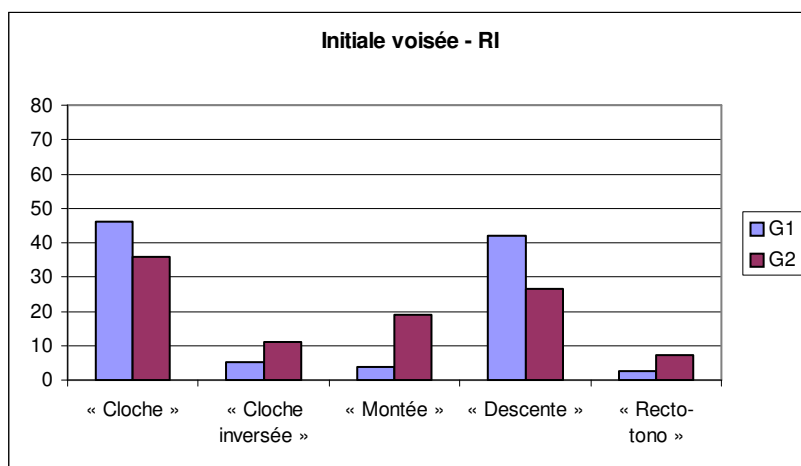


Figure n°96 : répartition en % des mots à l'initiale voisée selon la forme de leur contour mélodique en situation de répétition immédiate (RI) pour chacun des deux groupes (G1 pour « groupe musiciens », G2 pour « groupe non musiciens »).

La répartition des courbes dans les différentes catégories est beaucoup plus variable chez les enfants non musiciens. Ils produisent moins de contours en cloche (36,02 %) et beaucoup moins de contours descendants (26,47 %). En revanche, ils produisent plus de contours montants que les enfants de G1 (19,11 % contre 3,94 %) et plus de cloches inversées (11,02 % contre 5,26 %). Il est étonnant de trouver des contours montants en situation de répétition immédiate. Il nous semble que, dans notre corpus, ces contours montants correspondent généralement à une demande de confirmation, comme si l'enfant demandait à son interlocuteur « *Est-ce que c'est bien ça ?* ». La partie finale des contours en cloche inversée est montante et correspond généralement, comme la montée simple, à une demande de confirmation.

La situation d'interaction semble donc jouer un rôle assez important lors de la production des contours montants. A la lumière de ces explications, il est étrange que les enfants du groupe G2 produisent plus de contours montants en situation de répétition immédiate. En effet, en répétition différée, les enfants peuvent ne pas se rappeler un mot et solliciter l'appui du locuteur natif pour qu'il les « rassure » en quelque sorte sur le choix du mot ; la montée exprime donc une incertitude. Mais, en répétition immédiate, le modèle est donné à l'enfant qui n'a plus qu'à le répéter. Cette proportion plus grande de contours montants pourrait traduire un « manque de confiance » à reproduire des mots inconnus ayant un contour complexe (plus difficile à restituer car non présent dans la LM).

Un autre point intéressant apparaît par rapport à la proportion de contours en cloche dans les deux situations de discours. Cette proportion est en effet plus élevée en RI (46,05 % pour G1 ; 36,02 % pour G2) qu'en RD (30,48% pour G1 et 25,64 % pour G2). Nous avons vu que le contour en cloche n'existe pas en français et qu'il pourrait être difficile à reproduire en raison du crible prosodique. Il serait plus facilement restitué à la suite du locuteur natif qu'en répétition différée, situation où l'effet de filtre agirait de manière plus marquée.

Notre hypothèse initiale se vérifie en situation de répétition immédiate dans les mots à l'initiale voisée, car les enfants musiciens produisent plus de contours en cloche que les enfants non musiciens. Et s'ils ne produisent pas de contours en cloche, ils produisent des contours descendants, mais pas de contours en cloche inversée, ni de contours montants. Leur comportement est donc plus tranché que celui des enfants non musiciens, révélant peut-être une plus grande « sécurité » à reproduire des événements sonores non habituels. Par ailleurs, la proportion de contours en cloche est plus élevée en RI qu'en RD pour les deux groupes, sûrement en raison du fait que les enfants reproduisent plus facilement les contours caractéristiques de l'anglais à la suite du locuteur natif.

2.2.2. Variable « Exagération de la courbe »

Après avoir calculé l'intervalle en $\frac{1}{4}$ de tons séparant le point initial du point final de chaque inflexion pour chacun des contours des énoncés (deux inflexions pour les contours en cloche et en cloche inversée ; une inflexion pour les contours montants et descendants), nous avons déterminé l'intervalle moyen pour chaque type d'inflexion.

Les résultats sont exposés dans le tableau n°54 et les figures n°97 et n°98 pour les contours en cloche et en cloche inversée et dans le tableau n°55 et les figures n°99 et n°100 pour les contours descendants et montants (voir également tableaux détaillés, avec la totalité des contours analysés, dans le volume II, annexe n°9.2.2.).

	G1				G2			
	RD		RI		RD		RI	
Contour	A	B	A	B	A	B	A	B
Cloche	4,18	5,89	4,37	7,15	3,87	7,87	3,35	7,33
Cloche inversée	5,99	5,57	5,52	3,60	5,65	4,25	5,36	4,30

Tableau n°54 : intervalles moyens correspondant à la première inflexion initiale et à la deuxième inflexion (B) des contours en cloche et en cloche inversée, exprimés en 1/4 de tons (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

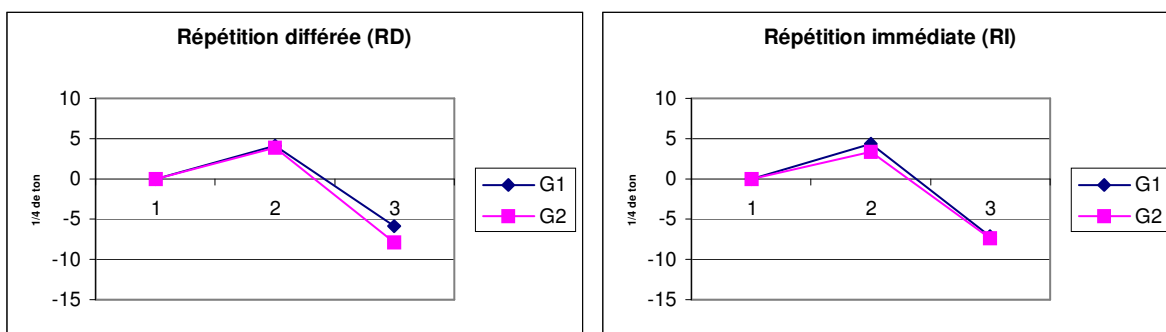


Figure n°97 : contour moyen « en cloche » en répétition différée (RD) et en répétition immédiate G1 pour « groupe musiciens », G2 pour « groupe non musiciens ».

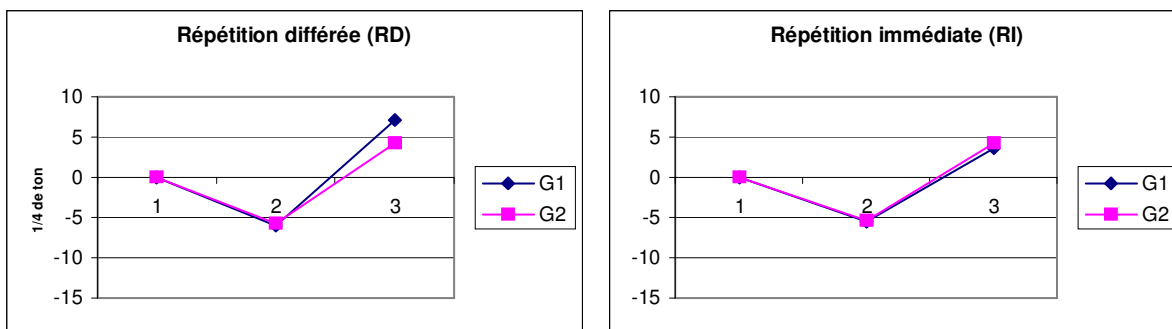


Figure n°98 : contour moyen « en cloche inversée » en répétition différée (RD) et en répétition immédiate pour G1 pour « groupe musiciens », G2 pour « groupe non musiciens ».

	G1		G2	
	RD	RI	RD	RI
Descendant	7,06	6,14	5,56	7,08
Montant	6,38	4,53	5,44	4,17

Tableau n°55 : intervalles moyens correspondant aux contours montants et descendants, exprimés en 1/4 de tons (G1 pour « groupe musiciens », G2 pour « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

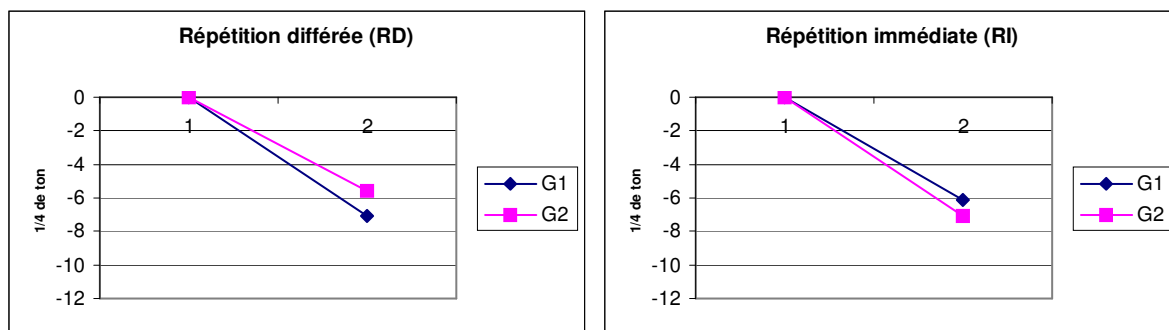


Figure n°99 : contour moyen « descendant » en répétition différée (RD) et en répétition immédiate pour G1, « groupe musiciens » et G2, « groupe non musiciens ».

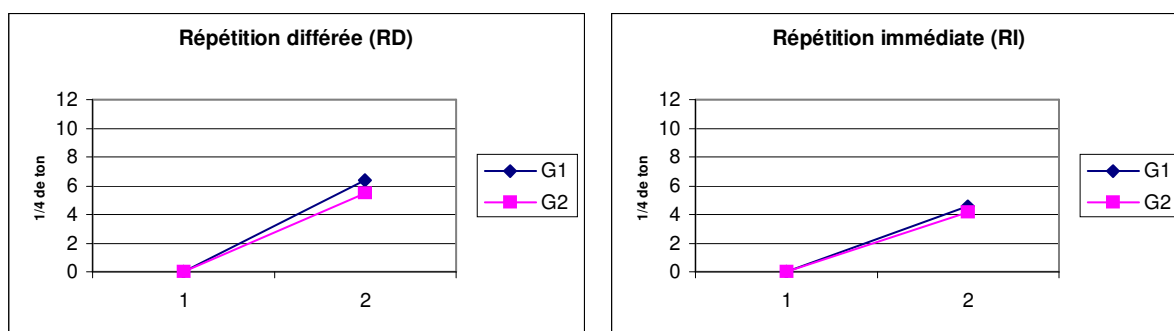


Figure n°100 : contour moyen « montant » en répétition différée (RD) et en répétition immédiate pour G1, « groupe musiciens » et G2, « groupe non musiciens ».

Une première analyse révèle que, pour l'ensemble des inflexions, les intervalles moyens les plus grands le sont beaucoup moins que ceux du locuteur natif. En effet, on relève pour ce locuteur une inflexion initiale de 8,21 ¼ de tons et une inflexion finale de 13,67 ¼ de tons pour le contour en cloche « moyen » (voir figure n°88) et une inflexion de 10,40 ¼ de tons pour le contour descendant « moyen » (voir figure n°89). Pour les enfants, le plus grand intervalle relevé est un intervalle moyen de 7,87 ¼ de tons pour la deuxième inflexion des contours en cloche (G2 en RD, voir tableau n°54). Ils ne produisent donc pas de contours plus amples que ceux du locuteur natif. Il est vrai que les moyennes lissent fortement les résultats et qu'elles font disparaître les particularités individuelles. Nous les avons donc complétées par le calcul des écarts-types pour les différents contours.

	G1				G2			
	RD		RI		RD		RI	
	A	B	A	B	A	B	A	B
Cloche	3,36	4,25	2,84	3,51	3,06	4,16	2,41	3,50
Cloche inversée	3,277	5,23	2,05	2,25	3,02	2,80	3,977	3,40

Tableau n°56 : écarts-types correspondant à la première inflexion initiale et à la deuxième inflexion (B) des contours en cloche et en cloche inversée, exprimés en ¼ de tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

	G1		G2	
	RD	RI	RD	RI
Descendant	3,40	2,97	3,776	4,27
Montant	4,40	3,55	3,30	3,59

Tableau n°57 : écarts-types correspondant aux contours montants et descendants, exprimés en ¼ de tons (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

Pour le locuteur natif, l'écart-type est de 4,95 pour l'inflexion initiale et de 6,18 pour l'inflexion finale pour les contours en cloche, et de 5,24 pour les contours descendants. Chez les enfants, les écarts-types sont beaucoup plus réduits que ceux du locuteur natif : entre 2,05 et 3,97 pour l'inflexion initiale et entre 2,25 et 4,16 pour les contours en cloche et en cloche inversée (voir tableau n°56) ; entre 2,97 et 4,27 pour les contours descendants et entre 3,59 et 4,40 pour les contours montants (voir tableau n°57). Ces écarts-types réduits révèlent que, s'il y a des contours exagérés, ils ne sont pas nombreux et que la tendance générale des groupes est à la modération de l'ampleur de la courbe mélodique. Effectivement, si on regarde les résultats détaillés figurant dans les tableaux de l'annexe n°9.1.2. (Volume II), on peut voir que, pour l'inflexion initiale des contours en cloche par exemple, l'intervalle le plus élevé est de 12,05 ¼ de tons (soit un intervalle approximatif de quarte augmentée) et qu'il n'y a que 7 productions sur un total de 63 qui dépassent un intervalle de 10 ¼ de tons (soit un intervalle d'une quarte juste), tous types de discours et groupes confondus, alors que locuteur natif dépasse 8 fois l'intervalle de 10 ¼ de tons sur un total de 37 productions et qu'il va jusqu'à produire un intervalle de 21 ¼ de tons (soit un intervalle approximatif d'une septième mineure). Le comportement d'exagération ne caractérise aucun des deux groupes, il ne caractérise donc pas le « groupe musiciens ». Il semble plutôt relever de stratégies employées par quelques individus appartenant aussi bien à G1 qu'à G2. Par ailleurs, sur les représentations graphiques exposées dans les figures n°97, 98, 99 et 100, on peut remarquer que les différences entre les groupes sont minimales pour l'ensemble des résultats et que les « contours moyens » se superposent presque parfaitement, quel que soit le type de contour et quel que soit le type de discours.

On peut toutefois relever quelques différences, même si celles-ci sont légères. Par exemple, la deuxième inflexion du contour en cloche en RD est légèrement plus ample pour G2 que pour G1, mais seulement de 1,98 ¼ de tons (soit approximativement un demi-ton). A l'inverse, la deuxième inflexion du contour en cloche inversée en RD est légèrement plus ample pour G1 que pour G2, mais seulement de 1,32 ¼ de tons (voir tableau n°54). Les différences sont légèrement plus importantes pour les contours montants et descendants.

Ainsi, en ce qui concerne les contours descendants, en RD, l'intervalle moyen de G1 est plus grand d'1,5 ¼ de tons que celui du groupe G2 (voir tableau n°55). Nous ne continuerons pas plus avant, ces différences n'étant pas pertinentes. On peut faire la même remarque par rapport aux deux situations de discours.

En ce qui concerne les écarts-types, on peut relever des différences plus intéressantes entre les groupes. Ainsi, pour les enfants du groupe G1, l'écart-type est plus élevé en situation de RD qu'en situation de RI et ce, dans les quatre types de contours (tableau n°56 et n°57). Initialement, nous avons posé l'hypothèse que les musiciens exagéreraient plus souvent les contours en RD qu'en RI. Les écarts-types révèlent qu'il y a plus de contours amples en situation de RD, même s'ils ne sont pas assez nombreux pour affecter la configuration du contour moyen. Pour les enfants du groupe G2, les résultats sont moins cohérents et il n'y a pas d'uniformité pour les contours en cloche et en cloche inversée. En revanche, pour les contours montants et descendants, la situation est inversée par rapport à G1 : les écarts-types sont plus élevés en RI qu'en RD, ce qui veut dire qu'on relève plus de productions amples chez les enfants du groupe G2 en répétition immédiate pour ces deux catégories de contour.

Notre hypothèse initiale n'est pas vérifiée et l'amplitude de la courbe mélodique ne semble pas affectée par la variable indépendante « musique », mais il semble cependant qu'en situation de répétition différée, les enfants musiciens produisent des contours plus variables qu'en situation de répétition immédiate. Il serait donc particulièrement intéressant de procéder à une étude qualitative des productions des enfants, pour déterminer si certains d'entre eux manifestent des stratégies d'exagération des éléments prosodiques. Mais une telle étude ne concernerait que quelques enfants et la tendance générale qui caractérise notre échantillon est à la production de contours peu amples sans distinction de groupe. Même en situation de discours immédiat, les enfants ne reproduisent pas l'amplitude des contours caractéristiques du locuteur natif. Pourtant, ils perçoivent correctement les contours prosodiques, comme le suggèrent les résultats obtenus aux tests de perception dans l'étude n°1 (chapitre VII). En réalité, la stratégie d'exagération pourrait être plutôt liée à la personnalité de l'apprenant. En effet, nous savons que les apprenants francophones adultes n'osent pas exagérer les contours en anglais, par peur du « ridicule », car la production de contours très amples n'entre pas dans leurs habitudes prosodiques maternelles. Ils hésitent à jouer avec les paramètres prosodiques devant les autres élèves dans une classe de langue par exemple, alors qu'ils le font plus facilement individuellement.

De la même façon, le fait de « mettre l'accent » sur une expression anglaise à l'intérieur d'un discours en français est généralement mal toléré, jugé comme prétentieux. Il semble donc que les « préjugés linguistiques » exercent une influence considérable sur le comportement prosodique des locuteurs francophones et que ces préjugés semblent déjà affecter les apprenants âgés de 7 ans. Il serait intéressant d'étudier ce phénomène avec des apprenants plus jeunes afin de déterminer si cette « timidité prosodique » se renforce avec l'âge.

2.2.3. Variable « Localisation de l'accent »

2.2.3.1. Mots bisyllabiques

La proportion d'accents bien localisés, c'est-à-dire placés sur la première syllabe (A), est plus élevée chez les enfants du groupe G1 en RD (73,17 % en RD, voir tableau n°58) que chez les enfants du groupe G2 (63,46 % en RD). Pour RI, cette différence se réduit et la proportion des productions correctes avoisine les 90 % pour les deux groupes (95,23 % pour G1 et 91,25 % pour G2). Les enfants montrent donc d'excellentes capacités, qu'ils soient musiciens ou non, à restituer la place de l'accent en situation de répétition immédiate. En revanche, en situation de répétition différée, comme nous l'avions supposé au départ, la tâche paraît être beaucoup plus difficile, sûrement à cause de la mise en jeu de la mémoire à long terme. En répétition immédiate, l'enfant peut solliciter sa mémoire à court terme et oublier le mot quelques instants après l'avoir prononcé. Les effets de l'apprentissage musical semblent donc se manifester plutôt dans les tâches faisant appel à la mémoire à long terme qu'à la mémoire à court terme. Notre hypothèse initiale se vérifie donc uniquement en situation de RD, situation où l'accent est mieux localisé dans les productions des enfants musiciens que dans celles des enfants non musiciens.

Bisyllabiques	G1		G2	
	RD	RI	RD	RI
A	30	40	33	73
B	11	2	19	7
% A	73,17	95,23	63,46	91,25
% B	26,82	4,76	36,53	8,75

Tableau n°58 : nombre et proportion (exprimée en pourcentage) de mots bisyllabiques classés en fonction de la localisation de l'accent (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

2.2.3.2. Mots trisyllabiques

Dans les mots trisyllabiques de notre corpus, l'accent est localisé sur la deuxième syllabe. Les résultats sont assez contrastés selon les deux groupes (voir tableau n°59), mais nous prendrons cependant beaucoup de précautions en interprétant ces résultats, le nombre d'énoncés étant trop réduit pour en tirer des tendances généralisables. Ainsi en RD, nous n'avons pu relever que 4 énoncés pour G1. Nous n'en tiendrons donc pas compte. Pour le reste et à l'inverse des résultats concernant les mots bisyllabiques, les enfants du groupe G2 se montrent très bons pour restituer l'accent en situation de RI (82,60 % des mots correctement accentués contre 65 % pour G1). En situation de RD, les résultats sont nettement moins bons pour les deux groupes, plus mauvais là-aussi pour G1 (38,88 % des mots correctement accentués) que pour G2 (50 %). Lorsqu'ils se trompent, les enfants de G2 ont tendance à accentuer la première syllabe (50 % de mots accentués sur la première syllabe) et rarement la troisième syllabe (11,11 % de mots accentués sur la troisième syllabe). On constate la même chose pour le groupe G1 en situation de RI (35 % de mots accentués sur la première syllabe), ceux-ci ne produisant en revanche jamais d'accent en finale (0 % de mots accentués sur la troisième syllabe). Ces derniers résultats sont intéressants car ils révèlent que la position finale de l'accent en français ne semble pas influencer les résultats des enfants lorsqu'ils reproduisent des mots en anglais. Bien au contraire, leurs erreurs de localisation concernent la première syllabe. Or, nous savons que l'anglais connaît une tendance à l'accentuation en initiale de mot (cf. chapitre V, 2.1.1.1.). Ainsi, même s'ils se trompent, leurs énoncés « sonnent » quand même anglais. Cependant ces arguments ne peuvent rester qu'à l'état d'hypothèses, le nombre d'énoncés étant trop réduit (65 énoncés au total).

Trisyllabiques	G1		G2	
	RD	RI	RD	RI
A	1	7	9	2
B	2	13	7	19
C	1	0	2	2
% A	25	35	50	8,69
% B	50	65	38,88	82,60
% C	25	0	11,11	8,69

Tableau n°59 : nombre et proportion (exprimée en pourcentage) de mots trisyllabiques classés en fonction de la localisation de l'accent (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RI pour « répétition immédiate » et RD pour « répétition différée »).

On peut quand même conclure que pour les mots bi et trisyllabiques, le taux de réussite est beaucoup plus élevé en situation de RI qu'en situation de RD. La place de l'accent serait donc plus difficile à restituer en situation de discours différé.

Quoiqu'il en soit, notre hypothèse initiale ne se vérifie pas pour les mots trisyllabiques car les enfants du groupe G2 localisent mieux l'accent que les enfants du groupe G1.

3. Etude n°4 « Production des noyaux vocaliques »

3.1. Plan expérimental

3.1.1. Hypothèse

Nous posons l'hypothèse particulière « Contrastes vocaliques » qu'une oreille entraînée par la pratique musicale induit une meilleure production des contrastes vocaliques d'une langue étrangère.

3.1.2. Variables indépendantes

3.1.2.1. La musique : deux modalités

La variable indépendante « musique » est composée de deux modalités : le groupe d'enfants « musiciens » G1 et le groupe d'enfants « non musiciens » G2 (voir 2.1.2.1.).

3.1.2.2. Le discours : deux modalités

La variable indépendante « discours » est composée de deux modalités : discours immédiat RI et discours différé RD (voir 2.1.2.2.).

3.1.3. Variables dépendantes

3.1.3.1. Timbre du noyau vocalique

Nous l'avons vu, l'apprenant tente de se rapprocher du système cible. Cette tâche est entravée par les problèmes de crible phonologique. Dans l'étude n°2 sur la perception des contrastes vocaliques, nous avons vu que les enfants qui avaient une oreille entraînée par la pratique musicale se montraient moins sensibles au phénomène de crible phonologique que les enfants qui n'avaient pas d'oreille entraînée (chapitre VII). Est-ce que cette capacité à percevoir les contrastes vocaliques non natifs de l'anglais va également engendrer une meilleure production de ces contrastes, en vertu des liens étroits qui existent entre perception et production ?

Pour le savoir, nous étudierons la qualité des différents contrastes vocaliques produits par les enfants dans notre corpus et nous les comparerons aux voyelles et aux diphtongues produites par le locuteur natif de façon à déterminer s'il existe une différence entre les deux groupes et dans les deux situations de discours (l'effet de filtre pouvant être plus fort en RD qu'en RI).

Dans le chapitre V (3.1.3.2.3.), nous avons choisi de caractériser le timbre d'une voyelle par les rapports entre ses trois premiers formants. Le fait de travailler avec des rapports et non pas avec des fréquences absolues résout le problème de normalisation. En effet, à cause de la variabilité inter locuteurs, il est très difficile de comparer une voix d'homme, celle de notre locuteur natif, à des voix d'enfants. Cette variabilité est principalement due à une différence de taille entre le tractus vocal de l'adulte, et celui, plus court, de l'enfant. Cependant, l'utilisation des rapports entre la fréquence centrale des trois premiers formants peut réduire et presque éliminer les différences entre locuteurs (cf. travaux exposés dans le chapitre V). Ce mode de calcul nous permet de comparer les productions du locuteur natif à celles des enfants de notre échantillon.

La figure n°101 nous donne la « charte des rapports de formants » du locuteur natif (pour la description de cette représentation et sa lecture, se reporter au Volume II, annexe n°3.2.) et le tableau n°60, les valeurs des trois premiers formants et les intervalles F2-F1 et F3-F2 pour chaque voyelle représentée sur la charte.

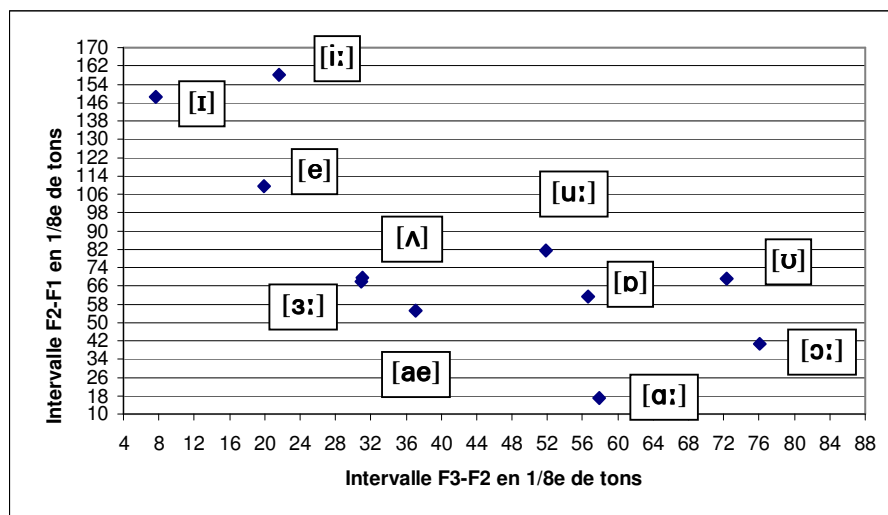


Figure n°101 : charte des rapports de formants du locuteur anglophone natif.

	F1	F2	F3	F1-F3	F2-F3
[ʊ]	343	931	2646	69,14	72,33
[æ]	686	1519	2597	55,04	37,13
[ɑ:]	923	1188	2722	17,01	57,88
[ʌ]	588	1568	2450	67,92	30,9
[ɜ:]	539	1470	2303	69,47	31,08
[ɪ]	343	2940	3283	148,7	7,64
[ɒ]	497	1208	2737	61,5	56,63
[i:]	245	2401	3283	158,05	21,66
[e]	462	2239	2986	109,29	19,93
[ɔ:]	490	882	2646	40,7	76,07
[u:]	392	1274	2695	81,62	51,88

Tableau n°60 : valeurs des trois premiers formants en Hz et des rapports F2-F1 et F3-F2 en 1/8^e de tons pour les 11 voyelles prononcées par le locuteur anglophone natif.

D'après les résultats de l'étude n°2 (chapitre VII), nous posons comme hypothèse opérationnelle que la capacité des enfants musiciens à mieux discriminer le timbre des contrastes vocaliques de l'anglais va engendrer une meilleure reproduction des contrastes vocaliques de l'anglais, c'est-à-dire des voyelles et des diphtongues, en raison des liens qui existent entre la perception et la production. La variable « timbre du noyau vocalique » comprend deux modalités : la modalité « voyelles » et la modalité « diphtongues ».

3.1.3.2. Durée du noyau vocalique

Nous avons posé comme hypothèse dans le chapitre 2.1.3.2. que les enfants musiciens, habitués à manipuler les éléments musicaux, pouvaient reporter cette stratégie dans la restitution de la prosodie d'une langue étrangère, en exagérant l'amplitude des contours mélodiques. Or, s'ils exagèrent ces contours, ils vont passer plus de temps à les réaliser et comme le noyau vocalique porte le contour mélodique, il devrait être également allongé. Non seulement, la durée absolue (désormais DA) du noyau sera plus élevée, mais sa durée relative (désormais DR) au sein de la syllabe sera plus grande.

Nous avons vu dans le chapitre V qu'il était possible de dresser une typologie des noyaux vocaliques de l'anglais en fonction de leur durée intrinsèque (3.1.3.2.1.). Cette typologie permet en particulier de différencier les voyelles brèves des voyelles longues. Quant aux diphtongues, elles sont encore plus longues en raison d'une trajectoire articulatoire beaucoup plus complexe (chapitre V, 3.2.2.1.1.).

La variable « durée du noyau vocalique » comprend donc trois modalités, la modalité « voyelles brèves », la modalité « voyelles longues » et la modalité « diphtongues » et nous posons comme hypothèse opérationnelle que la durée absolue et la durée relative des voyelles longues et des diphtongues seront plus élevées chez les enfants musiciens et spécialement en situation de répétition différée (RD). Il est probable que la différence entre les groupes soit moins marquée pour les voyelles brèves, mais celles-ci devraient cependant être mieux restituées par les enfants musiciens, c'est-à-dire que leur durée absolue et relative sera plus proche de celle des voyelles brèves prononcées par un locuteur natif. Le plan expérimental est exposé dans le tableau n°61.

	G1		G2	
	RD	RI	RD	RI
Voyelles brèves				
Voyelles longues				
Diphtongues				

*Tableau n°61 : plan expérimental pour la variable « durée du noyau vocalique »
(tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens »,
RD, répétition différée, RI, répétition immédiate).*

3.1.4. Mesures acoustiques

3.1.4.1. Mesures de timbre

Pour chaque voyelle de notre corpus, la fréquence des trois premiers formants F1, F2 et F3 a été relevée grâce à une coupe spectrale LPC et une coupe cepstrale éditée avec le logiciel Winsnoori (voir note n°7). Les coupes ont été appliquées au centre de la partie stable des voyelles.

Pour chaque diphtongue, la fréquence des trois premiers formants a été relevée sur chacune des deux cibles vocaliques. Les coupes ont été appliquées au centre de la partie stable de la cible, ou s'il n'y a pas de partie stable sur le point d'inflexion de la trajectoire des formants.

Les valeurs de F1, F2 et F3 ont ensuite été converties en huitièmes de tons selon la formule :

$$S = (\log(X) - \log(220)) / (\log(2) / 48)$$

X étant la valeur en Hz à convertir et 220, la fréquence de référence (fréquence choisie car elle est inférieure à la fréquence la plus basse que l'on puisse trouver pour F1 ; les chiffres obtenus après conversion seront donc toujours positifs).

Une fois la conversion faite, les intervalles en huitièmes de tons entre F1 et F2 (F2-F1) et entre F2 et F3 (F3-F2) ont été calculés.

Les 7 voyelles brèves et les 5 voyelles longues de l'anglais ont été analysées au sein de 35 mots de notre corpus (mots mono, bi et trisyllabiques) et ce, pour les productions des 25 enfants qui composent notre échantillon. 1225 prises de mesures ont ainsi été réalisées (au total, 49 occurrences différentes réparties dans 35 mots pour 25 enfants). La répartition des mots du corpus par voyelle étudiée est donnée dans le Volume II (annexe n°8.2.3.). Pour les noyaux vocaliques complexes, 5 diphtongues ont été analysées au sein de 10 mots de notre corpus (mots mono, bi et trisyllabiques). 275 prises de mesures ont été réalisées (au total 11 occurrences différentes réparties dans 10 mots pour 25 enfants).

Seuil de perception

Nous avons choisi les huitièmes de tons en fonction des seuils de perception donnés par la littérature. Il existe peu de travaux à ce sujet et ces seuils concernent souvent la variation d'un seul formant et non la variation simultanée de plusieurs formants. Ce seuil varie de 3 à 5 % de la fréquence du formant, mais une variation aussi large que 13 % a été démontrée par Nakagawa et al. (1982). Mermelstein (1978) a étudié la variation simultanée de la fréquence des formants à partir d'un seul point référence. Il relève un seuil de 3 % pour F2. Kewley-Port (1990) rapporte un seuil plus réduit, de 1 à 2 % pour la variation d'un seul formant et Hawks (1994), un seuil de 1,9 %. A titre d'exemple, par rapport à une valeur de 588 Hz correspondant au premier formant de la voyelle [ʌ], une variation de fréquence de 1 % représente une variation de $0,68 \frac{1}{8}^{\circ}$ de tons (593,88 Hz), une variation de fréquence de 2% représente une variation de $1,35 \frac{1}{8}^{\circ}$ de tons (599,76 Hz) et une variation de fréquence de 3 % représente une variation de $2,04 \frac{1}{8}^{\circ}$ de tons (605,64 Hz). Ces variations très faibles justifient le recours aux huitièmes de tons pour le calcul des intervalles entre les trois premiers formants. Mais ces auteurs utilisent une méthodologie qui place l'auditeur dans des conditions les plus favorables pour une tâche de discrimination. Les conditions d'écoute en parole naturelle sont plus dégradées, car soumises aux perturbations dues à l'environnement naturel.

De l'avis même de Hawks (1994 : 1080),

« discriminability of vowels in natural speech communication modes should be based on larger perceptual units ».

Par ailleurs, Hawks montre que le seuil de perception du changement simultané de la fréquence de plusieurs formants est plus petit que pour un seul formant, mais seulement pour certains patterns. La perception de la variation simultanée de plusieurs formants semble plus naturelle que la variation d'un seul formant :

« in natural speech, however, we would not anticipate shifting of only a single formant, but rather, the simultaneous shifting of multiple formants » (Hawks, 1994 : 1074).

Enfin, la discrimination de variations parallèles est significativement meilleure que pour des variations opposées et que pour la variation d'un seul formant (Hawks, 1994). Pour les mouvements parallèles, l'information provenant des formants séparés se combine pour améliorer la performance (Mermelstein, 1978). Il est donc probable que le seuil de perception de la variation des trois premiers formants au cours de la partie stable soit très réduit.

3.1.4.2. Mesures de durée

La durée absolue (DA exprimée en ms) de chaque noyau vocalique et de chaque syllabe a été relevée dans les 30 mots monosyllabiques qui composent le corpus (voir liste des mots dans le Volume II, annexe n°8.2.2.) et ce, pour les productions des 27 enfants de notre échantillon. 1620 prises de mesure ont ainsi été réalisées, dont 810 relevés de durée du noyau et 810 relevés de durée syllabique. Les étiquettes qui délimitent chacun des segments étudiés (vocaliques et syllabiques) ont été posées en couplant la représentation oscillographique à la représentation spectrographique à bandes larges du logiciel Winsnoori. La durée relative de la voyelle a été obtenue en calculant le pourcentage de la durée de la voyelle en fonction de la durée totale de la syllabe (DR exprimée en %).

3.2. Résultats

3.2.1. Timbre du noyau vocalique

Pour les productions des enfants, les résultats seront présentés par son, mais également par mot, car les contrastes vocaliques apparaissent au sein de contextes phonétiques très différents, mais surtout, par ce que chaque mot a un statut différent dans l'apprentissage.

Ainsi, des mots qui ressemblent aux mots français comme « *orange* » ou « *banana* » auront tendance à être prononcés à la française, c'est-à-dire avec un [ɔ] à l'initiale et la voyelle nasale [ɑ̃] en syllabe finale. Des mots très courants comme « *cat* » ou « *dog* » sont déjà connus des enfants et très souvent prononcés à la française dans leur environnement. En revanche, les mots inconnus posent moins de problèmes d'interférences. Les moyennes par son gommeraient les particularités qui nous intéressent dans le « parcours d'apprentissage » des enfants.

3.2.1.1. Voyelles

3.2.1.1.1. Voyelles brèves

[ɪ] : en syllabe accentuée, cette voyelle est très fermée (rapport F2-F1 pouvant atteindre 150 1/8^e de tons chez le locuteur natif (« *pink*, *eleven* »). Elle est plus fermée que la voyelle française de timbre correspondant [i] (120 1/8^e de tons). En syllabe non accentuée, la voyelle est beaucoup plus ouverte (F2-F1 = 100 1/8^e de tons), comme dans « *orange* » et « *rabbit* » (même aperture que [e]).

Pour les enfants, on ne relève pas de différences entre les groupes en RD ; ils produisent une voyelle équivalente [i] (120 1/8^e de tons). Il y a donc **assimilation** sans distinction de groupe. En revanche, en RI, G2 réalise une voyelle légèrement plus ouverte (113 1/8^e de tons) ; on relève également beaucoup de variabilité dans la réalisation de l'intervalle F2-F1 (σ = 21,36 pour G1 et σ = 21,46 pour G2 en RI ; σ = 19,07 pour G1 et σ = 22,48 pour G2 en RD). Cette variabilité reflète la différence d'aperture qui affecte les productions des enfants pour les deux groupes et dans les deux situations de discours.

Les résultats détaillés par mot font apparaître que les enfants, sans distinction de groupe ni de discours, produisent un [i] dans les mots « *six* », « *pink* » et « *fish* ». On relève une grande variabilité dans l'intervalle F2-F1 dans « *six* » (σ = 22,28 pour G1 et σ = 23,22 pour G2) et dans « *pink* » pour G1 (σ = 20,27). Dans « *teddy* », en RI, ils produisent une voyelle également proche de [i] quoique légèrement plus fermée pour G1 (136 1/8^e de tons). Si cette voyelle est aussi fermée dans une syllabe accentuée, c'est qu'ils ne reproduisent pas correctement la rythmique anglaise de ce mot. Dans « *baby* » en RI, G1 produit une voyelle équivalente à [i], mais G2, une voyelle plus légèrement plus ouverte (109 1/8^e de tons).

On note beaucoup plus de variabilité dans la réalisation de cette voyelle pour l'intervalle F2-F1 ($\sigma = 21,03$ pour G1 et $\sigma = 29,36$ pour G2) et pour l'intervalle F3-F2 pour G2 ($\sigma = 9,15$ pour G1 et $\sigma = 22,14$ pour G2). Dans « rabbit », en RI, la voyelle est légèrement plus ouverte ($110 \frac{1}{8}^\circ$ de tons pour G1 et $114 \frac{1}{8}^\circ$ de tons pour G2), plus proche de la production du locuteur natif. On relève pour les deux groupes une grande variabilité de l'intervalle F2-F1 ($\sigma = 25,99$ pour G1 et $\sigma = 21,57$ pour G2). Dans le mot « eleven », la voyelle est plus fermée qu'un [i] français pour les deux groupes en RD ($130 \frac{1}{8}^\circ$ de tons pour G1 et $135 \frac{1}{8}^\circ$ de tons pour G2), qu'en RI ($106 \frac{1}{8}^\circ$ de tons pour G1 et $99 \frac{1}{8}^\circ$ de tons pour G2) où la voyelle se rapproche du [e] français (**assimilation**). Pour G2 en RD, on relève par ailleurs une grande variabilité dans l'intervalle F2-F1 ($\sigma = 14,19$ pour G1 et $\sigma = 24,03$ pour G2). Enfin, dans le mot « orange », les enfants des deux groupes produisent une voyelle de timbre similaire à la voyelle française [i] ($116-21 \frac{1}{8}^\circ$ de tons pour G1 et $116-23 \frac{1}{8}^\circ$ de tons pour G2). En revanche, on note une différence entre les groupes en RI, où, tandis que les enfants de G1 produisent une voyelle de même timbre qu'en RD ($126-23 \frac{1}{8}^\circ$ de tons), les enfants de G2 produisent une voyelle moins fermée, mais surtout plus postérieure ($87-44 \frac{1}{8}^\circ$ de tons). La variabilité qui affecte l'intervalle F2-F1 est plus élevée en RD ($\sigma = 25,39$ pour G1 et $\sigma = 23,12$ pour G2) qu'en RI ($\sigma = 18,97$ pour G1 et $\sigma = 18,18$ pour G2).

Ces résultats montrent que la voyelle anglaise [ɪ] est le plus souvent assimilée à la voyelle française [i] sans distinction de groupe, ni de discours et ce, même dans les syllabes non accentuées où la voyelle devrait être plus centralisée.

[ʊ] : au sein de notre corpus, il n'y a qu'une occurrence de cette voyelle dans le mot « book ». Le LN produit une voyelle très postérieure (F3-F2 : $72 \frac{1}{8}^\circ$ de tons) et une aperture aussi marquée que pour la voyelle française [ɔ].

Dans les productions des enfants, on relève beaucoup plus de variabilité pour les deux intervalles en situation de RD. Dans cette situation, G1 réussit à produire un [ʊ] (F2-F1 : $64 \frac{1}{8}^\circ$ de tons et F3-F2 : $71 \frac{1}{8}^\circ$ de tons), tandis que G2 produit une voyelle légèrement plus antérieure (F3-F2 : $64 \frac{1}{8}^\circ$ de tons) et plus ouverte (F2-F1 : $61 \frac{1}{8}^\circ$ de tons), mais qui reste proche malgré tout de [ʊ]. En situation de RI, G1 produit un son de timbre similaire. En revanche, G2 produit une voyelle aussi postérieure que [ʊ] (F3-F2 : $70 \frac{1}{8}^\circ$ de tons), mais plus ouverte (F2-F1 : $54 \frac{1}{8}^\circ$ de tons), qui la rapproche de la voyelle française [o].

Par ailleurs, on relève beaucoup de variabilité dans l'intervalle F2-F1 pour G2 ($\sigma = 14,62$ pour G1 et $\sigma = 22,26$ pour G2), mais ce qui est plus frappant, une grande variabilité dans l'intervalle F3-F2 pour les deux groupes, spécialement pour G1 ($\sigma = 23,80$ pour G1 et $\sigma = 18,03$ pour G2). Cette variabilité, qui n'est pas habituelle pour l'intervalle F3-F2 sur l'ensemble des noyaux vocaliques de notre corpus, montre que le lieu d'articulation de la voyelle [ʊ] varie beaucoup dans les productions des enfants.

Ces résultats montrent que les enfants ne produisent pas l'assimilation avec la voyelle française [u], et que, s'il y a une différence de timbre, elle va dans le sens d'une exagération de type [bok] qui « sonne anglais ».

[ʌ] : dans le mot « duck », le locuteur natif produit une voyelle centrale (F2-F1 : $67 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $30 \frac{1}{8}^{\circ}$ de tons).

Les moyennes générales des enfants indiquent qu'en RD, G1 et G2 produisent une voyelle de timbre proche (pas de crible) et en RI, une voyelle légèrement plus postérieure (avec un intervalle F3-F2 plus élevé, $49,49 \frac{1}{8}^{\circ}$ de tons pour G1 et $44,43 \frac{1}{8}^{\circ}$ de tons pour G2 ; timbre proche de la voyelle française [ɔ]).

Dans le détail des trois mots qui contiennent cette voyelle, la voyelle de « one » est correctement reproduite pour G1 et G2 en RD (pas d'assimilation). En revanche, dans le mot « duck », en situation de RI, G1 produit une voyelle un peu plus fermée que la voyelle française [ɔ] (F2-F1 : $71 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $43 \frac{1}{8}^{\circ}$ de tons) et G2 une voyelle proche de la voyelle française [ɔ] (F2-F1 : $59 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $45 \frac{1}{8}^{\circ}$ de tons). Dans l'**extrait sonore n°64**, l'enfant n°6 produit ce type d'assimilation c'est-à-dire [dɔk] en situation de répétition immédiate, alors que l'enfant n°4 produit le mot correctement, c'est-à-dire [dʌk]. La prononciation du mot « umbrella » est un peu particulière pour le locuteur natif puisque celui-ci produit une voyelle plus ouverte et plus postérieure que le [ʌ] produit dans « duck » (F2-F1 : $40 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $49 \frac{1}{8}^{\circ}$ de tons), dont le timbre est proche de [ɑ] français. En RI, les enfants de G1 produisent une voyelle plus postérieure que celle du LN, dont le timbre est proche de la voyelle anglaise [ɒ] (F2-F1 : $56 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $56 \frac{1}{8}^{\circ}$ de tons).

Les enfants de G2 produisent une voyelle proche de [ɔ], peut-être [ɔ̃] (F2-F1 : 66 1/8° de tons et F3-F2 : 42 1/8° de tons), à cause du mot français « ombrelle » (**extrait sonore n°65** en RI, pas de correction, finale de mot prononcée « à la française » : « LN : *umbrella* – E1 et E2 : *umbrella* – LN : *umbrella* – E1 : *umbrelle* ; c'est une ombrelle ou un parapluie ? LN : *umbrella*, repeat – E2 : *umbrelle* »). Pour les deux groupes, la variabilité de l'intervalle F2-F1 est élevée, en particulier pour G1 ($\sigma = 25,02$ pour G1 et $\sigma = 18,63$ pour G2).

[ɒ] : dans les mots « orange » et « doll », le locuteur natif produit la voyelle [ɒ] de même aperture, mais plus antérieure que la voyelle française [o] (F2-F1 : 48 1/8° de tons et F3-F2 : 60 1/8° de tons). Dans le mot « dog », la voyelle est légèrement plus fermée, mais localisée au même endroit sur la charte des rapports de formants (F2-F1 : 61 1/8° de tons et F3-F2 : 56 1/8° de tons).

En RD, les enfants des deux groupes produisent des voyelles de timbre proche, voire similaire à [ɒ]. En RI, G2 produit le même type de voyelle et G1, une voyelle légèrement plus antérieure (F2-F1 : 69 1/8° de tons et F3-F2 : 48 1/8° de tons). On ne relève donc pas d'effet de filtre au profit de la voyelle française [ɔ].

Dans le détail des mots, la voyelle initiale [ɒ] de « orange » est correctement reproduite par les deux groupes en RD et en RI, par G2. En revanche, on relève un effet de filtre en faveur de la voyelle française [ɔ] pour G1. Pour les deux groupes, on relève une grande variabilité de l'intervalle F2-F1 ($\sigma = 24,98$ pour G1 et $\sigma = 18,34$ pour G2). Dans le mot « doll », la voyelle est correctement reproduite par les deux groupes en RI, ainsi qu'en RD pour G1. C'est cette fois G2 en RD qui produit une voyelle encore plus fermée que la voyelle française [ɔ] (F2-F1 : 76 1/8° de tons et F3-F2 : 39 1/8° de tons). On relève également une grande variabilité de l'intervalle F2-F1 pour les deux groupes ($\sigma = 21,04$ pour G1 et $\sigma = 22,11$ pour G2). Dans le mot « dog », l'effet de filtre est massif en RD et en RI pour les deux groupes (plus fermé que la voyelle [ɔ] pour G1 et légèrement plus postérieure que [ɔ] pour G2). Alors que le mot « doll » n'est pas courant et pas connu des enfants, le mot « dog » est très répandu. L'assimilation doit provenir de la prononciation « à la française » de ce mot.

[æ] : dans « rabbit » et « apple », la voyelle [æ] produite par le locuteur natif est au centre de l'axe F3-F2 (F2-F1 : 55 1/8° de tons et F3-F2 : 37 1/8° de tons pour « rabbit » ; F2-F1 : 38 1/8° de tons et F3-F2 : 36 1/8° de tons pour apple », légèrement plus ouverte). La voyelle de « parrot » est plus antérieure et se rapproche du [a] français dans « parrot ».

En moyenne, les enfants des deux groupes et dans les deux types de discours produisent la même voyelle que le locuteur natif dans « rabbit » et « apple ».

Dans le détail des mots, on ne relève pas de différence entre les productions des deux groupes pour « black » en RD, « cat » en RD, « rabbit » en RI. L'**extrait sonore n°66** nous donne un exemple de production du mot « cat » en répétition différée par deux enfants du groupe G2 (enfants n°16 et n°23). La même voyelle est produite par G1 en RD dans « apple ». En revanche, dans le mot « apple », la voyelle est beaucoup plus ouverte, proche de la voyelle française [ɑ] pour les deux groupes en situation de RI (F2-F1 37 1/8° de tons et F3-F2 : 44 1/8° de tons pour G1 ; F2-F1 : 45 1/8° de tons et F3-F2 : 43 1/8° de tons pour G2). Pour « parrot », on relève une différence entre les groupes : tandis que G1 produit une voyelle proche de [æ], G2 produit une voyelle proche de la voyelle française [ɑ].

[e] : on relève des différences de timbre pour cette voyelle chez le locuteur natif ; alors qu'elle est très fermée et très antérieure dans un mot comme « red » (F2-F1 109 1/8° de tons et F3-F2 : 19 1/8° de tons), elle est plus ouverte dans « eleven » (F2-F1 92 1/8° de tons et F3-F2 : 25 1/8° de tons, entre les deux voyelles françaises [e] et [ɛ]), elle est proche du [ɛ] français dans le mot « umbrella ».

Dans les productions des enfants, le timbre est similaire pour les deux groupes et dans les deux situations de discours. Le timbre de cette voyelle se situe entre les deux voyelles françaises [e] et [ɛ], comme la voyelle produite dans « eleven » par le locuteur natif.

Dans le détail des mots, on retrouve cette voyelle sans distinction de groupe ni de discours dans « ten », « yellow », « seven », avec peu de variabilité. On la retrouve également dans « red » en RD pour les deux groupes, avec une plus forte variabilité dans l'intervalle F2-F1 (σ =17,36 pour G1 et σ =25,23 pour G2). En revanche, le groupe G2 produit une voyelle proche de [ɛ] en RI ; il y a donc **assimilation**.

Dans le mot « twelve », la voyelle est proche de celle produite par le locuteur natif dans le même mot, c'est-à-dire plus postérieure et aussi ouverte que [ɛ]. On relève une forte variabilité dans l'intervalle F2-F1 pour G2 en RI ($\sigma = 14,15$ pour G1 et $\sigma = 20,28$ pour G2). En RI, dans « teddy », les deux groupes produisent une voyelle aussi fermée que le [e] du locuteur natif dans « red », sans doute en raison d'une harmonisation vocalique due à la présence de la voyelle [ɪ] en deuxième syllabe. Si de plus, cette deuxième syllabe est accentuée comme elle le serait en français, le phénomène d'harmonisation est renforcé. Dans « umbrella », la voyelle est également aussi fermée pour G1 en RI. En revanche pour G2, elle se situe entre les deux voyelles françaises [e] et [ɛ], comme dans la production du locuteur natif. Enfin, dans le mot « eleven », la voyelle est très fermée en RD pour G1 et G2, surtout si la première voyelle est correctement prononcée, c'est-à-dire [ɪ] (influence fermante). En RI cependant, la voyelle est plus ouverte, localisée au même endroit que la voyelle française [œ].

[ə] : la voyelle [ə] varie beaucoup chez le locuteur natif et son timbre est fortement influencé par le contexte phonétique dans lequel elle apparaît. Ainsi, dans des mots comme « purple » et « apple », elle est très postérieure, entre la voyelle anglaise [ʊ] et la voyelle [o], dans « apple » (F2-F1 : $64 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $66 \frac{1}{8}^{\circ}$ de tons) et elle est localisée au même endroit que la voyelle [o] dans « purple ». Cette postériorisation provient de la présence de la latérale [l], qui provoque également une labialisation de la voyelle. Dans des mots comme « banana » et « umbrella », la voyelle [ə] est très centralisée (F2-F1 : $67 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $36 \frac{1}{8}^{\circ}$ de tons pour « umbrella » par exemple). Dans « parrot », la voyelle est légèrement plus fermée et plus antérieure, proche de la voyelle française [ɛ] (influence de l'articulation de l'occlusive dentale [t]). Dans le mot « eleven », la voyelle est aussi fermée que [e] et aussi antérieure que [ɛ] (F2-F1 : $99 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $26 \frac{1}{8}^{\circ}$ de ton), sûrement influencée par les voyelles fermées se trouvant dans les deux syllabes précédentes.

En situation de répétition différée, les enfants des deux groupes produisent l'équivalent de la voyelle française [œ] (F2-F1 : $80 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $36 \frac{1}{8}^{\circ}$ de ton pour G1 et F2-F1 : $81 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $35 \frac{1}{8}^{\circ}$ de ton pour G2) ; il y a donc un effet de filtre assez marqué (voir **extrait sonore n°67**, « seven » prononcée par l'enfant n°6 du groupe G1).

En situation de répétition immédiate, la voyelle est plus ouverte et légèrement plus postérieure, proche de la voyelle française [ɔ] (F2-F1 : $66 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $40 \frac{1}{8}^{\circ}$ de ton pour G1 et F2-F1 : $59 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $49 \frac{1}{8}^{\circ}$ de ton pour G2) ; par rapport à la situation de discours précédente, l'effet de crible est donc moins marqué.

On retrouve cette voyelle dans les mots « parrot », « apple » et « purple » dont toutes les occurrences sont prononcées en RI. Il est à noter dans les deux derniers mots une très forte variabilité de l'intervalle F2-F1 pour G2 ($\sigma = 30,48$ dans « apple » et $\sigma = 22,15$ dans « purple »). En RI, dans « umbrella », le timbre de la voyelle est proche de la voyelle [æ] (F2-F1 : $55 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $37 \frac{1}{8}^{\circ}$ de ton pour G1 et F2-F1 : $48 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $36 \frac{1}{8}^{\circ}$ de ton pour G2) ; à cause de l'accentuation finale du français, les enfants doivent prononcer cette voyelle de façon pleine (on trouve aussi pour ce mot l'influence du mot français « ombrelle », **extrait sonore n°65**). Dans « seven », c'est la voyelle [œ] qui est produite ; il y a donc assimilation de timbre, d'autant plus marquée si l'enfant produit l'accentuation en finale comme en français. La première voyelle du mot « banana » est proche de [ɔ] en RI (F2-F1 : $64 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $34 \frac{1}{8}^{\circ}$ de ton pour G1 et F2-F1 : $63 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $43 \frac{1}{8}^{\circ}$ de ton pour G2). La seconde voyelle est produite avec le même timbre par G2. En revanche, elle est plus proche de [ʌ] pour G1 (F2-F1 : $70 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $35 \frac{1}{8}^{\circ}$ de ton), mais également extrêmement variable (F2-F1 : $\sigma = 29,62$, ce qui représente l'un des écart-type les plus élevés de l'ensemble de nos mesures, avec celui de la voyelle [ə] de « apple »).

En RD, dans « eleven », la voyelle est proche de [œ] pour les deux groupes (F2-F1 : $88 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $33 \frac{1}{8}^{\circ}$ de ton pour G1 et F2-F1 : $86 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $34 \frac{1}{8}^{\circ}$ de ton pour G2) ; il en est de même en RI pour G2. En revanche, dans la même situation de discours, G1 produit une voyelle plus fermée dont le timbre se situe entre les deux voyelles françaises [e] et [ø] (F2-F1 : $97 \frac{1}{8}^{\circ}$ de tons et F3-F2 : $27 \frac{1}{8}^{\circ}$ de ton). Une forte variabilité affecte l'intervalle F2-F1 pour G1 et G2 ($\sigma = 18,64$ pour G1 et $\sigma = 22,58$ pour G2).

3.2.1.1.2. Voyelles longues

[i:] : le locuteur natif produit une voyelle avec un rapport F3-F2 moins important que la voyelle la plus antérieure de l'anglais [ɪ], mais avec un rapport G2-F1 beaucoup plus élevé, comme dans « green » (158-21 1/8° de tons), « sheep » (164-13 1/8° de tons) et « thirteen » (162-17 1/8° de tons).

Dans les productions des enfants, on relève un effet de filtre massif en direction de la voyelle française [i] sans distinction de groupe, ni de discours (118-20 1/8° de tons pour G1 et 121-24 1/8° de tons pour G2 en RD ; 122-17 1/8° de tons pour G1 et 127-18 1/8° de tons pour G2).

Dans le détail des mots, c'est dans le mot « sheep » que la voyelle est produite avec le rapport F2-F1 le plus élevé (132-20 1/8° de tons pour G1 et 125-17 1/8° de tons pour G2 en RD), mais même pour G1, son timbre reste proche de la voyelle française [i]. Dans le mot « green » en RD, la voyelle produite est un [i] pour les deux groupes (voir **extrait sonore n°68**, « green » prononcé par l'enfant n°16 du groupe G2). En revanche, pour « green » en RI et dans tous les autres mots, c'est-à-dire « three » en RD, « thirteen » en RD et en RI, on relève une différence entre G1 et G2. En effet, les enfants de G1 produisent une voyelle plus ouverte que le [i] français qui se rapproche de la voyelle française [e] (pour « green » en RI, 110-23 1/8° de tons pour G1 et 131-19 1/8° de tons pour G2), alors que les enfants de G2 produisent un [i]. Par ailleurs, on relève une forte variabilité de l'intervalle F2-F1 pour G2 dans « green » en RI ($\sigma = 20,10$) et dans « thirteen » en RD ($\sigma = 25,14$). Dans « three » en RD et « thirteen » en RI, cette variabilité concerne les deux groupes ($\sigma = 20,52$ pour G1 et $\sigma = 19,73$ pour G2 dans « three » ; $\sigma = 25,85$ pour G1 et $\sigma = 19,05$ pour G2 dans « thirteen »).

[ɜ:] : dans les mots « bird », « thirteen » et « purple », le locuteur natif produit une voyelle légèrement plus postérieure et plus ouverte que la voyelle française [ɛ] (69-31 1/8° de tons pour « bird » ; 78-30 1/8° de tons pour « thirteen » ; 76-30 1/8° de tons pour « purple »).

Dans les productions des enfants, on relève un effet de filtre massif en direction de la voyelle française [œ], c'est-à-dire plus fermée que la voyelle anglaise [ɜ:] produite par le locuteur natif (81-38 1/8° de tons pour G1 en RD ; 82-34 1/8° de tons pour G1 et 86-33 1/8° de tons pour G2). En RD, G2 produit une voyelle encore plus fermée, proche de la voyelle française [ø] (94-31 1/8° de tons).

Dans le détail des mots, on relève le même effet de filtre pour « bird » en RI, pour « thirteen » en RI. En RD, pour le même mot, on note une différence entre les deux groupes : tandis que G2 produit une voyelle proche de la voyelle française [ø] (101-33 1/8^e de tons), G1 produit une voyelle de même aperture que [ɜ:], mais plus postérieure (71-43 1/8^e de tons) (voir **extrait sonore n°69**, « bird » prononcé par l'enfant n°6 du groupe G1). En RI, dans le mot « purple », on remarque une différence de même type entre les deux groupes : G2 produit une voyelle équivalente à la voyelle française [œ] (81-37 1/8^e de tons), alors que G1 produit la même voyelle que dans « thirteen » en RD (71/41 1/8^e de tons). Enfin, en RD, G1 produit une voyelle proche de la voyelle française [e] (102-33 1/8 de tons) (pas d'occurrences de ce mot en RD pour G2).

[ɑ:] : dans les mots « car », le locuteur produit une voyelle proche de [æ] (58-38 1/8^e de tons) et dans le mot « banana », une voyelle plus fermée, proche de [ʌ] (70-36 1/8^e de tons dans « banana »). Dans le chapitre V, la locutrice anglophone avait produit une voyelle plus ouverte et plus postérieure (32-51 1/8^e de tons dans « banana »), proche de la voyelle française [ɑ] produite par la locutrice francophone (36-45 1/8^e de tons).

Dans les productions des enfants, la voyelle [ɑ:] n'apparaît qu'en situation de RI. Son aperture est équivalente à celle du locuteur natif, mais elle est plus postérieure (67-43 1/8^e de tons pour G1 et 59-42 1/8^e de tons pour G2). Dans le mot « car » et dans le mot « banana », on ne relève aucune différence entre les groupes, mais une différence de timbre entre ces deux mots, similaire à celle produite par le locuteur natif (57-46 1/8^e de tons pour G1 et 51-41 1/8^e de tons pour G2 dans « car » ; 79-38 1/8^e de tons pour G1 et 66-42 1/8^e de tons pour G2 dans « banana »). Dans chacun de ces deux mots, l'aperture est identique à celle du locuteur natif, mais la voyelle est aussi postérieure que la voyelle française [ɑ]. Dans « banana », on relève cependant une grande variabilité de l'intervalle F2-F1 pour G1 ($\sigma = 21,91$ pour G1 et $\sigma = 16,40$ pour G2).

[ɔ:] : dans le mot « horse », le locuteur natif réalise la voyelle la plus postérieure et la plus ouverte de toutes ses productions (40-76 1/8^e de tons) ; sur la charte des rapports de formants, elle est localisée à l'extrême droite, tout en bas de la représentation.

Pour les productions des enfants, nous exposerons directement les résultats par mot car toutes les occurrences en RD sont produites avec le mot « four » et toutes les occurrences en RI, avec le mot « horse ». Dans le mot « horse », les enfants produisent une voyelle légèrement plus postérieure et plus ouverte que la voyelle française [ɔ] (49-52 1/8° de tons pour G1 et 51-53 1/8° de tons pour G2), et encore plus postérieure dans « four » (49-60 1/8° de tons pour G1 et 49-58 1/8° de tons pour G2), mais jamais aussi postérieure que la voyelle produite par le locuteur natif. Dans ce mot, l'intervalle F2-F1 est assez variable pour G1 ($\sigma = 20,06$ pour G1 et $\sigma = 12,75$ pour G2). L'**extrait sonore n°70** nous donne un exemple d'une très bonne reproduction du mot « horse » par un des enfants du groupe G1 (voir **extrait sonore n°70**, enfant n°5).

[u:] : dans le mot « blue », le locuteur natif produit une voyelle de même aperture, mais plus antérieure (81-51 1/8° de tons) que la voyelle française [u] (locutrice francophone chapitre V, 86-62 1/8° de tons).

Les deux mots « blue » et « two » n'apparaissent dans notre corpus qu'en situation de discours différé. Dans les productions des enfants, on ne relève aucune différence entre les groupes (72-48 pour G1 et 71-50 pour G2). Le timbre de la voyelle est proche de celui de la voyelle produite par le locuteur natif (voir **extrait sonore n°71**, « two » prononcé par les enfants n°4 et 6 du groupe G1). Le détail des mots ne révèle pas de différence et c'est toujours la même voyelle qui est produite (75-48 1/8° de tons pour G1 et 72-52 1/8° de tons pour G2 dans « two » ; 74-48 1/8° de tons pour G1 et 71-48 1/8° de tons pour G2 dans « blue »). On relève une forte variabilité dans l'intervalle F2-F1 pour G1 par rapport à G2 ($\sigma = 24,47$ pour G1 et $\sigma = 13,12$ pour G2) dans le mot « two ».

3.2.1.2. Diphtongues

Les résultats exposés dans le détail des mots ne seront pas traités en fonction de la variable « discours » puisque chaque mot contenant une diphtongue dans le corpus n'apparaissait que dans une seule situation de discours, soit en répétition différée, soit en répétition immédiate.

[aɪ] : dans le mot « nine », le locuteur natif produit une diphtongue dont le timbre de la première cible correspond à la voyelle [æ] (46-34 1/8° de tons) et le timbre de la seconde cible, à la voyelle [i:] (160-14 1/8° de tons), voyelle très fermée et très antérieure.

Comme pour le locuteur natif, les enfants des deux groupes produisent la première cible de la diphtongue [aɪ] avec un timbre proche de la voyelle [æ], et ce, dans les deux situations de discours (47-38 1/8^e de tons pour G1 et 54-37 1/8^e de tons pour G2 en RD ; 55-44 1/8^e de tons pour G2 en RI), bien qu'en situation de RI, les enfants du groupe G1 produisent une voyelle un peu plus postérieure (55-44 1/8^e de tons). La seconde cible est beaucoup moins fermée que celle du locuteur natif. Ainsi, dans les deux situations de discours, les enfants du groupe G1 produisent en moyenne une voyelle avec un timbre proche de la voyelle française [i] (122-21 1/8^e de tons en RI et 124-20 1/8^e de tons en RD), tandis que les enfants produisent une voyelle légèrement plus ouverte, proche de la voyelle française [e] (106-23 1/8^e de tons en RI et 114-22 1/8^e de tons en RD). Pour G2, on note une grande variabilité dans l'intervalle F2-F1 pour la première cible en situation de RI (σ =10,59 pour G1 et σ =25,24 pour G2 en RI) et pour la seconde cible dans les deux situations de discours (σ =10,38 pour G1 et σ =28,28 pour G2 en RI ; σ =13,35 pour G1 et σ =23,01 pour G2 en RD). Les cibles sont donc atteintes avec plus de précision par les enfants du groupe G1.

Dans le détail des mots « five » et « nine », on ne relève pas de différences entre les groupes. Ainsi dans ces deux mots, la première cible de la diphtongue correspond à la voyelle [æ] (47-38 1/8^e de tons pour G1 et 53-37 1/8^e de tons pour G2 dans « five » ; 47-37 1/8^e de tons pour G1 et 55-37 1/8^e de tons pour G2 dans « nine » en RD). Dans le mot « white », le timbre de la première cible est cependant légèrement plus postérieur, proche de la voyelle [ɒ] pour G1 (55-49 1/8^e de tons) et proche de la voyelle française [ɔ] pour G2 (54-39 1/8^e de tons). En ce qui concerne la deuxième cible, dans le mot « nine », son timbre correspond à celui de la voyelle française [i] pour les deux groupes (128-20 1/8^e de tons pour G1 et 119-20 1/8^e de tons pour G2). En revanche, on relève la même différence entre les groupes dans les mots « five » et « white » que celle qui apparaissait au niveau des moyennes générales. Les enfants du groupe G1 produisent une voyelle plus fermée, proche de la voyelle française [i] (122-21 1/8^e de tons pour G1 dans « white » et 122-20 1/8^e de tons pour G1 dans « five »).

Dans le mot « white », les enfants du groupe G2 produisent un intervalle F2-F1 très variable dans la première cible (σ =10,59 pour G1 et σ =25,24 pour G2). On relève la même variabilité pour la seconde cible (σ =10,38 pour G1 et σ =28,28 pour G2).

Ainsi, on trouve des productions allant de la simple voyelle, qui correspond à la voyelle [æ] (**extrait sonore n°72**, enfant n°15 ; 32-48 1/8° de tons) à une très bonne reproduction de la diphtongue (**extrait sonore n°73**, enfant n°16 ; 27-53 1/8° de tons pour la première cible et 115-27 1/8° de tons pour la seconde cible). Pour G1, les productions se rapprochent de celle de l'enfant n°6, qui réalise la meilleure production (**extrait sonore n°74** ; 38-52 1/8° de tons pour la première cible et 126-15 1/8° de tons pour la seconde cible). En revanche, dans les mots « five » et « nine », la différence entre les groupes est nettement moins marquée, même si on observe une plus grande variabilité pour G2.

[ei] : dans les mots « grappe » et « baby », le locuteur natif produit des diphtongues semblables et dont la seconde cible est très fermée, aussi fermée qu'un [i] et aussi antérieure qu'un [i:] (respectivement 136-22 1/8° de tons et 140-23 1/8° de tons). Il est normal que le point d'arrivée soit aussi fermé car la cible de départ est elle-même relativement fermée (proche de la voyelle anglaise [e] ; 101-14 1/8° de tons dans « grappe » 101-20 1/8° de tons dans « baby ») et il faut une trajectoire minimale pour que la diphtongue soit perçue en tant que telle.

Dans les productions des enfants, la première cible est similaire pour les deux groupes et dans les deux situations de discours. Son timbre est approximativement localisé entre la voyelle française [e] et la voyelle française [ɛ] (97-28 1/8° de tons pour G1 et 103-25 1/8° de tons pour G2 en RI ; 84-27 1/8° de tons pour G1 et 90-23 1/8° de tons pour G2 en RD). La seconde cible est proche du timbre de la voyelle française [i] pour les deux groupes en RD (118-22 1/8° de tons pour G1 et 123-23 1/8° de tons pour G2), ainsi que pour G2 en RI (127-23 1/8° de tons). Elle est en revanche légèrement plus antérieure pour G1 en RI avec un intervalle F3-F3 plus faible (124-15 1/8° de tons). En ce qui concerne la variabilité dans les productions des enfants, elle n'est pas très marquée sauf pour l'intervalle F2-F1 pour G1 en RD ($\sigma = 24,66$ en RD et $\sigma = 14,75$ en RI) et pour G2 en RD ($\sigma = 22,10$ en RD et $\sigma = 16,64$ en RI).

Dans le détail des mots, la diphtongue de « eight » est réalisée par tous les enfants. Les enfants du groupe G2 produisent une première cible dont le timbre se situe entre ceux des deux voyelles françaises [ɛ] et [e] (90-23 1/8° de tons), tandis que celui produit par les enfants du groupe G1 est plus proche de la voyelle française [ɛ] (84-27 1/8° de tons).

On relève plus de variabilité pour l'intervalle F2-F1 pour G1 que pour G2 ($\sigma = 24,66$ pour G1 et $\sigma = 15,74$ pour G2), à cause de la production divergente d'un des enfants, l'enfant n°8 (38-64 pour la première cible et 99-20 1/8^e de tons pour la seconde cible. Les autres enfants produisent des diphtongues très marquées comme l'enfant n°5 (38-64 pour la première cible et 99-20 1/8^e de tons pour la seconde cible ; **extrait sonore n°75**) et l'enfant n°7 (112-17 pour la première cible et 132-17 1/8^e de tons pour la seconde cible ; **extrait sonore n°76**). Dans cet extrait, la production de cet enfant est d'ailleurs précédée de celle, toute aussi bonne, d'un des enfants du groupe G2, l'enfant n°15 (110-11 pour la première cible et 129-19 1/8^e de tons pour la seconde cible). En ce qui concerne le timbre de la seconde cible, il correspond à celui de la voyelle française [i] pour G2 (123-23 1/8^e de tons) et il est légèrement plus ouvert pour G1, se situant entre le timbre des deux voyelles françaises [e] et [i] (118-22 1/8^e de tons). On relève plus de variabilité pour l'intervalle F2-F1 pour G2 ($\sigma = 16,17$ pour G1 et $\sigma = 22,10$ pour G2).

Si la diphtongue était plus ou moins correctement réalisée dans le mot « eight », il n'en va pas de même dans les mots « baby », « cake » et « grappe », où les enfants ont tendance à produire une voyelle à la place de la diphtongue. Cette tendance est particulièrement marquée dans le mot « grappe » (85,71 % de voyelles pour G1 et 100 % de voyelles pour G2), dans le mot « cake » (71,42 % pour G1 et 76,92% pour G2) et un peu moins dans le mot « baby » (33,33 % pour G1 et 57,14 % pour G2). Cette tendance est d'autant plus étonnante que ces trois mots soient produits en RI, c'est-à-dire tout de suite après le locuteur natif, alors que « eight » était produit en RD. Il semble que les effets de filtre peuvent être importants en RI, surtout si les mots sont assimilés à un mot français leur ressemblant. C'est le cas de « cake », terme utilisé en française pour désigner une sorte de gâteau ; il est prononcé [kɛk] (**extrait sonore n°77** ; dans l'**extrait n°78**, l'enfant n°16 produit également une voyelle, mais dont la durée est très allongée) On trouve le même type d'assimilation avec le mot « grappe » qui ressemble au mot français équivalent « grappe » prononcé [grap] (enfant n°1, **extrait sonore n°79**, en RD). En revanche, dans un mot comme « baby », il n'y a pas de correspondance avec le français si ce n'est avec le mot « bébé », mais celui-ci est très différent de la prononciation anglaise. Cependant, le mot « baby » est très utilisé par les français et on peut l'entendre très souvent prononcé de la manière suivante : [bebi]. La fréquence de cette prononciation a pu influencer la production des enfants, même en RI (enfant n°4, à la fin de l'**extrait sonore n°80**).

Si nous nous intéressons au timbre de cette voyelle, comment est-elle réalisée dans les trois mots ? Dans « baby » et dans « cake », elle équivaut à la voyelle française [e] (99-22 1/8° de tons pour G1 et 96-25 1/8° de tons pour G2 dans « baby » ; 105-22 1/8° de tons pour G1 et 105-18 1/8° de tons pour G2 dans « cake »), certains enfants produisant la voyelle [ɛ] (dans ce cas, l'assimilation au français est totale, comme dans l'exemple n°X). L'intervalle F2-F1 est relativement variable dans le mot « baby » (intervalle F2-F1 : $\sigma = 18,34$ pour G1 et $\sigma = 20,30$ pour G2) et dans le mot « cake » pour G1 (intervalle F2-F1 : $\sigma = 21,27$ pour G1 et $\sigma = 10,61$ pour G2). Cette variabilité s'explique par le fait que les enfants produisent soit une voyelle de timbre correspondant à [ɛ], soit une voyelle de timbre correspondant à [e], la différence entre les deux voyelles étant essentiellement une différence d'aperture (ce qui affecte l'intervalle F2-F1).

Enfin, dans le mot « grappe », le timbre de la voyelle est légèrement plus fermé, entre les voyelles françaises [e] et [i] (113-19 1/8° de tons pour G1 et 113-20 1/8° de tons pour G2). Comme dans les deux mots précédents, on relève une forte variabilité (intervalle F2-F1 : $\sigma = 18,34$ pour G1 et $\sigma = 20,30$ pour G2). Dans ce mot, il n'y a qu'un enfant qui produise une diphtongue, l'enfant n°2 appartenant au groupe G1 (**extrait sonore n°79**) et cette production est très proche de celle du locuteur natif (106-17 1/8° de tons pour la première cible et 139-13 1/8° de tons pour la seconde cible). Elle suit la production de l'enfant n°1, [grap]. Il est à noter que seuls 13 enfants sur 25 ont produit ce mot, dont 10 en G1 et 3 en G2). En ce qui concerne les diphtongues réalisées dans « cake », la première cible est proche de la voyelle française [e] dans les deux groupes (113-18 1/8° de tons pour G1 et 101-23 1/8° de tons pour G2) et la seconde cible est proche de la voyelle anglaise [ɪ] (136-14 1/8° de tons pour G1 et 137-13 1/8° de tons pour G2). On relève peu de variabilité sur la première cible, alors que la seconde cible est plus dispersée (intervalle F2-F1 : $\sigma = 17,46$ pour G1 et $\sigma = 16,87$ pour G2). Lorsque la diphtongue est réalisée, elle est correctement réalisée et proche de productions du locuteur natif. Dans le mot « baby », la diphtongue n'est pas produite exactement de la même façon selon les groupes. Ainsi, G1 produit une première cible dont le timbre est proche du timbre de la voyelle française [œ] et une seconde cible dont le timbre se situe entre ceux des deux voyelles françaises [e] et [i] (86-35 1/8° de tons pour la première cible et 115-16 1/8° de tons pour la seconde cible).

G2 produit une première cible dont le timbre est proche de celui de la voyelle française [e] (intervalle F2-F1 variable, $\sigma = 16,19$ pour G2 alors que pour G1, $\sigma = 9,05$) et une seconde cible de timbre proche de la voyelle française [i] (105-26 1/8^e de tons pour la première cible et 117-24 1/8^e de tons pour la seconde cible).

[au] : dans le mot « brown », le locuteur natif produit une diphtongue dont la première cible est plus ouverte que la voyelle [æ] (48-45 1/8^e de tons) et la seconde cible, de même aperture, mais plus postérieure à la voyelle [ø] (101-10 1/8^e de tons). Dans le mot « cow », la trajectoire est encore plus marquée puisque la première cible est proche de la voyelle [ʌ] (61-28 1/8^e de tons) et la seconde cible est plus postérieure que la voyelle [ɔ:] (44-83 1/8^e de tons).

Il n'y a que deux mots comportant la diphtongue [au] dans notre corpus, nous les traiterons donc directement sans passer par une moyenne générale, comme c'était le cas pour les précédentes diphtongues. Dans le mot « cow », la diphtongue n'est pas aussi marquée qu'elle l'était dans la production du locuteur natif, bien au contraire. La première cible équivaut à la voyelle [ʌ] pour G1 (60-35 1/8^e de tons) et à la voyelle [ɔ] pour G2 (62-41 1/8^e de tons). La seconde cible est plus postérieure pour G1, équivalente à la voyelle anglaise [ɒ] (52-53 1/8^e de tons) ; elle est très peu différenciée de la première cible pour G2 (61-45 1/8^e de tons) ce qui signifie qu'elle n'est pratiquement pas réalisée. Un des enfants produit la suite [kal] en RI, puis finit par se corriger après avoir entendu à nouveau le locuteur natif (enfant n°16 de G2 ; **extrait sonore n°81**). Dans cet extrait, on remarque à quel point la trajectoire du locuteur natif est marquée par rapport à celles des deux enfants. L'enfant n°23 (G2) réussit à produire l'une des plus belles diphtongues relevée pour ce mot (70-37 1/8^e de tons pour la première cible et 63-55 1/8^e de tons pour la seconde cible ; **dans l'extrait sonore n°82**, sa production est précédée par celle de l'enfant n°3, de G1).

Dans le mot « brown », la trajectoire de la diphtongue est marquée par une augmentation d'aperture (c'est-à-dire une montée de F2 et du rapport F2-F1), alors que le lieu d'articulation ne change guère.

Ainsi, pour les enfants de G1, le timbre de la première cible est proche de la voyelle anglaise [ʌ] et le timbre de la seconde cible, aussi fermé que le timbre de la voyelle française [œ] (65-32 1/8^e de tons), mais aussi postérieure que la voyelle anglaise [ɔ:] (80-42 1/8^e de tons). La trajectoire est moins marquée pour les enfants de G2, puisque le timbre de la première cible est proche de la voyelle française [ɔ] (62-46 1/8^e de tons) et le timbre de la seconde cible est proche de la voyelle française [œ] (74-38 1/8^e de tons). On relève également une très forte variabilité dans l'intervalle F2-F1 pour G2 alors que l'écart type est déjà élevé pour G1 (σ =30,83 pour G2 et σ =16,19 pour G1). Certains enfants ne produisent aucune diphtongue, mais une voyelle plus postérieure que [ɔ] pour G1 (55-46 1/8^e de tons) et une voyelle équivalente à [ɔ] pour G2 (44-39 1/8^e de tons). L'enfant n°4 (G1) produit la suite [brɔn] suivie par l'enfant n°6 (G2) qui produit la suite [brañ] (73-54 1/8^e de tons pour la première cible et 58-55 1/8^e de tons pour la deuxième cible). Cette dernière production est l'une des meilleures réalisations de cette diphtongue (à la fin de l'**extrait sonore n°83**) avec celle de l'enfant n°23 du groupe G2 (71-66 1/8^e de tons pour la première cible et 69-48 1/8^e de tons pour la seconde cible) (**extrait sonore n°84**).

[ɔɪ] : dans le mot « boy », le locuteur natif produit une diphtongue dont le timbre de la première cible est proche de la voyelle [ɔ:] (40-78 1/8^e de tons) et la deuxième cible, proche de la voyelle anglaise [e], mais plus antérieure, comme la voyelle [i:] (110-14 1/8^e de tons).

En répétition immédiate, les enfants produisent également une diphtongue, mais comme pour la voyelle [ɔ:], ils produisent une voyelle trop antérieure sur la première cible, mais quand même plus postérieure que la voyelle française [ɔ] (42-64 1/8^e de tons pour G1 et 51-56 1/8^e de tons pour G2), surtout pour les enfants de G1. Pour la deuxième cible, la voyelle produite par les deux groupes est aussi fermée que celle du locuteur natif, mais plus postérieure (108-23 1/8^e de tons pour G1 et 106-26^e de tons pour G2), tendant vers la voyelle française [i]. On relève une grande variabilité dans l'intervalle F2-F1 pour la seconde cible par rapport à la première cible, surtout dans les productions des enfants de G2 (σ =21,15 pour G1 et 31,67 pour G2).

On peut dire qu'il y a un léger effet de filtre sur les deux cibles, mais les enfants essaient d'atteindre une première cible aussi postérieure que celle du locuteur natif, même s'ils n'y parviennent pas (il y aurait un effet de filtre marqué s'ils avaient produit la voyelle française [ɔ]). La seconde cible tend vers la voyelle française [i], elle est plus postérieure que la voyelle produite par le locuteur natif. Cependant, leur timbre n'est pas très éloigné. Les deux meilleures productions ont été relevées dans chacun des deux groupes (38-99 1/8^e de tons et 144-13 1/8^e de tons pour l'enfant n°3 du groupe G1, **extrait sonore n°85** ; 44-73 1/8^e de tons et 153-22 1/8^e de tons pour l'enfant n°23 du groupe G2, **extrait sonore n°86**).

[əʊ] : dans le mot « yellow », le locuteur natif ne produit pas de diphtongue, mais la voyelle [ɔ:] (39-87 1/8^e de tons). En répétition différée, les enfants ne produisent pas non plus de diphtongue, mais une voyelle un peu plus postérieure que [ɔ] (66-49 1/8^e de tons pour G1 et 58-53 1/8^e de tons pour G2). Sur les chartes de rapports de formants (Volume II, annexe n°10.3.3.), on peut voir que pour G1, toutes les productions correspondent à [ɔ], tandis qu'un des enfants produit une voyelle proche de celle du locuteur natif (44-82 1/8^e de tons). Pour G2, les productions se répartissent plus progressivement entre les deux sons, un des enfants atteignant également la voyelle [ɔ:] (45-75 1/8^e de tons). La production de ces deux enfants intervient au cours du même entretien, l'enfant du groupe G2 (enfant n°15) étant influencé par la production de l'enfant du groupe G1 (enfant n°7) (production des deux enfants dans l'**extrait sonore n°87**, l'enfant n°7 figurant en premier).

3.2.2. Durée du noyau vocalique

3.2.2.1. Voyelles brèves

Dans notre corpus, les six voyelles brèves de l'anglais [ɒ], [ɪ], [æ], [e], [ʌ] et [ʊ] apparaissent au sein de 13 mots monosyllabiques (« *dog, doll, pink, six, fish, black, cat, red, ten, twelve, book, one, duck* »). La structure syllabique de ces mots est de type CVC. Il n'y a pas de consonne finale allongeante, sauf dans le mot « *twelve* ». La durée de la voyelle brève [ə] n'a pas été relevée car cette voyelle n'apparaît que dans les mots multisyllabiques en anglais (dans les syllabes atones).

Nous avons vu dans le chapitre V (3.1.2.1.1.1.) que les voyelles brèves de l'anglais se caractérisent par une durée absolue brève, généralement inférieure à 210 ms (comprise entre 180 et 200 ms pour les voyelles brèves de l'A.N.A. d'après les données de Lehiste et Peterson, 1960) et une durée relative faible par rapport aux voyelles longues et aux diphtongues (de 24,47 % à 29,52 % dans les syllabes CVC). Or, d'après nos résultats, les enfants produisent ces voyelles avec une durée absolue encore plus réduite et ce, pour les deux groupes et dans les deux situations de discours (tableau n°62, moyennes allant de 133,25 ms à 162,25 ms). Nous avons relevé des durées similaires dans le chapitre V pour les voyelles françaises [i], [e], [ɛ], [a], [u], [ɔ] et [œ] (inférieures à 139 ms) et nous avons posé l'hypothèse que les francophones, influencés par leurs voyelles maternelles, raccourcissaient les voyelles brèves de l'anglais (chapitre V, 3.2.1.1.4.), notamment les voyelles de timbre correspondant (à cause du phénomène de crible phonologique, assimilation aux voyelles natives et réalisation de la durée des voyelles natives ; par exemple [ɪ] à [i] ; [ʌ] à [œ] et [ʊ] à [u]). Cette hypothèse semble se vérifier d'une manière générale pour les enfants de notre corpus. A l'échelle de la syllabe, la durée relative n'est en revanche pas modifiée.

Si nous regardons les résultats de manière plus détaillée (moyennes par voyelle disponibles dans le Volume II, annexe n°10.3.), les voyelles les plus courtes de l'anglais, c'est-à-dire [ɪ], [ʌ] et [ʊ] sont produites de manière encore plus brève par les enfants francophones, notamment la voyelle [ʌ] (respectivement 98,7 ms et 93,33 ms pour G1 et G2 en RD). Pour les deux groupes, la durée absolue est cependant plus élevée en situation de RI, notamment pour G2 (162,25 ms). Il paraît logique qu'en RI, la durée des voyelles de l'anglais soit plus facilement restituable, car l'enfant répète tout de suite après le locuteur natif.

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	134,05	148,72	133,26	162,25
DR (en %)	27,51	28,30	28,84	31,48

Tableau n°62 : résultats pour les voyelles brèves dans les mots monosyllabiques (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Nous avons posé comme hypothèse initiale que les enfants du groupe « musiciens » (G1) produiraient des voyelles brève avec une durée absolue et relative proche des voyelles prononcées par un locuteur natif. Cette hypothèse ne se vérifie pas et l'ensemble des enfants produit des voyelles encore plus courtes que ne le sont les voyelles anglaises.

3.2.2.2. Voyelles longues

Les 5 voyelles longues de l'anglais [i:], [ɑ:], [ɔ:], [u:] et [ɜ:] apparaissent au sein de 9 mots monosyllabiques dans notre corpus (« *car, horse, four, blue, two, bird, green, sheep, three* »). La structure syllabique de ces mots est soit de type CVC, soit de type CV. Les voyelles longues de l'anglais se caractérisent par une grande durée absolue (chapitre V, 3.2.1.1.2.) pouvant aller de 233 à 325 ms dans les syllabes de type CVC (entre 240 et 260 ms pour les voyelles longues de l'A.N.A. d'après les données de Lehiste et Peterson, 1960). Dans le même type de syllabe, la durée absolue des voyelles longues produites par les enfants de notre corpus correspond à ces données. Ils respectent la différence de durée entre ces voyelles et les voyelles brèves (les voyelles longues étant en moyenne deux fois plus longues que les voyelles brèves). Leur durée relative correspond à la durée la plus longue relevée pour la locutrice native dans le chapitre V. En effet, la proportion moyenne de la voyelle longue au sein de la syllabe varie pour les enfants de 42 à 46 % (tableau n°63), alors que la durée relative la plus longue était de 42,09 % (voyelle [ɑ:]) pour la locutrice dans les syllabes CVC et pouvait descendre jusqu'à 27,9 % (voyelle [i:]).

En ce qui concerne la différence entre les deux groupes, on relève un comportement inverse de G1 par rapport à G2. Nous avons déjà noté que les voyelles brèves étaient plus longues en situation de RI pour G2. Ce phénomène se confirme pour les voyelles longues (durée moyenne de 261,55 ms en RI contre 218,91 ms en RD). En revanche, pour G1, c'est en situation de RD que les voyelles longues sont plus longues (262 ms) par rapport à la situation de RI (229,05 ms). Cependant, pour les deux groupes, la durée relative des voyelles longues est plus élevée en RD qu'en RI (46,4 % pour G1 et 46,94 % pour G2 en RD contre 42,08 % pour G1 et 42,48 % pour G2 en RI).

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	262	229,05	218,91	261,55
DR (en %)	46,4	42,08	46,94	42,48

Tableau n°63 : résultats pour les voyelles longues dans les mots monosyllabiques de type CVC (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée,

RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Les voyelles longues qui apparaissent dans des syllabes de type CV sont plus longues chez les enfants du groupe G1 que chez les enfants du groupe G2 (tableau n°64). C'est en situation de répétition immédiate que la différence est la plus marquée (332 ms pour G1 contre 105,33 ms pour G2), mais il y a trop peu d'occurrences en RI pour déterminer une tendance générale, les mots « *three* », « *blue* » et « *two* » étant connus des enfants et produits majoritairement en situation de répétition différée (voir tableaux figurant dans le Volume II, annexe 10.3.).

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	192,62	332	175,62	105,33
DR (en %)	49,54	47,29	48,06	30,89

Tableau n°64 : résultats pour les voyelles longues dans les mots monosyllabiques de type CV (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Nous ne commenterons pas les résultats par mots (disponibles dans le Volume II, annexe 10.3.), mais nous prendrons deux exemples pour illustrer deux comportements différents. La durée absolue de la voyelle longue [i:] ainsi que sa durée relative semblent plus courtes dans les mots « *sheep* » et « *green* » produits par les enfants (Volume II, annexe n°10.3.) que ce qui avait été relevé pour la locutrice francophone dans le chapitre V. Ce « raccourcissement » pourrait indiquer un phénomène d'assimilation avec la voyelle française de timbre correspondant [i] et c'est effectivement ce qu'indiquent les mesures de timbre réalisées dans la partie précédente (voir 3.2.1.1.2.). A l'inverse, la voyelle longue [ɑ:] est très allongée dans le mot « *car* » pour les deux groupes (309 ms pour G1 et 380 ms pour G2 en RI). Or, cette voyelle sonne « anglais », d'autant plus qu'elle est suivie de la consonne [r] qui n'existe pas en français. En répétition immédiate les enfants perçoivent bien ce son et essaient de le restituer, notamment en amplifiant sa durée et en diphtonguant (**extrait sonore n°88**). Le mot « *car* » est connu en France, même par les non anglicistes et il est habituellement prononcé à la française, [kaɛr]. Sa « version » anglaise surprend l'un des deux enfants : « *on aurait dû penser à [kar], mais [kaewə]* » (**extrait sonore n°89**). En revanche, les caractéristiques spécifiquement « anglaises » de la voyelle [i:] sont beaucoup moins saillantes que celles de la voyelle [ɑ:], donc moins repérables, ce qui la rend plus vulnérable au phénomène de crible phonologique.

3.2.2.3. Diphtongues

Parmi les 8 diphtongues que compte le système vocalique anglais, les 4 diphtongues [aɪ], [eɪ], [ɔɪ] et [aʊ] apparaissent dans notre corpus au sein de 6 mots monosyllabiques (« *white, five, nine, eight, cake, brown* »). La structure syllabique de ces mots est de type CVC. Dans le chapitre V, nous avons vu que les diphtongues figurent parmi les noyaux vocaliques les plus longs de l'anglais avec une grande durée absolue (jusqu'à 510 ms). Dans les productions des enfants, les diphtongues sont également les noyaux vocaliques possédant les durées absolues les plus élevées par rapport aux voyelles brèves et aux voyelles longues. On ne relève aucune différence entre G1 et G2 (voir tableau n°65). En revanche, le discours semble affecter la durée absolue et relative des diphtongues de la même manière dans les deux groupes. Ainsi, on relève une durée absolue en RD pour G1 (293,15 ms) et pour G2 (295,86) plus élevée qu'en RI (248,26 ms pour G1 et 243,40 ms pour G2). De même, la durée relative est plus élevée en RD pour G1 (53,66 %) et pour G2 (52,59 %) qu'en RI (41,06 % pour G1 et 40,47 % pour G2).

	G1		G2	
	RD	RI	RD	RI
DA (en ms)	293,15	248,26	295,86	243,40
DR (en %)	53,66	41,06	52,59	40,47

Tableau n°65 : résultats pour les diphtongues dans les mots monosyllabiques de type CVC (tons (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens », RD pour répétition différée, RI pour répétition immédiate, DA pour durée absolue et DR pour durée relative).

Pour les diphtongues présentes dans des syllabes de type CV ([ɔɪ] dans « *boy* » et [aʊ] dans « *cow* »), on ne relève pas de différence entre les groupes (voir tableaux figurant dans le Volume II, annexe n°10.3.). Comme ces deux mots ne sont pas connus des enfants et ont été en majorité produits en situation de répétition immédiate, nous ne pouvons en tirer une tendance générale sur la différence entre les deux types de discours. Cependant, nous pouvons relever la durée absolue la plus élevée de tous les noyaux vocaliques, avec une durée relative proche des 80 %.

En ce qui concerne la différence entre les groupes, nous avons posé comme hypothèse que les enfants du « groupe musiciens » (G1) produiraient des voyelles longues et des diphtongues d'une durée absolue et relative plus grande que les enfants du « groupe non musiciens » (G2), spécialement en situation de répétition différée (RD). Cette hypothèse est en partie confirmée puisque les voyelles produites par G1 sont plus longues en RD dans les syllabes de type CVC.

Quant aux voyelles produites dans les syllabes de type CV, elles sont plus longues pour G1 en RD, mais aussi en RI. La tendance du groupe G1 à allonger les voyelles en situation de RI est confirmée pour les voyelles longues dans les syllabes de type CVC. Les diphtongues constituent les noyaux vocaliques les plus spécifiques de l'anglais par rapport au français ; ils sont également les plus allongés et cet allongement est correctement reproduit par l'ensemble des enfants de notre corpus. Aucune différence pertinente n'a été relevée entre les groupes. En revanche, la variable indépendante « discours » influence la durée des diphtongues, puisque celle-ci est plus allongée en situation de répétition différée.

4. Etude n°5 « Stabilité » de timbre

4.1. Plan expérimental

4.1.1. Hypothèse

Les locuteurs dont l'oreille a été entraînée par la pratique musicale rencontreront plus de facilités à restituer l'instabilité de timbre des voyelles anglaises.

4.1.2. Variables indépendantes

4.1.2.1. La Musique : deux modalités

La variable indépendante « musique » est composée de deux groupes, le groupe d'enfants musiciens G1 et le groupe d'enfants non musiciens G2 (voir 2.1.2.1.).

4.1.2.2. Le discours : deux modalités

La variable indépendante « discours » est composée de deux modalités : la modalité de discours immédiat RI et la modalité de discours différé RD (voir 2.1.2.2.).

4.1.3. Variable dépendante : durée de la zone de stabilité de timbre de la voyelle

Dans le chapitre V (3.1.3.2.2.), nous avons vu que les voyelles brèves de l'anglais se caractérisent par une zone de stabilité de timbre étendue et des zones de transition réduites, alors que les voyelles longues se caractérisent par une zone de stabilité de timbre réduite par rapport à des zones de transition étendues. A cause du phénomène de crible phonologique, certaines voyelles longues peuvent être assimilées à des voyelles brèves et en conséquence, être produites de la même façon, c'est-à-dire avec une trop grande stabilité de timbre.

Par ailleurs, à cause de l'excès de tension articulatoire qui caractérise la production de sa LM et qu'il va transposer à l'anglais (rythmicité syllabique), un apprenant francophone éprouvera des difficultés à restituer l'instabilité de timbre des voyelles longues de l'anglais. En revanche, il ne devrait pas rencontrer de difficultés avec les voyelles brèves car celles-ci se caractérisent par une grande phase de stabilité.

Afin d'étudier le paramètre de stabilité, il faut déterminer pour chaque type de son, c'est-à-dire pour les voyelles longues, mais également pour les voyelles brèves, la proportion de stabilité et la proportion d'instabilité. Par exemple, dans le mot « *horse* » prononcé par le locuteur natif (figure n°102 et **extrait sonore n°90**), la proportion de stabilité de la voyelle [ɔ:] est de 20,45 %, alors que la proportion d'instabilité est de 79,54 % (avec une transition initiale de 10,25 % et une transition finale de 66,23 %).

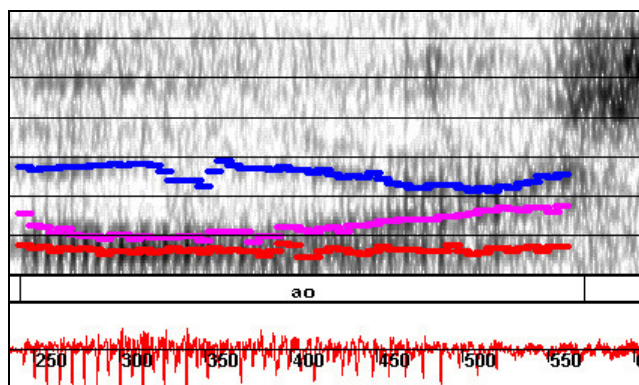


Figure n°102 : représentation spectrographique du mot « *horse* » avec tracé des trois premiers formants de la voyelle [ɔ:].

A l'inverse, dans le mot « *duck* » (figure n°103), la proportion de stabilité est de 88,17 %, alors que celle d'instabilité est de 11,82 % (absence de phase de transition initiale et présence d'une phase de transition finale réduite, de 11,82 %).

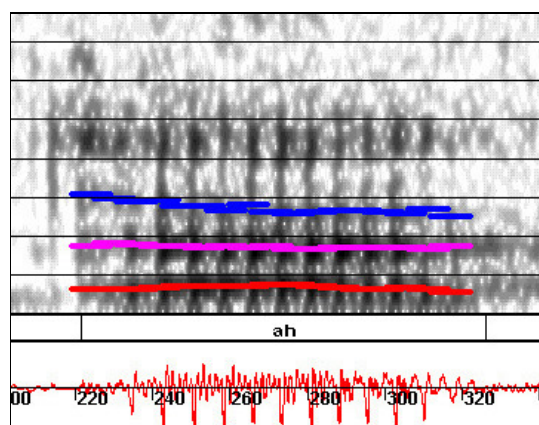


Figure n°103 : représentation spectrographique du mot « *duck* » avec tracé des trois premiers formants de la voyelle [ʌ].

La variable « stabilité de timbre » comprend donc deux modalités : la modalité « stabilité » et la modalité « instabilité ». La proportion de stabilité correspond à la durée relative de la zone de stabilité de timbre et la proportion d'instabilité, à la somme de la durée relative de la zone de transition initiale et de la durée de la zone de transition finale.

Nous posons l'hypothèse opérationnelle que les locuteurs dont l'oreille a été entraînée par la pratique musicale restitueront mieux l'instabilité de timbre des noyaux vocaliques longs de l'anglais, car ils pourraient reproduire plus facilement la rythmique spécifique de l'anglais et ils sont moins sensibles au phénomène de crible phonologique (cf. résultats de l'étude n°2 en perception). Quant aux voyelles brèves, elles seront produites avec une grande stabilité de timbre par les enfants des deux groupes. Le plan expérimental de l'étude est présenté dans le tableau n°66.

	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
Répétition différée (RD)				
Répétition immédiate (RI)				

Tableau n°66 : tableau expérimental de l'étude n°5 (G1 pour le « groupe musiciens », G2 pour le « groupe non musiciens »).

Matériel analysé

Notre corpus est constitué de 43 mots prononcés par 25 enfants (l'un des entretiens n'ayant pu être analysé en raison d'un fort bruit de fond), soit un total de 1075 mots. Les noyaux vocaliques étudiés sont les 5 voyelles longues de l'anglais, c'est-à-dire [i:], [ɜ:], [u:], [ɔ:] et [ɑ:] et les 7 voyelles brèves de l'anglais, c'est-à-dire [ɪ], [e], [æ], [ɒ], [ʌ], [ʊ] et [ə]. Cette dernière voyelle sera étudiée à part car elle dépend du phénomène d'accent. Sur la totalité des noyaux vocaliques appartenant à ces mots, nous avons lancé une détection automatique des trois premiers formants avec le logiciel Winsnoori (voir note n°X et annexe n°X pour un exemple de fichier de sortie édité avec ce logiciel). Cependant, cette détection a échoué pour un grand nombre de fichiers et seulement 260 fichiers ont pu être édités, soit 24,18 % de la totalité des mots composant le corpus. Outre des problèmes de qualité du signal pour certains enregistrements (rapport signal/bruit élevé), le problème principal a été le traitement des voix d'enfants. La détection automatique est en effet beaucoup plus facile pour des voix graves, en particulier les voix d'hommes.

Ainsi, toutes les productions du locuteur natif ont pu être analysées, même lorsque le signal était bruité. En revanche, la détection était beaucoup plus difficile pour les productions des enfants : leurs voix sont bien sûr beaucoup plus aiguës, mais de plus, certains enfants s'expriment avec une voix faible car ils sont intimidés par la situation d'entretien. Par ailleurs, nous avons relevé des problèmes spécifiques de détection concernant les voyelles dont les second et troisième formants sont élevés, comme [i:] et [I] ou concernant les voyelles compactes dont les premier et second formants sont très rapprochés, comme [u] et [u:].

A cause de ces problèmes de détection, la fiabilité des 260 fichiers retenus a été vérifiée scrupuleusement grâce aux mesures manuelles réalisées sur la totalité du corpus dans le cadre de l'étude n°4 (détail des fichiers, tableau n°67). Ils ont ensuite été traités par le programme informatique créé spécialement à cet effet (pour une description de la méthode et une description détaillée de la conception du programme, se reporter au Volume II, annexe n°11.1.). Celui-ci fournit la durée relative des différents événements composant le noyau vocalique, c'est-à-dire la transition initiale, la zone de stabilité de timbre et la transition finale (exprimée en pourcentage par rapport à la durée totale du noyau vocalique).

	G1		G2	
	RD	RI	RD	RI
Nombre de fichiers	46	61	89	85
Total	107		174	

Tableau n°67 : nombre de fichiers par mots et par discours (avec RD pour répétition différée et RI pour répétition immédiate).

3.2. Résultats

3.2.1. Résultats généraux

Les voyelles longues de l'anglais se caractérisent par une plus grande zone variabilité de timbre que les voyelles brèves et cette caractéristique est respectée par les enfants des deux groupes en situation de RD. En effet, la proportion d'instabilité est de 54,02 % pour G1 et de 51,09 % pour G2, alors que pour les voyelles brèves, elle est de 43,31 % pour G1 et de 35,69 % pour G2 (voir tableau n°67). En situation de RI, on observe la même tendance pour G1 avec une proportion d'instabilité de 50,17 % pour les voyelles longues et de 46,55 % pour les voyelles brèves. En revanche pour G2, dans la même situation de discours, les voyelles longues sont produites avec trop de stabilité (57,27 % de stabilité) et se rapprochent de la structure des voyelles brèves (42,72 % d'instabilité pour les voyelles longues contre 39,49 % d'instabilité pour les voyelles brèves).

On remarque donc que dans toutes les situations de discours, les voyelles longues sont produites avec une proportion d'instabilité plus grande pour G1 sauf pour G2 en RI, où elles sont produites avec trop de stabilité. L'hypothèse initiale est donc confirmée, les enfants musiciens restituent mieux l'instabilité de timbre qui caractérise les voyelles longues anglaises.

Les voyelles brèves sont produites avec une proportion de stabilité plus grande que les voyelles longues et cette structure est respectée par les enfants du groupe G2 (64,30 % en RD et de 60,50 % en RI). En revanche, nous relevons un **résultat inattendu** pour les enfants du groupe G1. Ils produisent les voyelles brèves avec une proportion d'instabilité plus élevée par rapport aux enfants du groupe G2 et ce, dans les deux situations de discours (46,55 % d'instabilité pour G1 contre 39,49 % pour G2 en RI et 43,31 % d'instabilité pour G1 contre 39,49 % pour G2 en RD). Nous pensions relever des résultats identiques pour les deux groupes car les voyelles brèves devraient être plus facilement reproduites que les voyelles longues. La stabilité qui les caractérise correspond en effet à la stabilité de timbre des voyelles françaises. Une interprétation possible de ces résultats est que les enfants musiciens ont repéré que l'instabilité de timbre caractérisait certaines voyelles anglaises. En effet, ils la reproduisent correctement au sein des voyelles longues. Cette caractéristique n'existe pas en français et elle contribue à donner à l'anglais sa « couleur » sonore spécifique. Il est possible qu'ils généralisent cette instabilité à des sons qui ne devraient pas l'être pour que leurs énoncés « sonnent » anglais. Nous avons déjà évoqué cette hypothèse à deux reprises dans l'étude n°3, pour les résultats concernant la variable « forme de la courbe » et la variable « localisation de l'accent ». En effet, les enfants du groupe G1 ont tendance à produire plus de contours spécifiquement anglais (les contours en cloche) sur des mots qui ne devraient pas être affectés par de tels contours (les mots commençant par une initiale non voisée). De la même manière, dans les mots trisyllabiques qui portent l'accent sur la seconde syllabe, les erreurs de localisation de l'accent produites par les enfants du groupe G1 affectent toujours la première syllabe et jamais la dernière. Or l'anglais se caractérise justement par une tendance à l'accentuation de la première syllabe. Le comportement des musiciens consisterait donc plutôt à « généraliser » des caractéristiques spécifiques à l'anglais plutôt qu'à les « exagérer ».

		G1		G2	
		Stabilité	Instabilité	Stabilité	Instabilité
RD	Voyelles brèves	56,68	43,31	64,30	35,69
	Voyelles longues	45,97	54,02	48,90	51,09
RI	Voyelles brèves	53,44	46,55	61,45	38,54
	Voyelles longues	49,82	50,17	60,50	39,49

Tableau n°68 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles brèves et des voyelles longues de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition immédiate (RI) et de situation différée (RD).

4.2.2. Résultats détaillés

4.2.2.1. Voyelles longues

En RD, la voyelle [i:] est marquée par une grande instabilité pour les deux groupes, surtout pour G2 (81,23 % pour G2 et 66,66 % pour G1 ; voir tableau n°69). En RI, cette instabilité est conservée pour G2 (66,72 %) et plus modérée pour G1 (47,53). Or cette voyelle figure parmi les voyelles les plus stables des voyelles longues de l'anglais (avec 63,79 % de stabilité chez la locutrice francophone étudiée dans le chapitre V). De même, en RD, la voyelle [ɜ:] est produite avec beaucoup d'instabilité par G2 (65,13 % d'instabilité), alors qu'elle est beaucoup plus stable pour G1 (38,55 % d'instabilité). Or, comme la voyelle [i:], cette voyelle est produite de manière stable par la locutrice anglophone (voir chapitre V). En revanche, la stabilité de cette voyelle est plus grande en RI (47,53 % d'instabilité pour G1 et 33,26 % d'instabilité pour G2). Pour ces deux voyelles, c'est G2 qui les produit avec le plus d'instabilité, surtout en RD.

Les voyelles [u:] ; [ɔ:] sont très instables pour G1 en RD et en RI. La même instabilité affecte la voyelle [u:] en RD pour G2 ; en revanche, on remarque une très grande stabilité en RI (93,82 % de stabilité, voir tableau n°70). A l'inverse, la voyelle [ɔ:] est marquée par une instabilité plus grande en RI (52,73% d'instabilité) qu'en RD (42,35 % d'instabilité). En RI, la voyelle [ɑ:] est instable pour G1 (57,57 % d'instabilité). On obtient les mêmes résultats pour G2 en RI (55,18 % d'instabilité). En revanche, on observe une grande stabilité en RD pour G2 (88,74 % de stabilité). Pour G1, la comparaison n'est pas possible en raison de l'absence d'occurrences pour cette voyelle en RD.

Répétition différée (RD)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[i:]	33,33	66,66	18,76	81,23
[ɜ:]	61,44	38,55	34,87	65,13
[u:]	44,22	55,77	44,49	55,50
[ɔ:]	44,90	55,08	57,64	42,35
[ɑ:]			88,74	11,25
Moyenne totale	45,97	54,02	48,90	51,09

Tableau n°69 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles longues de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition différée.

Répétition immédiate (RI)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[i:]	52,46	47,53	33,26	66,72
[ɜ:]	58,41	41,58	67,19	32,80
[u:]			93,82	6,18
[ɔ:]	46	53,99	47,25	52,73
[ɑ:]	42,41	57,57	44,81	55,18
Moyenne totale	49,82	50,17	57,27	42,72

Tableau n°70 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles longues de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition immédiate.

3.2.2.2. Voyelles brèves

Les voyelles brèves de l'anglais se caractérisent par une durée absolue réduite et surtout une grande stabilité de timbre (plus de 50 % de la durée de la voyelle pour la locutrice anglophone dans le chapitre V). Les voyelles [ɪ], [e], [æ] et [ɒ] sont très stables en RD pour G2. Pour G1, il n'y a que la voyelle [ɒ] qui soit produite avec beaucoup de stabilité (76 %, voir tableau n°71). Les autres voyelles sont plus instables et en particulier, la voyelle [æ] (38,04 % de stabilité contre 64,44 % pour G2). Quant à la voyelle [ʌ], elle est instable pour les deux groupes (57,82% d'instabilité pour G1 et 52,53 % d'instabilité pour G2).

En situation de RI, les voyelles brèves sont très stables pour G2 sauf la voyelle [ɒ] (57,19 % d'instabilité, voir tableau n°72). On peut remarquer que la voyelle [ʌ] est plus stable, mais pas aussi stable que les autres voyelles (51,26 % de stabilité, alors que pour [æ], 58,02 % de stabilité, mais surtout pour [ɪ], 75,13 % de stabilité et pour [e], 80,51 % de stabilité). Pour G1, on observe les mêmes résultats que G2 pour la voyelle [æ] (59,02 % de stabilité). On relève également plus de stabilité pour G1 en RI qu'en RD.

La voyelle [ʌ] est plus instable pour G1 que G2 (55,14 % d'instabilité contre 48,73 % pour G2). Pour la voyelle [ɒ], elle est plus stable pour G1 (50,23 % de stabilité) que pour G2 (42,31 %). On remarque plus d'instabilité pour G2 en RI par rapport à RD.

La voyelle [ə] est spécifique à étudier parce que sa stabilité dépend beaucoup de l'accentuation. En effet, dans les syllabes non accentuées, on observe une réduction du timbre et une grande instabilité. Mais, les enfants produisent ce son avec trop de stabilité. En effet, on observe dans la réalisation de ce son une proportion plus forte de stabilité que d'instabilité (la proportion de stabilité allant de 52,56 à 55,76 %) et ce, pour les deux groupes et dans les deux situations de discours.

Répétition différée (RD)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[ɪ]	56,92	43,0644,24	79,24	20,75
[e]	55,74	44,24	68,02	31,98
[æ]	38,04	61,95	64,44	35,55
[ɒ]	76,00	23,98	67,59	32,40
[ʌ]	42,16	57,82	47,46	52,53
[ʊ]			58,58	41,41
[ə]	52,56	47,43	64,80	35,19
Moyenne totale	56,68	43,31	64,30	35,69

Tableau n°71 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles brèves de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition différée.

Répétition immédiate (RI)	G1		G2	
	Stabilité	Instabilité	Stabilité	Instabilité
[ɪ]	49,13	50,86	75,13	24,86
[e]	61,64	38,35	80,51	19,49
[æ]	59,02	40,97	58,02	41,97
[ɒ]	50,23	49,75	42,31	57,68
[ʌ]	44,84	55,14	51,26	48,73
[ʊ]	55,70	44,28		
[ə]	53,48	46,51	55,79	44,20
Moyenne totale	53,44	46,55	60,50	39,49

Tableau n°72 : proportion de stabilité et d'instabilité de timbre (exprimée en pourcentage moyen) au sein des voyelles brèves de l'anglais prononcées par les enfants du « groupe musiciens » (G1) et du « groupe non musiciens » (G2) en situation de répétition immédiate.

5. Résumé des résultats et discussion générale

5.1. Production de la prosodie

Dans l'étude n°3, nous avons posé comme hypothèse que les locuteurs dont l'oreille a été entraînée par la pratique musicale rencontreraient plus de facilités à reproduire les éléments musicaux de l'anglais, c'est-à-dire les éléments prosodiques. Cette hypothèse se vérifie partiellement pour deux variables sur les trois variables étudiées.

5.1.1. Forme de la courbe

En ce qui concerne la forme de la courbe, il apparaît en effet qu'en situation de répétition immédiate, la proportion de contours spécifiquement anglais est plus grande chez les enfants musiciens. Inversement, la proportion de ces contours dans les productions des non musiciens est plus faible, révélant peut-être un phénomène de crible prosodique avec le français en situation de répétition immédiate (réalisation de contours descendants à la place des contours en cloche). Pour des enfants de 7 ans, il ne suffit donc pas de répéter des mots après un locuteur natif pour pouvoir les reproduire correctement. En revanche, le fait d'avoir une oreille entraînée par la pratique musicale leur permet de mieux repérer les caractéristiques prosodiques non natives et donc de mieux les reproduire. Par ailleurs, les enfants musiciens généralisent la production des contours en cloche aux énoncés où ces contours ne devraient pas apparaître, c'est-à-dire dans les mots monosyllabiques à l'initiale non voisée (normalement affectés d'un contour descendant). Les enfants se montrent donc « bons imitateurs », non pas dans le sens où ils imitent fidèlement la réalité, mais dans le sens où ils sont capables d'isoler les principales caractéristiques d'une langue et de les incorporer à leurs productions pour avoir au final des énoncés qui « sonnent » anglais, en tout cas aux oreilles des français. Ce faisant, ils se comportent exactement comme le ferait un caricaturiste lorsqu'il restitue les traits caractéristiques du visage d'une personne sur un dessin ou sa façon de parler dans une imitation vocale ou sa façon de bouger dans une imitation corporelle et gestuelle. Si les contours en cloche contribuent à la couleur « anglaise » de l'énoncé, alors pourquoi ne pas les généraliser ?

5.1.2. Localisation de l'accent

On retrouve un phénomène similaire dans la restitution de l'accent dans les mots trisyllabiques. Nous avons vu qu'il était difficile de tirer une conclusion quant à la différence entre G1 et G2 et ce, à cause du petit nombre d'occurrences analysées. Cependant, lorsque les enfants des deux groupes se trompent et localisent l'accent sur une mauvaise syllabe, c'est le plus souvent la première syllabe qui est concernée et non la syllabe finale, ce qui aurait traduit une interférence avec la rythmique du français. Or, l'accentuation initiale constitue l'une des caractéristiques distinctives de l'anglais par rapport au français (cf. Chapitre V, 2.1.1.2.). Comme pour les contours en cloche, les enfants pourraient avoir identifié cette caractéristique distinctive et la généraliser. Ainsi, même s'ils font des erreurs, leurs productions ressemblent à des mots anglais. Cette stratégie de généralisation est utilisée par les enfants des deux groupes, sûrement à cause du fait que l'accent en anglais est un phénomène sonore saillant. Il n'y a donc pas besoin d'avoir une oreille particulièrement entraînée pour le repérer. Nous avons vu dans le chapitre VII que l'accent était perçu facilement dans le test de perception T5 et ce, par l'ensemble des enfants de notre échantillon (étude n°1). Les résultats concernant leur restitution de l'accent dans les mots bisyllabiques confirment cette interprétation. En effet, les enfants des deux groupes le localisent quasiment parfaitement (plus de 90 % de productions correctes dans les deux groupes) en situation de répétition immédiate. En revanche, la tâche semble beaucoup plus difficile en répétition différée et ce sont les musiciens qui s'en acquittent le mieux. Dans ce type de répétition, c'est la mémoire à long terme qui est mise en jeu puisque ces enfants restituent des énoncés qu'ils ont appris en classe. La capacité des musiciens à reproduire correctement des énoncés stockés à long terme pourrait provenir d'une mémoire musicale plus optimale. La mémoire musicale de la parole est définie par Llorca (1992 : 47-48) comme :

« celle qui repose seulement sur les données concrètes, données phonétiques et prosodiques d'une séquence entendue, indépendamment de leur valeur linguistique : il s'agit purement de la mémoire des sons perçus et de leur agencement ».

La mémoire musicale est très importante dans une LE, puisqu'elle permet la mémorisation d'une forme sonore avant l'accès à son interprétation linguistique et à son sens. La prosodie joue un rôle essentiel dans cette mémorisation :

« On découvre que l'intégration des intonations et des schémas rythmiques se fait davantage par des souvenirs sonores que par des connaissances théoriques. Par exemple, on ne retient vraiment la place de l'accent dans un mot anglais que lorsqu'on a le souvenir de sa prononciation par une personne particulière, avec la voix de cette personne et l'intervalle mélodique qu'elle a placé entre la syllabe accentuée et la syllabe inaccentuée » (Llorca, 1992 : 60).

Les musiciens semblent donc capables de conserver des « souvenirs sonores » précis des énoncés qu'ils ont appris, probablement parce qu'ils ont été entraînés à mémoriser des extraits sonore au cours de leur formation musicale. En revanche, les énoncés seraient moins bien stockés par les non musiciens et pourraient être plus facilement altérés par des effets de filtre avec la langue maternelle. Il est donc nécessaire d'opérer un « rafraîchissement » fréquent au contact du modèle natif pour éviter ces phénomènes d'assimilation « à long terme ».

5.1.3. Exagération de la courbe

Contrairement aux deux autres variables, notre hypothèse ne se vérifie pas pour la variable « exagération de la courbe ». Non seulement les enfants musiciens n'utilisent pas de stratégie d'exagération, mais de plus, leurs contours moyens sont étonnamment identiques à ceux des enfants non musiciens. En effet, les contours des deux groupes se superposent presque parfaitement et ce, sans distinction de situation de discours. Par ailleurs, ces contours moyens sont beaucoup moins amples que les productions du locuteur natif, même en répétition immédiate. Pourtant, nous avons trouvé, dans un corpus rassemblant les productions de 10 enfants du même âge (Dodane, 1997, 2000), un enfant musicien, Anthony, qui manifestait une stratégie d'exagération très marquée. Il est possible que cette stratégie ne se manifeste que chez des enfants ayant un très bon niveau musical, ce qui était le cas de cet enfant. Il est également possible que d'autres facteurs interviennent dans la formation de ce type de stratégie comme la personnalité de l'apprenant par exemple. En effet, est-ce que l'apprenant va oser exagérer ses productions en anglais ou au contraire, va-t-il les « tasser » par peur d'avoir l'air ridicule devant ses camarades ? L'âge pourrait également se révéler déterminant, les enfants plus jeunes n'ayant pas peur de jouer avec les sons ; le poids social est déjà identique à 7 ans pour tous les enfants mais il est possible qu'une étude avec des enfants plus jeunes fasse apparaître des différences entre les groupes. Dans la présente étude, nous avons choisi d'étudier les productions des enfants par groupe, « groupe musiciens » et « groupe non musiciens ».

Ainsi, les contours étudiés sont des contours moyens, c'est-à-dire qu'ils gommement les différences entre les individus. Avec cette méthode, si l'un des enfants de notre corpus manifeste une stratégie d'exagération, nous ne pouvons pas le remarquer. Il serait donc intéressant, dans le futur, de compléter notre analyse par une étude qualitative des productions de chaque enfant de notre corpus. Cette étude sera facilitée par le fait que nous disposons dorénavant des contours moyens produits par l'ensemble des enfants, contours que nous pourrons comparer aux productions individuelles. Si nous n'avons pas fait apparaître de stratégie d'« exagération », nous avons en revanche fait apparaître, et de manière plus marquée chez les enfants musiciens que les enfants non musiciens, une stratégie de « généralisation » des caractéristiques prosodiques spécifiquement non natives de l'anglais.

5.2. Production des contrastes vocaliques

Dans l'étude n°4, nous avons posé l'hypothèse que le fait d'avoir une oreille entraînée par la pratique musicale amènerait les apprenants à mieux percevoir et donc mieux restituer le timbre et la durée des contrastes vocaliques de l'anglais, c'est-à-dire les voyelles brèves, les voyelles longues et les diphtongues.

5.2.1. Timbre du noyau vocalique

5.2.1.1. Voyelles brèves

L'ensemble de nos résultats ne fait guère apparaître de différences entre les groupes G1 et G2 ; on relève quelques différences en faveur de l'un ou l'autre groupe mais pas de tendance générale qui pourrait les différencier. S'il y a une assimilation, elle est réalisée la plupart du temps par les deux groupes pour une situation de discours déterminée ou pour les deux situations de discours. On relève plusieurs types de fautes :

- Soit on relève une **assimilation de la voyelle anglaise à une voyelle française** comme c'est le cas pour [ɪ] et [i] (bien que le groupe G1 ait tendance à produire une voyelle plus fermée, donc plus proche de la production du locuteur natif dans les mots « *teddy*, *orange*, *baby* »).

- On relève également des **assimilations provoquées par la transposition de la rythmique française à l'énoncé anglais**, en particulier dans les mots multisyllabiques : dans « **umbrella** », la voyelle [ʌ] devient [ɒ] et la voyelle [ə] devient [æ] ; dans « **seven** », la voyelle [ə] devient [œ]. La voyelle [ə] devrait être centrale car elle apparaît au sein d'une syllabe non accentuée (réduction de durée et de qualité). Or, les petits francophones ont tendance à produire des syllabes de durée identique à cause de la régularité syllabique et une syllabe allongée en syllabe finale à cause de la rythmique du français. Le timbre de la voyelle est donc trop marqué.
- **La prononciation de certains mots anglais semble influencée par leur prononciation « à la française »** (mots courants et connus même des non anglicistes), comme dans le mot « **dog** » (G1 et G2) et le mot « **orange** » (G1 en RI) où la voyelle [ɒ] est assimilée à la voyelle française [ɔ].
- **La voyelle est correctement reproduite**, comme c'est le cas de la voyelle [ɒ] dans « **orange** » (G1) et « **doll** » (G1 et G2). De la même manière, la voyelle [æ] est correctement produite, sauf dans « **apple** » (G1 et G2) et dans « **parrot** » (G2). C'est également le cas pour la voyelle [e] sauf dans le mot « **red** » (G2) (« **twelve**, comme LN). La voyelle [ʌ] est bien reproduite dans le mot « **one** » en RD pour les deux groupes.
- La voyelle peut être correctement reproduite, mais ses **caractéristiques non natives sont exagérées**. C'est le cas de la voyelle [ʊ] qui est prononcée comme la voyelle française [o] (pas d'assimilation avec la voyelle française [u] comme on aurait pu le prévoir), et de la voyelle [ʌ] dans le mot « **duck** » en RI qui est prononcée comme la voyelle française [ɔ].

5.2.1.2. Voyelles longues

Comme pour les voyelles brèves, on ne relève pas de différences marquées entre les groupes dans la prononciation des voyelles longues :

- La **voyelle anglaise [u:]** est **bien réalisée** et proche de la production du locuteur natif. La voyelle [ɔ:] n'est pas aussi postérieure que celle du locuteur natif, mais

elle est plus postérieure que la voyelle française [ɔ]. On ne peut donc pas dire qu'il y a assimilation car les enfants des deux groupes tentent d'« atteindre » le timbre de la voyelle produite par le locuteur natif, surtout dans le mot « horse ».

- La voyelle [ɑ:] est produite comme la voyelle [æ] par les enfants, mais le locuteur natif les prononce exactement de la même manière. En revanche, dans les mots « car » et « banana » en situation de répétition immédiate, les enfants produisent une voyelle plus postérieure que celle du locuteur natif en situation de répétition immédiate.
- **On relève une assimilation de la voyelle anglaise à une voyelle française dans deux cas :** la voyelle anglaise [i:] est assimilée à la voyelle française [i] et la voyelle anglaise [ɜ:] est assimilée à la voyelle française [œ]. Cependant, on remarque que les enfants musiciens, même s'ils sont sensibles à cette assimilation, produisent des voyelles dont le timbre est légèrement plus proche du timbre des voyelles produites par le locuteur natif.

5.2.1.3. Diphtongues

On relève quelques différences entre les groupes dans la production des diphtongue [aɪ] et [aʊ] ; les autres diphtongues sont produites de façon relativement homogène :

- Dans la diphtongue [aɪ], bien qu'on relève un effet de filtre sur la seconde cible pour les deux groupes ([i] ou [e]), les **deux cibles sont atteintes avec plus de précision par les enfants de G1**. De même, **ils réalisent mieux la diphtongue [aʊ]** que les enfants du groupe G2. Mais, même en fonction de ces différences, la trajectoire de l'ensemble des diphtongues produites par les enfants est beaucoup moins marquée que celle du locuteur natif.
- Dans la diphtongue [eɪ], on relève un **effet de filtre sur chaque cible** : la première cible est assimilée aux voyelles françaises [e] ou [ɛ] et la seconde cible, à la voyelle française [i]. Le même phénomène se produit pour la diphtongue [ɔɪ], où la première cible est assimilée à la voyelle française [ɔ] et la deuxième cible, à la voyelle française [i].

- Dans les mots « baby », « cake » et « grappe », **les enfants ont tendance à produire une voyelle à la place de la diphtongue**. Il est possible qu'ils soient influencés par la prononciation française de ces mots, comme nous l'avons déjà relevé pour certaines voyelles brèves. Dans les mots « baby » et « cake », le timbre de la voyelle est proche de celui de la voyelle française [e] et dans le mot « grappe », il se situe entre celui des deux voyelles françaises [e] et [i]. Dans le mot « yellow », tous les enfants produisent une voyelle équivalente à la voyelle française [ɔ], mais légèrement plus postérieure. Cependant, le locuteur natif ne produit pas non plus de diphtongue, mais la voyelle [ɔː].

L'ensemble de ces résultats ne laisse guère apparaître de différences entre les groupes de musiciens et de non musiciens, mis à part quelques diphtongues et quelques voyelles qui sont mieux reproduites par les musiciens. Cependant, comme nous l'avons déjà évoqué plus haut, l'analyse par groupe atténue les différences individuelles et il est possible qu'une analyse des productions par individu contribue à mettre à jour des différences entre musiciens et non musiciens, notamment si nous limitons cette analyse à la comparaison des productions des enfants ayant les meilleures oreilles musiciennes à celles des enfants ayant les plus mauvaises oreilles musiciennes de notre corpus.

Si nous avons échoué à faire apparaître une différence entre musiciens et non musiciens, les productions relativement homogènes des enfants révèlent des tendances générales qui pourraient caractériser les productions en anglais des enfants francophones (et peut-être des apprenants francophones d'une manière plus générale pour certaines d'entre elles) :

- assimilation de certaines voyelles anglaises aux voyelles françaises de timbre proche et effets de filtre sur les cibles des diphtongues;
- assimilation provenant de la transposition de la rythmique du français à l'anglais ;
- assimilation provenant de l'influence de la « prononciation à la française » de certains mots anglais utilisés très couramment dans l'environnement de l'enfant ;
- exagération des caractéristiques non natives de certaines voyelles.

5.2.2. Durée du noyau vocalique

En ce qui concerne les voyelles brèves, nous pensions que les musiciens produiraient des voyelles d'une durée absolue et relative proche de celle des voyelles produites par le locuteur natif. Cette hypothèse ne se vérifie pas et on ne relève pas de distinctions entre les groupes. Bien au contraire, l'ensemble des enfants produit des voyelles encore plus courtes que ne le sont les voyelles anglaises. Ce phénomène de « réduction » de durée pourrait s'expliquer par le fait que les voyelles anglaises sont assimilées aux voyelles françaises de timbre proche. En effet, les voyelles françaises étant plus courtes que les voyelles anglaises « correspondantes », ces dernières sont « raccourcies ». Ce phénomène de réduction est moins marqué en situation de répétition immédiate, notamment pour les enfants du groupe G2, sans doute parce que les enfants restituent les mots tout de suite après le modèle. La prononciation du mot, parce qu'elle est stockée à court terme, n'a pas eu le temps de s'altérer au contact des représentations maternelles stockées à long terme. Il est donc très important de fournir à l'enfant un modèle natif, authentique, en répétition immédiate, puis de le répéter régulièrement au cours de l'apprentissage de manière à éviter les effets de « filtre » dus au stockage à long terme. Ce phénomène confirme qu'il ne suffit pas de placer un enfant au contact d'une LE, pour que sa prononciation soit « native ». Les effets de filtre sont déjà présents à l'âge de 7 ans et il est nécessaire de prévoir une « rééducation », même légère, lors de l'apprentissage. Par ailleurs, les effets d'une prononciation « à la française » chez le professeur de langue seraient désastreux car le modèle fourni à l'enfant en répétition immédiate serait déjà altéré.

Par rapport aux voyelles brèves, les voyelles longues et les diphtongues de l'anglais possèdent des caractéristiques spécifiques que l'on ne retrouve pas en français, ce qui les rend particulièrement saillantes. Leur grande longueur absolue et relative est correctement reproduite par l'ensemble des enfants du corpus. Cependant, on relève une différence entre les groupes car les enfants musiciens produisent des voyelles longues d'une durée encore plus grande que celles produites par les enfants non musiciens. C'est le cas pour les voyelles apparaissant dans les syllabes de type CVC et dans les syllabes de type CV dans les deux situations de discours, mais de manière plus marquée en situation de répétition dans ce dernier contexte, l'allongement apparaît également en situation de répétition immédiate).

D'après ces résultats, il semble que plus un noyau vocalique possède des caractéristiques spécifiquement « natives », plus il sera facilement reproduit par les enfants et moins les effets de filtre n'en altéreront la prononciation. Par ailleurs, sa durée sera allongée plus facilement en situation de répétition différée. Le fait d'avoir une oreille entraînée paraît faciliter ce processus pour les voyelles longues. En revanche, les noyaux proches de ceux de la langue maternelle de l'apprenant seront assimilés à ces derniers, comme c'est le cas pour les voyelles brèves de l'anglais, sans qu'il y ait de différence entre les groupes. L'exagération des caractéristiques « natives » pourrait donc être un principe efficace pour éviter les phénomènes de filtre. La prononciation du professeur doit donc non seulement être native, mais très exagérée.

5.2.3. Stabilité de timbre

Dans l'étude n°5, nous avons posé l'hypothèse que les locuteurs dont l'oreille est entraînée par la pratique musicale restitueraient mieux l'instabilité de timbre spécifique aux voyelles longues de l'anglais à cause d'une meilleure restitution de la rythmique de cette langue et d'une sensibilité moindre au phénomène de crible phonologique. Cette hypothèse se vérifie en situation de répétition immédiate, puisque les enfants musiciens produisent les voyelles longues avec plus d'instabilité que les enfants non musiciens. L'instabilité de timbre affectant les noyaux vocaliques de longue durée constituant l'une des caractéristiques spécifiques de l'anglais, il est possible que les enfants l'ait « repérée » et qu'ils la généralisent à des sons qui ne sont normalement pas concernés. Si c'est le cas, ils feraient preuve de la même stratégie que la stratégie de généralisation mise à jour pour les caractéristiques prosodiques spécifiques de l'anglais, c'est-à-dire les contours en cloche et l'accentuation initiale. Or, c'est justement le comportement qu'ils adoptent pour reproduire les voyelles brèves de l'anglais. Celles-ci sont produites avec une grande instabilité et ce, dans les deux situations de discours, alors qu'elles se caractérisent habituellement par une grande phase de stabilité.

Bibliographie chapitre VIII

- Baken, Ronald J. and Daniloff, Raymond G. (1991). *Readings in Clinical Spectrography of Speech*. San Diego: Singular, 566 p.
- Beaumont et Desmoinaux (2000). *J'apprends, je dessine, je joue : Les Couleurs*. Ed. Cerf Volant, 10 p.
- Butterfield et O'Neill (1997). *Peekaboo, I know !* Instructional Fair TS Denison, 10 p.
- Dodane, Christelle (1996). *Traitement Comparé de la Musique et de la Langue*. Maîtrise. Université de Franche-Comté, Besançon.
- Dodane, Christelle (1997). *La Langue en Harmonie*. Diplôme d'Etudes Approfondies. Université de Franche-Comté, Besançon.
- Dodane, Christelle (2000). L'apprentissage précoce d'une langue étrangère : une solution pour la maîtrise de l'intonation et de la prononciation. In Guimbretière, 229-248.
- Guimbretière, Elisabeth (2000). *Apprendre, Enseigner, Acquérir : la Prosodie au Cœur du Débat*. Rouen : Publications de l'Université de Rouen, 312 p.
- Hawks, John W. (1994). Difference limens for formant patterns of vowel sounds. In *Journal of Acoustical Society of America*, 95/2, 1074-1084.
- Imagier en Anglais* (1997). Père Castor. Paris : Flammarion, 258 p.
- Kewley-Port, Diane (1990). Thresholds for formant frequency discrimination in isolated vowels. In *Journal of Acoustical Society of America*, 87/1, S159.
- Konopczynski, Gabrielle et Mac Carthy, Brian (1977). Les effets de la lecture à voix haute sur l'intonation. In *International Review of Applied Linguistic in Language Teaching*, 15/1, 230-238.
- Lehiste Ilse and Peterson Gordon E. (1960). Transitions, glides, and diphthongs. In Baken and Daniloff (1991), 286-295.
- Llorca, Régine (1992). Le rôle de la mémoire musicale dans la perception d'une langue étrangère. Dans *Revue de Phonétique Appliquée*, 102, 45-67.
- Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. In *Journal of Acoustical Society of America*, 85, 2114-2134.
- Nakagawa, T., Saito, S. and Yoshino, T. (1982). Tonal difference limens for second formant frequencies of synthesized Japanese vowels. In *Ann. Bull. RILP*, 16, 81-88.
- Piaget Jean (1937). *La Construction du Réel chez l'Enfant*. Paris : Delachaux et Niestlé, 399 p.
- Zenatti, Arlette (1980). *Tests Musicaux pour Jeunes Enfants, avec Applications en Psychopathologie de l'Enfant et de l'Adulte*. Issy Les Moulineaux : Editions Scientifiques et Psychologiques, 53 p.

Conclusion

Nous avons posé comme hypothèse générale dans notre travail qu'il existait un transfert de compétence de la pratique musicale à la pratique d'une langue étrangère. Cette hypothèse se fonde sur une revue de la littérature consacrée à ce sujet et dont nous tirerons les conclusions suivantes :

- On relève de grandes proximités de formes entre la musique et la langue, en particulier au niveau prosodique, ainsi que de grandes similitudes structurales, notamment par le fait qu'elles répondent toutes deux à une organisation hiérarchique.
- Le bébé fait preuve de capacités de traitement communes pour la langue et pour la musique, ce qui amène à se poser la question d'une éventuelle capacité cognitive commune. Le bébé montre également une grande sensibilité aux éléments musicaux, sensibilité à laquelle s'adapte son entourage par des ajustements vocaux spécifiques.
- L'environnement musical et l'environnement linguistique exercent le même type d'influences en provoquant d'une part une spécialisation de l'attention du jeune enfant sur sa langue maternelle et son système musical maternel dans un but d'acquisition, d'autre part une baisse simultanée de ses capacités de discrimination des contrastes non natifs. Ensuite, après l'émergence du babillage canonique, la langue et la musique empruntent des chemins divergents, mais on continue à relever des caractéristiques communes dans le développement linguistique et musical du jeune enfant (notamment le passage simultané d'un traitement de type global à un traitement de type analytique).

Les similitudes entre la langue et la musique nous ont amenée à penser que les « compétences » qui caractérisaient une oreille musicienne devaient se reporter sur le traitement de la langue, notamment sur ses caractéristiques musicales, c'est-à-dire la prosodie. Nous avons également posé comme hypothèse qu'une meilleure perception devait engendrer une meilleure production en vertu des liens qui existent entre perception et production.

Une grande partie de nos résultats suggère un transfert de compétence de la pratique musicale à la pratique d'une langue étrangère. Cependant, nous prendrons quelques précautions en ce qui concerne les interprétations qui pourraient en être tirées. Dans les années 60, il était courant de penser qu'il suffisait de placer un enfant au contact d'une langue étrangère avant un certain âge pour qu'il se mette à la parler couramment.

C'est naturellement inexact. De même, nous ne prétendons pas qu'il suffise qu'un enfant soit musicien pour qu'il apprenne facilement et correctement une langue étrangère. Une telle conclusion irait à l'opposé des objectifs que nous sommes fixés dans ce travail. Par ailleurs, il est possible que d'autres facteurs interviennent sur les résultats des enfants qui participent à notre corpus, comme l'influence de l'expérimentateur et de l'environnement familial et scolaire.

Cela étant dit, les musiciens se montrent significativement meilleurs que les enfants non musiciens dans la discrimination des contrastes vocaliques de l'anglais et ce, pour les trois tests relatifs aux « Contrastes vocaliques » (contrastes NN, contrastes N/NN, diphtongues et triphthongues), bien que ces tests soient plus difficiles que les tests de prosodie. Ils semblent donc moins sensibles au phénomène de crible phonologique. Nous avons vu qu'une oreille musicienne se caractérise par l'utilisation d'une stratégie analytique dans le traitement des contours mélodiques et par une capacité à traiter la dimension harmonique (accords musicaux) qui met également en jeu un traitement de type analytique. Ce traitement semble se transférer à la perception des contrastes vocaliques puisque les enfants musiciens font preuve d'une meilleure discrimination des contrastes vocaliques de timbres proches.

Nous ne pouvons toutefois pas déterminer si ces capacités analytiques proviennent d'un meilleur entraînement général ou d'une réorganisation perceptive provoquée par la pratique musicale. Les travaux en neurosciences et en sciences cognitives font apparaître une latéralisation plus à gauche de la musique chez les musiciens ayant l'oreille absolue, alors que l'hémisphère gauche est précisément dédié au traitement de type analytique de l'information auditive. Par ailleurs, le traitement des voyelles serait également localisé dans l'hémisphère gauche. Il est donc probable qu'il existe un transfert de compétence des capacités analytiques qui caractérisent une oreille musicienne à la discrimination des contrastes vocaliques, tâche qui nécessite justement un traitement de type analytique. Ces questions ne peuvent toutefois être complètement élucidées et il serait intéressant de réaliser une étude en imagerie cérébrale avec des sujets musiciens et des sujets non musiciens afin de déterminer si la/les zone(s) activée(s) pendant le traitement des accords musicaux se superposent à/aux zone(s) activée(s) pendant le traitement des voyelles.

En ce qui concerne les productions des enfants, nous avons relevé que les musiciens utilisaient une stratégie de généralisation des particularités prosodiques de l'anglais. Il est possible qu'en raison de leur oreille entraînée, ils repèrent plus facilement ces caractéristiques mais qu'ils les produisent dans des énoncés où elles ne devraient pas apparaître, même en situation de répétition immédiate. Cette stratégie contribue à donner une couleur « anglaise » à leurs productions. Ainsi, les contours en cloche spécifiques de l'anglais n'apparaissent que sur les mots à l'initiale voisée sont généralisés aux mots ne possédant pas cette caractéristique. On relève la même stratégie en ce qui concerne les voyelles brèves de l'anglais qui ne devraient pas être concernées par le phénomène. Celles-ci sont produites avec beaucoup d'instabilité de timbre par les enfants musiciens comme s'ils voulaient généraliser l'instabilité caractéristique des voyelles longues et des diphtongues de l'anglais à des voyelles dont le timbre devrait pourtant être stable. De la même façon, la tendance à l'accentuation initiale en anglais est généralisée aux mots trisyllabiques de notre corpus qui ne devraient être accentués que sur la deuxième syllabe. Mais, ce comportement concerne également les enfants non musiciens, en raison peut-être du fait que l'accent est un phénomène facile à repérer.

Si la stratégie de généralisation semble caractéristique d'un apprenant possédant une oreille musicienne, ce n'est pas le cas de la stratégie d'exagération des éléments prosodiques, comme nous l'avions supposé au départ. Non seulement les contours mélodiques ne sont pas exagérés, mais ils sont « minorés » par rapport aux productions du locuteur natif et ce, pour les deux groupes. Il est probable que ce comportement apparaisse au niveau individuel et que notre approche par groupe ait contribué à le noyer dans nos résultats. Il serait utile pour compléter notre analyse, d'étudier les stratégies de type individuel utilisées par les enfants de notre population. Il est aussi possible qu'une certaine retenue empêche la plupart des enfants d'exagérer les contours mélodiques devant leurs camarades ou des tierces personnes. Il serait intéressant de mener le même type d'étude avec des populations plus jeunes, car il est possible que des enfants plus jeunes soient moins sensibles à cette gêne et qu'ils n'hésitent pas à jouer avec les éléments prosodiques de la langue.

Nous avons également relevé que les musiciens restituent plus facilement l'accent dans les mots bisyllabiques, en situation de répétition différée. La tâche est en effet plus difficile qu'en situation de répétition immédiate car elle fait appel à la mémoire à long terme.

Pour que la prononciation soit correcte, il faut que le stockage initial respecte la forme sonore du mot et que cette forme ne s'altère pas avec le temps, notamment au contact des représentations de la langue maternelle. Il est alors possible que les exercices pratiqués pendant la formation musicale contribuent à améliorer la qualité de ce « stockage », en améliorant la mémoire musicale.

Ces résultats justifient a priori notre intuition sur le rapprochement que l'on peut faire entre la musique et la langue et le transfert de compétence de l'un à l'autre. Ils permettent également de mettre en lumière au moins quatre points, pouvant s'avérer d'une grande importance dans la pédagogie de l'apprentissage des langues étrangères :

- il ne suffit pas de mettre un enfant au contact d'une production native pour qu'il produise des énoncés de qualité « native » et ce, à cause des interférences avec la langue maternelle qui interviennent aussi bien en situation de répétition différée qu'en situation de répétition immédiate ;
- l'enseignant de langue doit avoir une production si ce n'est native, en tout cas de très bonne qualité pour que l'enfant puisse stocker des énoncés dont la prononciation est authentique ;
- la formation de l'oreille permet une meilleure mémorisation sensorielle de la « trace sonore » (Llorca, 1992) des énoncés en L2, améliorant ainsi leur stockage en mémoire à long terme.
- la stratégie de « généralisation » des caractéristiques « distinctives » d'une langue étrangère constitue un indice sur la progression de l'apprentissage de la prononciation. En effet, si un apprenant utilise une telle stratégie, c'est qu'il a identifié les spécificités sonores de cette langue. Il conviendrait cependant de veiller à ce que ces exagérations ne se fixent pas, mais constituent seulement une étape au cours de l'apprentissage de la langue étrangère.

S'il est certain d'après les résultats de cette thèse qu'il existe un transfert de compétence de la formation musicale à la pratique d'une langue étrangère, il s'agit maintenant d'exploiter ce transfert dans l'enseignement des langues étrangères. Nombre de pédagogues ont compris toute la richesse qui pouvait être tirée du rapprochement entre la musique et la didactique des langues.

Ainsi, Ribière-Raverlat (1997) insiste sur la primauté de l'oreille et sur la sensibilisation à la musique de la langue maternelle et des langues étrangères. Rittaud-Huttinet (2000) souligne la grande parenté qui existe entre musique et prononciation et la met à profit dans une série d'activités fondées sur le jeu et les chansons (les chansons permettent non seulement de travailler sur des points de phonétique particuliers, mais également d'introduire des notions de grammaire et de travailler le vocabulaire). Llorca travaille plus spécifiquement sur le rythme. Elle introduit la pratique de la langue dans la pratique du spectacle à travers ce qu'elle a baptisé le « théâtre rythmique » et elle propose une méthode tirée de ces principes (Llorca, 1997). Dans sa thèse, Aubin (1996) a transposé les techniques pédagogiques utilisées par les professeurs d'instrument de musique à l'apprentissage des langues étrangères pour fonder la « didactique de la musique du français ». Enfin, certains orthophonistes incorporent des techniques musicales dans leurs pratiques de rééducation, comme Daubisse-Kattan et Chalot-Benincà (1995) qui utilisent la notation musicale et associent les rythmes verbaux et les rythmes musicaux dans la rééducation des enfants sourds.

Certains musiciens se sont également intéressés au transfert des techniques musicales dans la pratique des langues, notamment le pianiste de jazz Kaddouch¹, qui a créé une méthode d'éveil à l'apprentissage des langues étrangères à destination des enfants de 3 à 5 ans. Cette méthode est directement inspirée de son expérience de musicien et d'improvisateur. À l'inverse, la méthode créée par Susuky en 1945 s'inspire de l'acquisition de la langue maternelle pour l'apprentissage de la musique : l'enfant apprend de manière très précoce (3-4 ans) à écouter, à mémoriser et à reproduire des extraits musicaux, puis des œuvres musicales, avant de débiter l'apprentissage du solfège. Cet inventaire est loin d'être exhaustif, mais souligne la diversité et la richesse des applications qui sont tirées des similitudes entre la musique et la langue.

Il serait souhaitable d'étudier l'impact des techniques musicales avec des apprenants adultes. En effet, avec l'âge, les effets du crible phonologique se renforcent et il devient de plus en plus difficile d'accéder à une prononciation « authentique ». Une véritable « rééducation » aux sons de la langue étrangère s'impose donc, dont l'objectif est d'amener progressivement l'apprenant à prendre conscience de ses fautes.

¹ Disponible à l'adresse suivante : http://www.kaddouch-music.com/ecoles/language_school.html

Les techniques de formation de l'oreille musicienne peuvent se montrer précieuses dans un tel processus et la dimension ludique apportée par la musique amène les apprenants à se « dépasser » et à oser jouer avec les sons, comme le ferait naturellement un jeune enfant (Dodane, 2001 ; Dodane et Konopczynski, 2001). Il semble prometteur de les coupler aux techniques de visualisation du signal et notamment de la mélodie (Germain et Martin, 2000), qui permettent de soumettre à l'observation de l'apprenant les éléments musicaux de la langue, jusque là perçus, mais « non visibles ». Les « mouvements » que l'oreille doit accomplir pour passer du phénomène « perçu » à sa représentation visuelle sont en effet similaires à ceux qui sont accomplis en musique pour passer de l'écoute d'un extrait musical à sa visualisation sur une partition. Dans ce processus, les systèmes de transcription de la prosodie constituent également un outil d'un intérêt majeur (cf. Jones, Fonagy et Magdics,.... Chapitre I, 2.1.1.2.).

Cette conclusion ne vient donc pas clore un projet qui se dépasse de lui-même et nous pouvons rêver d'un matériel pédagogique qui passerait par l'utilisation de la musique en vue d'une véritable « éducation » (enfants jeunes) ou « rééducation » (apprenants plus âgés, dès 12 ans) de l'oreille grâce à des techniques d'inspiration musicale (techniques utilisées en solfège notamment) ; la notation musicale (écoute-notation-visualisation) et les techniques de production musicale seraient également exploitées (technique vocale ; chant / pratique instrumentale).

La méthode pourrait s'inspirer de certains principes exposés dans cette thèse :

- Elle accorderait la primauté à la formation de l'oreille, avant de commencer le travail sur la production ;
- Elle accorderait la priorité aux éléments musicaux dont la mise en place est nécessaire pour une bonne production du niveau segmental, mais cette priorité ne se ferait pas aux dépens des phonèmes (cf. Galazzi-Matasci et Pedoya, 1983). Le parallèle entre la musique et la prosodie pourrait être utilisé dans des sons « intermédiaires », comme les diphtongues par exemple.
- Elle se fonderait sur un catalogue d'interférences adapté à chaque couple « langue de départ- langue cible » (cf. le catalogue français-anglais figurant dans cette thèse, voir Volume II, annexe n°3.5.).

En débutant cette recherche, nous n'imaginions pas toute la richesse que le parallèle entre la langue et de la musique allait contribuer à nous révéler, ni la somme de travaux concernée par cette problématique. Ces cinq années de recherche ont véritablement constitué pour nous une découverte, découverte dans le sens où le rapprochement de la langue et de la musique a changé notre regard, y compris sur des phénomènes qui nous paraissaient jusque-là familiers et au terme de ce travail, nous prenons conscience qu'il nous reste encore un vaste monde à explorer.

« Un vrai voyage de découverte n'est pas de chercher de nouvelles terres, mais d'avoir un œil nouveau », Edgard Morin, Transversales Sciences Culture, n°29, 8.

Bibliographie générale

- Adamczewski, Henri et Keen, Denis (1973). *Phonétique et Phonologie de l'Anglais Contemporain*. Paris : Armand Colin, 253 p.
- Aiello, Rita and Sloboda, John A. (1994). *Musical Perceptions*. New-York: Oxford University Press, 304 p.
- Aimard, Paule (1996). *Les Débuts du Langage chez l'Enfant*. Dunod : Paris, 226 p.
- Andrews, M. W. and Dowling, Jay W. (1991). The developments of perception of interleaved melodies and control of auditory attention. In *Music Perception*, 8, 349-368.
- Andrews, Moya L. and Madeira, Shelley S. (1977). The assessment of pitch discrimination ability in young children. In *Journal of Speech and Hearing Disorders*, 42/2, 279-286.
- Apfelstadt, Hilary E. (1983). *An investigation of the effect of melodic perception instruction on the pitch discrimination and vocal accuracy of kindergarten children*. PhD, Université de Wisconsin-Madison, 1983.
- Armitage, S.E., Baldwin, B.A. et Vince, M.A. (1980). The foetal sound environment of sheep. In *Science*, 208, 1173-1174.
- Astesano, Corinne (2001). *Rythme et Acentuation en Français : Invariance et Variabilité*. Paris : L'Harmattan, 317 p.
- Aubin, Sophie (1996). *La Didactique de la Musique du Français : sa Légitimité, son Interdisciplinarité*. Thèse de doctorat, Université de Rouen, 495 p.
- Awake ! (1997). Do african drums really talk ? In *Watchtower Bible and Tract Society of New-York*, July 22, 21-23.
- Ayotte, J., Peretz, Isabelle, Rousseau, I., Bard, C. and Bojanowski, M. (2000). Pattern of music agnosia associated with middle cerebral artery artefact. In *Brain*, 123, 1926-1938.
- Baken, Ronald J. and Daniloff, Raymond G. (1991). *Readings in Clinical Spectrography of Speech*. San Diego: Singular, 566 p.
- Balzano, G. J. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. In *Computer Music Journal*, 4, 66-84.
- Barraud, Henry (2002). La mélodie. Dans *Encyclopedia Universalis en ligne*.
- Bartlett, J. C. et Dowling, Jay W. (1980). The recognition of transposed melodies: a key-distance effect. In *Journal of Experimental Psychology: Human Perception and Performance*, 6, 501-515.
- Bates, Elizabeth, Dale, Philips and Thal, Donna (1995). Individual differences and their implications for theories of language development. In Fletcher and Mac Whinney, 96-151.
- Beaumont et Desmoinaux (2000). *J'apprends, je dessine, je joue : Les Couleurs*. Ed. Cerf Volant, 10 p.
- Beckman, Mary (1992). Evidence for speech rhythms across languages. In Tohkura, Vatikotis-Bateson and Sagisaka, 457-463.
- Beier, U. (1976). The talking drum of the Yoruba. In Umiker and Sebeok, 690-692.
- Benedict, H. (1979). Early lexical development: comprehension and production. In *Journal of Child Language*, 6, 183-200.
- Bentley, A. (1966). *Musical Ability in Children and its Measurement*. Londres: Harrap.
- Benveniste, Emile (1951). *La Notion de " Rythme " dans son expression linguistique*. Paris : P.U.F., 402 p.
- Bergeson, Tonya R. et Trehub, Sandra E. (1999). Mothers' singing to infant and preschool children. In *Infant Behavior and Development*, 22, 1, 51-64.
- Berio, Luciano (1998). *Sequenzas*. Livret d'accompagnement du CD, Deutsche Grammophon (457 038-2). Ensemble Intercontemporain.
- Bernstein Ratner, Nan (1986). Durational cues which mark clause boundaries in mother-child speech. In *Journal of Phonetics*, 14/2, 303-309.
- Bertoncini, Josiane (1993). Initial capacities for speech processing: infants' attention to prosodic cues for segmentation. In De Boysson-Bardies, De Schonen, Jusczyk, MacNeilage and Morton, 249-257.

- Bertoncini, Josiane et de Boysson-Bardies, Bénédicte (2000). La perception et la production de la parole avant deux ans. Dans Kail et Fayol, 95-136.
- Best, Catherine T. (1993). Learning to perceive the sound pattern of English. In Rovee-Collier and Lipsitt,
- Best, Catherine, T. (1993). Emergence of language-specific constraints in perception of non-native speech a window on early phonological development. In Boysson-Bardies, Schonen, Jusczyk, MacNeilage and Morton, 289-304.
- Best, Catherine T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In Goodman and Nusbaum, 167-224.
- Best, Catherine T., McRoberts, Gerald W. and Sithole, Nomathemb N. (1988). The phonological basis of perceptual loss of nonnative contrasts: maintenance of discrimination among Zulu clicks by english-speaking adults and infants. In *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345-360.
- Bever, Thomas G. and Chiarello, R. J. (1974). Cerebral dominance in musicians and non musicians. In *Science*, 185/537, 137-139.
- Bigand, Emmanuel (1994). Contributions de la musique aux recherches sur la cognition auditive humaine. Dans Mac Adams et Bigand, 249-298.
- Bigand, Emmanuel, Tillmann, B., Poulin, B., D'Adamo, D. A. and Madurell (2001). The effect oh harmonic context on phoneme monitoring in vocal music. In *Cognition*, 81, B11-B20.
- Bijeljac-Babic, Ranka (2000). Acquisition de la phonologie et bilinguisme précoce. Dans Kail et Fayol, 169-192.
- Bloom, B. S. (1985). *Developing Talent in Young People*. New-York : Bellentine Books.
- Bolinger, Dwight L. (1978). Intonation across languages. In Greenberg, Ferguson, and Moravcsik, 471-524.
- Bolinger, Dwight L. (1998). Intonation in American English. In Hirst and Di Cristo, 45-55.
- Bornstein, Marc H., Kessen, W. and Weiskopf, S. (1976). Color vision and hue categorisation in young human infants. In *Journal of Experimental Psychology: Human Perception and Performance*, 2, 115-129.
- Bot, K. de (1986). The transfer of intonation and the missing data base. Kellerman, E. and Smith, M. S., Crosslinguistic Influence. In *Second Language Acquisition*, 110-119.
- Boulez, Pierre (1963). *Dire, Jouer, Chanter. La Musique et ses Problèmes Contemporains*. Paris : Juillard.
- Boysson-Bardies, Bénédicte de (1993). Ontogeny of language-specific syllabic productions. In Boysson-Bardies, Schonen de, Jusczyk, Mac Neilage and Morton, 353-363.
- Boysson-Bardies, Bénédicte de, Sagart, Laurent and Durand, Catherine (1984). Discernible differences in the babbling of infant according to target language. In *Journal of Child Language*, 11, 1-15.
- Boysson-Bardies, Bénédicte de, Hallé, Pierre A., Sagart, Laurent and Durand, Catherine (1989). A cross-linguistic investigation of vowel formants in babbling. In *Journal of Child Language*, 16, 1-17.
- Boysson-Bardies, Bénédicte de and Vihman, Marylin M. (1991). Adaptation to language: evidence from babbling and first words in four language. In *Journal of Child Language*, 67, 297-319.
- Boysson-Bardies, Bénédicte de, Schonen, Scania, Jusczyk, Peter, MacNeilage, Peter and Morton, J. (1993). *Developmental Neurocognition: speech and face processing in the first year of life*. Boston: Kluwer Academic Publishers.
- Bradshaw, J. and Nettleton, N. (1981). The nature of hemispheric specialization in man. In *The Behavioral and Brain Sciences*, 4, 51-91.
- Brand, M. (1985). Development and validation of the home musical environment scale for use at the early elementary level. In *Psychology of Music*, 13, 40-48.
- Bregman, Albert S. (1978). Auditory streaming is cumulative. In *Journal of Experimental Psychology: Human Perception and Performance*, 4, 380-387.
- Bregman, Albert S. (1981). Asking the " what for " question in auditory perception. In Kubovy and Pomerantz, 99-118.
- Bregman, Albert S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge : MIT Press.

- Bregman, Albert S. (1994). L'analyse des scènes auditives. Dans McAdams et Bigand, 11-39.
- Brelet, Gisèle (1949). *Le Temps Musical : Essai d'une Esthétique Nouvelle de la Musique*. Paris : P.U.F., 842 p.
- Broen, P. A. (1972). The verbal environment of the language-learning child. In *American Speech and Hearing Association Monographs*, 17 (décembre).
- Burnet, James (1776). *Of the Origine and Progress of Language*. Réédité en 1974. New-York: G. Olms, vol. III.
- Burns, E. M. and Ward, W. D. (1982). Intervals, scales and tunings. In Deutsch, 241-269.
- Busnel, Marie-Claire (1994). Is there prenatal culture ? In Gardner, Chiarelli, Gardner and Plooij, 285-314.
- Busnel, Marie-Claire (1997). Audition foetale et réactivité prénatale à la voix maternelle " adressée ". In Busnel et al., 35-49.
- Busnel, Marie-Claire, Daffos, Fernand, Dolto-Tolitch, Lecanuet, Jean-Pierre et Negri, Romana (1997). *Que savent les Fœtus ?* Ramonville Saint-Agne : Erès, 103 p.
- Busnel, René Guy and Classe, André (1976). *Whistled Language*. New-York: Springer, 117 p.
- Butterfield et O'Neill (1997). *Peekaboo, I know !* Instructional Fair TS Denison, 10 p.
- Carlson, Rolf, Friberg, Anders, Frydén, Lars, Granström, Björn et Sundberg, Johan (1989). La parole et l'exécution de la musique : parallèles et contrastes. Dans Mac Adams et Deliège, 551-571.
- Carton, Fernand (1974). *Introduction à la Phonétique Française*. Paris : Bordas (1ère édition), 253 p.
- Castell, K. C. (1982). Children's sensitivity to stylistic differences in " classical " and " popular " music. In *Psychology of Music*, Special Issue, 22-25.
- Caterette, E. C. and Friedman, M. P. (1976). *Handbook of Perception*, VII. New-York: Academic Press.
- Catford, John C. (1988). *A Practical Introduction to Phonetics*. Oxford University Press.
- Changeux, Jean-Pierre (1983). *L'Homme Neuronal*. Paris : Fayard, 419 p.
- Chapman, J. S. (1975). *The Relation between Auditory Stimulation of Short Gestation Infants and Their Gross Motor Limb Activity*. PHD, New-York University.
- Chion, Michel (1983). *Guide des Objets Sonores : Pierre Schaeffer et la Recherche Musicale*. Paris : Buchet-Chasnel, 187 p.
- Chomsky, Noam and Halle, Morris (1968). *The Sound Pattern of English*. New-York: Harper and Row, 470 p.
- Chouard, Claude-Henri (2001). *L'Oreille Musicienne*. Paris : Gallimard, 347 p.
- Christophe, Anne, Dupoux Emmanuel, Bertoncini, Josianne and Mehler, Anne (1994). Do infants perceive word boundaries ? An empirical study of the bootstrapping of lexical acquisition. In *Journal of the Acoustical Society of America*, 95, 1570-1580.
- Clarke, F. E. (1989). Considérations sur le langage et la musique. Dans Mac Adams et Deliège, 23-43.
- Cohen, A. and Baird, R. (1990). Acquisition of absolute pitch: The question of critical periods. In *Psychomusicology*, 9, 31-37.
- Cohen, A. J., Thorpe, Leigh A. and Trehub, Sandra, S. E. (1987). Infants' perception of musical relations in short transposed tone sequences. In *Canadian Journal of Psychology*, 41, 33-47.
- Cohen, Antonie and Nooteboom, Sieb G. (1975). *Structure and Process in Speech Perception*. Berlin: Springer, 351 p.
- Cohen, Rachel (1997). *Plaidoyer pour les Apprentissages Précoces : Stratégies Educatives pour la Réalisation des Potentialités Humaines*. Paris : P.U.F., 317 p.
- Condillac (1746). *Essai sur l'Origine des Connaissances Humaines*. Paris : Armand Colin, ed. de 1924, 225 p.
- Cooke, Deryck (1959). *The Language of Music*. Londres: Oxford University Press.
- Cooper, G. and Meyer, L. B. (1960). *The Rhythmic Structure of Music*. University of Chicago Press.
- Cooper, R. P., & Aslin, Richard N. (1990). Preference for infant-directed speech in the first month after birth. In *Child Development*, 66, 1584-1595.
- Cooper, W. E. and Paccia-Cooper (1980). *Syntax and Speech*. Cambridge: Harvard University Press.

- Cowan, Georges M. (1948). Mazateco whistled speech. In *Language*, 24/3, 280-286.
- Cruttenden, Alan (1986). *Intonation*. Cambridge: Cambridge University Press, 214 p.
- Crystal, David (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press, 381 p.
- Crystal, David (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press, 472 p.
- Cuddy, L. L., Cohen, Anthony J. and Miller, Joanne (1979). Melody recognition: The experimental application of musical rules. In *Canadian Journal of Psychology*, 33, 148-157.
- Cutting, James E. and Rosner, B. S. (1974). Categories and boundaries in speech and music. In *Perception and Psychoacoustics*, 16, 564-570.
- Dalgalian, Gilbert, Lieutand, S. and Weiss, F. (1981). *Pour un Nouvel Enseignement des Langues et une Nouvelle Formation des Enseignants*. Paris : Clé International, 143 p.
- Daubisse-Kattan, Rita et Chalot-Beninc, Florence (1995). Utilisation de la musique pour l'élaboration et la structuration du langage oral de l'enfant sourd. Dans *Glossa*, 43, 36-41.
- Dauer, Rebecca M. (1983). Stress-timing and syllable-timing reanalysed. In *Journal of Phonetics*, 11, 51-62.
- Davidson, L. (1994). Songsinging by young and old: a developmental approach to music. In Aiello and Sloboda, 99-130.
- DeCasper, A. J. and Fifer, W. P. (1980). Of human bonding : Newborns prefer their mothers' voices. In *Science*, 208, 1174-1176.
- DeCasper, A. J. and Prescott, P. A. (1984). Human newborns' perception of male voices: preference, discrimination, and reinforcing value. In *Developmental Psychobiology*, 17, 481-491.
- DeCasper, A. J. and Spence, M. J. (1986). Prematernal speech influence newborns' perception of speech sounds. In *Infant Behavior and Development*, 9, 133-150.
- DeCasper, A.J., Lecanuet, Jean-Pierre, Busnel, Marie-Claire, Granier-Deferre, Carolyn et Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. In *Infant Behavior and Development*, 17, 159-164.
- Dehaene-Lambertz, Ghislaine, Christophe, Anne et Van Ooijen, Britt (1997). Bases cérébrales de l'acquisition du langage. Dans Kail et Fayol, 61-93.
- Delattre, Pierre (1961). La leçon d'intonation de Simone de Beauvoir. Etude d'intonation déclarative comparée. Dans *French Review*, 35/1, 59-67.
- Delattre, Pierre (1965). *Comparing the Phonetic Features of English, French, German and Spanish*. Heidelberg: Julius Gross Verlag, 118 p.
- Delattre, Pierre (1969). L'intonation par les oppositions. Dans *Le Français dans le Monde*, 64, 6-13.
- Delcloque, Philippe (1999). *Speech Technology Applications in CALL*. Dundee: University of Albertay Dundee, 87 p.
- Deliège, Irène et Sloboda, John A. (1995). *Naissance et Développement du Sens Musical*. Paris : PUF, 307 p.
- Demany, Laurent (1982). Auditory stream segregation in infancy. In *Infant Behavior and Development*, 5, 261-276.
- Demany, Laurent, McKenzie, B. and Vurpillot, Eliane (1977). Rhythm perception in early infancy. In *Journal of Acoustic Society of America*, 76, 57-66.
- Deutsch, D. (1969). Music recognition. In *Psychological Review*, 76, 300-307.
- Deutsch, D. (1982). *The Psychology of Music*. New-York: Academic Press.
- Di-Cristo, Albert (1975). *Soixante et Dix Ans de Recherches en Prosodie : Bibliographie Alphabétique, Thématique et Chronologique*. Aix-en-Provence : Editions de l'Université de Provence, Travaux de l'Institut de Phonétique d'Aix-en-Provence, 351 p.
- Di-Cristo, Albert (1998). Intonation in French. In Hirst et Di Cristo, 195-218.
- Di Cristo, Albert (1999). Le cadre accentuel du français contemporain : essai de modélisation. Dans *Langues*, 3/2, 184-205.

- Didier, Béatrice (1982). Condillac et la Musique. Dans Sgard, 395-405.
- Dodane, Christelle (1996). *Traitement Comparé du Langage et de la Musique : Contribution de la Musique au Développement Linguistique de l'Enfant*. Mémoire de Maîtrise, Université de Franche-Comté, Besançon.
- Dodane, Christelle (1997). *La Langue en Harmonie*. Diplôme d'Etudes Approfondies. Université de Franche-Comté, Besançon.
- Dodane, Christelle (2000). L'apprentissage précoce d'une langue étrangère : une solution pour la maîtrise de l'intonation et de la prononciation. In Guimbretière, 229-248.
- Dodane, Christelle (2001). Les nouvelles technologies multimédias dans l'apprentissage de la prosodie d'une langue étrangère : une révolution dans la pratique ? *Actes du VIIe Congrès International de Communication Sociale*, Santiago de Cuba, Vol. I, 176-182.
- Dodane, Christelle, Konopczynski, Gabrielle, Santi, Serge et Espesser Robert (1999). Fo declination line: more evidence from infant's and children's speech (9-48 months). *Actes du Colloque ICPH's*, San Francisco, 2161-2164.
- Dodane, Christelle et Konopczynski, Gabrielle (2001). " Utilisation des technologies multimédias dans la correction prosodique de l'adulte (L2) ". Dans Küpers et Souchon, 37-48.
- Dom de Malherbe (1946). *Le Chant Grégorien*. Paris : Lyra Dei.
- Dom Gajard, Joseph (1951). *La Méthode de Solesmes : ses Principes Constitutifs, ses Règles Pratiques d'Interprétation*. Paris : Desclée et Cie, 94 p.
- Dowling, Jay W. (1978). Scale and contour: two components of a theory for melodies. In *Psychological Review*, 85, 341-354.
- Dowling, Jay W. (1994). La structuration mélodique. Dans Zenatti, 145-176.
- Dowling, Jay W. and Harwood, Dane L. (1986). *Music Cognition*. Orlando: Academic Press, 258 p.
- Dowling, Jay W. (1982). Melodic information processing and its development. In Deutsch, 413-429.
- Drake, Carolyn (2002). Temporal organisation of sound sequences: Same processes in music, environmental scenes and speech ? *ISCA Workshop "Temporal Integration in the Perception of Speech"*, 8-10 avril, Aix en Provence.
- Drake, Carolyn and Penel, Amandine (1999). Learning to play music: rhythm and timing. In *Parole*, 9/10, 49-62.
- Drake, Carolyn, Penel, Amandine and Bigand, Emmanuel (2000a). Tapping in time with mechanically and expressively performed music. In *Music Perception*, 18, 1-24.
- Drake, Carolyn, Jones, Riess, Mari and Baruch, Clarisse (2000b). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. In *Cognition*, 77, 251-288.
- Eimas, Peter D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. In *Perception and Psychophysics*, 16, 513-521.
- Eimas, Peter D. (1975). Auditory and phonetics coding of the cues for speech: discrimination of the r-l distinction by young infant. In *Perception and Psychophysics*, 18, 341-347.
- Elbert, T. Pantev, C., Wienbruch, C., Rockstroh, B. and Taub, E. (1996). Increased cortical representation of the fingers of the left hand in string player. In *Science*, 270, 305-307.
- Encyclopædia Universalis (2002). *Base de données Universalis en ligne* (consultée en février 2002) : <http://www.universalis-edu.com>.
- Essien, Jimmy Akpan (2000). *Contribution à la Recherche sur la Perception des Tons du Yoruba : Evidences Expérimentales à partir des Tambours, des Signaux de la Parole et de la Synthèse*. Thèse de Doctorat, Université Paris III, non publiée, 426 p.
- Fassbender, Christoph (1993). *Auditory Grouping and Segregation Processes in Infancy*. Norderstedt: Kaste Verlag.
- Fassbender, Christoph (1995). La sensibilité auditive du nourrisson aux paramètres acoustiques du langage et de la musique. In Deliège et Sloboda, 63-99.
- Faure, Georges (1948). *Manuel Pratique d'Anglais Parlé*. Paris : Hachette, 211 p.

- Faure, Georges (1962). *Recherches sur les Caractères et le Rôle des Eléments Musicaux dans la Prononciation Anglaise (Essai de Description Phonologique)*. Thèse de Doctorat d'Etat. Paris : Didier, 380 p.
- Favard, Jean (1992). *Les langues étrangères à l'école primaire : la problématique française. Dans L'Apprentissage des Langues Etrangères à l'Ecole Primaire*. Paris : Didier.
- Feagans, C. Garvey, R. Golinkoff, M. T. Greenberg, C. Harding, & J. Bohannon (1984). *The Origins and Growth of Communication*. Norwood: New Jersey, Ablex Publishing Corporation.
- Feijoo, Jean (1975). Ut conscientia Noscatur. Dans *Cahiers de Soprologie*, 13, 14-20.
- Feijoo, Jean (1983). Le Fœtus, Pierre et le loup. In Herbinet et Busnel, 193-206.
- Ferguson, Charles A. (1977). Baby talk as a simplified register. In Snow and Ferguson, 219-235.
- Fernald, Anne (1976). The mother's speech to the newborn. Paper presented at the Max-Planck Institute for Psychiatry, Munich (non published).
- Fernald, Anne (1984). The perceptual and affective salience of mothers' speech to infants. In Feagans, Garvey, Golinkoff, Greenberg, Harding & Bohannon, 5-29.
- Fernald, Anne (1985). Four-month-old infants prefer to listen to motherese. In *Infant Behavior and Development*, 8, 181-195.
- Fernald, Anne (1989). Intonation and communication intent in mother's speech to infants: is the melody the message ? In *Child Development*, 60, 1497-1510.
- Fernald, Anne and Kuhl, Patricia (1981). Fundamental frequency as an acoustic determinant of infant preference for " Motherese ". Boston: Paper presented at the *Biennial Meeting of the Society for Research in Child Development*.
- Fernald, Anne and Kuhl, Patricia (1987). Acoustic determinants of infant preference for motherese speech. In *Infant Behavior and Development*, 10, 279-293.
- Fernald, Anne and Mazzie, C. (1991). Prosody and focus in speech to infants mothers' speech to new-borns. In *Developmental Psychology*, 27, 209-221.
- Fernald, Anne and Simon, Thomas (1984). Expanded intonation contours in mothers' speech to newborns. In *Developmental Psychology*, 20/1, 104-113.
- Fisher, Cynthia L. and Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: cross-linguistic evidence. In *Child Development*, 67, 3192-3218.
- Flege, James E. (1989). Using visual information to train foreign-language vowel production. In *Language Learning*, 38, 365-407.
- Fletcher, Paul and Mac Whinney, Brian (1995). *The Handbook of Child Language*. Cambridge: Blackwell.
- Fonagy, Yvan and Léon, Pierre (1980). L'Accent en Français Contemporain. Dans *Studia Phonetica*, 15.
- Fonagy, Yvan (1980). L'accent en français : accent probabilitaire. Dans Fonagy et Léon, 123-233.
- Fonagy, Yvan and Magdics, Klara (1963). Emotional patterns in intonation and music. In *Zeitschrift für Phonetik, Sprach-Wissenschaft und Kommunikations-Forschung*, 16 (1-3), 293-326.
- Fonagy, Yvan, Bérard, Eva et Fonagy, Judith (1983). Clichés mélodiques. Dans *Folia Linguistica*, XVII, 153-185.
- Fonagy, Yvan, Guzman, Marina et Bérard, Eva (1976). Comment mesurer l'accent d'intonation ? Dans *Travaux de l'Institut d'Etudes Linguistiques et Phonétiques*, 2, 41-61.
- Fraisse, Paul (1974). *La Psychologie du Rythme*. Paris : PUF, 244 p.
- Francès, Robert (1958). *La Perception de la Musique*. Paris : Vrin, 408 p.
- Fréron, Alain (2002). Matériau. *Encyclopedia Universalis en ligne*.
- Fridman, R. (1974). Los comienzos de la conducta musical. In *Biblioteca de Psicología Evolutiva Serie 2* (Vol. 13). Buenos Aires: Paidós.
- Gagnard, Madeleine (1987). *La Voix dans la Musique Contemporaine et Extra-Européenne*. Fondettes : Van de Velde, 117 p.

- Gajewski, Boreslas (1902). *Grammaire du Solr  sol ou Langue Universelle de Fran  ois Sudre*. Paris, 44 p. (disponible    l'adresse suivante : <http://www.ptialaska.net/~srice/solresol/intro.htm>).
- Galazzi-Matasci, Enrica et Pedoya, Elisabeth (1983). Et la p  dagogie de la prononciation ? Dans *Le Fran  ais dans le Monde*, 180, 39-44.
- Garab  dian, Mich  le (1996). Apprendre les langues   trang  res le plus t  t possible    l'  cole primaire : pourquoi ? Pourquoi faire ? Dans *Les Langues    l'Ecole, un Apprentissage ?* Actes du Colloque IUFM, 13 et 14 mars, Dijon, 27-37.
- Gardiner, Martin F, Fox, Alan, Knowles, Faith and Jeffrey, Donna (1996). Learning improved by arts training. In *Nature*, 381/6580, 284.
- Gardner, H. (1973). Children's sensitivity to musical styles. In *Merrill-Palmer Quaterly*, 19, 67-7.
- Gardner, R.A., Chiarelli, B.T., Gardner, B.T. and Plooij, F.X. (1994). *The Ethological Roots of Culture*. Kluwer Acad. Publish. Dordrecht Netherlands.
- Gates, A. and et Bradshaw, J. (1977). Music perception and cerebral asymmetries. In *Cortex*, 13, 390-401.
- Gathercole, Susan E. and Baddeley, Alan D. (1993). *Working memory and language*. Hove: Lawrence Erlbaum Associates.
- Gentil-Neymon, V. (1993). *Influence de la Voix de la M  re aupr  s des Pr  matur  s en Service de Soins Intensifs*. Th  se de Doctorat de M  decine, Universit   Paris VII.
- Germain, Aline et Martin, Philippe (2000). Pr  sentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde. Dans *ALSIC*, 3/1, 61-75.
- Geschwind, N. and Levitsky, W. (1968). In *Science*, 161 (186).
- Giacomi, A. et De Heredia, C. (1986). R  ussites et   checs dans les communications linguistiques entre locuteurs francophones et locuteurs immigr  s. In *Language*, 84, d  cembre.
- Ginesy, Michel (1995). *M  mento de Phon  tique Anglaise*. Paris : Nathan Universit  , 288 p.
- Girard, Denis (1991). Bilan didactique qualitatif des exp  riences d'enseignement pr  coce d'une langue   trang  re 1960-80. Dans *Les Langues Vivantes    l'Ecole El  mentaire*, INRP.
- Girard, Denis (1995). *Enseigner les Langues : M  thodes et Pratiques*. Paris : Bordas, 175 p.
- Gleitman, Lila R. and Wanner, E. (1982). *Language Acquisition: the State of the Art*. Cambridge. Cambridge: Cambridge University Press.
- Gleitman, Lila R., Gleitman, Harry, Landau, Barbara and Wanner, E. (1988). Where the learning begins: initial representations for language learning. In Newmeyer.
- Goodman, Judith C. and Nusbaum, Howard C. (1994). *The Development of Speech Perception: the Transition from Speech Sounds to Spoken Words*. Cambridge: MIT press.
- Gordon, Edwin E. (1975). Fourth-year and fifth-year final results of a longitudinal study of the musical achievement of culturqllly-disadvantaged students. In *Experimental Research in the Psychology of Music*, 10, 24-52.
- Gordon, Edwin E. (1979). Developmental music aptitude as measures by the primary measures of music audition In *Psychology of Music*, 7/1, 42-49.
- Greenberg, J. P., Ferguson, Charles A. and Moravcsik, E. A (1978). *Universals of Human Language*. Standford: Stanford University Press.
- Grieser, and Kuhl, Patricia K. (1989). Categorization of speech by infants: support for speech-sound prototypes. In *Developmental Psychology*, 25, 577-588.
- Grosjean, Fran  ois (1983). How long is the sentence ? Prediction and prosody in the on-line processing of language ? In *Linguistics*, 21, 501-529.
- Grosjean, Fran  ois et Hirt, Cendrine (1996). Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects. In *Language and Cognitive Processes*, 11 (1-2), 107-134.
- Grover, C., Jamieson, D. G. and Dobrovolsky, M. B. (1987). Intonations in English, French and German. In *Perception and Speech*, 30/3, 277-295.

- Guaïtella, Isabelle (2000). Les systèmes de notation de la prosodie et du geste : pourquoi et comment les utiliser ? Dans Guimbretière, 165-184.
- Guberina, Pierre (1991). Rôle de la perception auditive dans l'apprentissage précoce des langues. Dans *Le Français dans le Monde*, août-septembre, 65-70.
- Guierre, Lionel (1984). *Drills in English Stress-Patterns*. Paris : A. Colin-Longman, 127 p.
- Guimbretière, Elisabeth (1994). *Phonétique et Enseignement de l'Oral*. Paris : Didier / Hatier, 96 p.
- Guimbretière, Elisabeth (2000). *Apprendre, Enseigner, Acquérir : la Prosodie au Cœur du Débat*. Rouen : Publications de l'Université de Rouen, 312 p.
- Guimbretière, Elisabeth (2000). Entretien avec Ivan Fonagy : Mes rencontres avec l'intonation. Dans Guimbretière, 151-163.
- Hagège, Claude (1996). *L'Enfant aux Deux Langues*. Paris : Odile Jacob, 298 p.
- Halford, Brigitte K. and Pilch, Herbert (1994). *Intonation*. Tübingen: Gunter Narr Verlag.
- Hallé, Pierre A. and De Boysson-Bardies, Bénédicte de (1994). Emergence of an early receptive lexicon: infants' recognition of words. In *Infant Behavior and Development*, 17, 119-129.
- Handel, Stephen (1989). *Listening: an Introduction to the Perception of Auditory Events*. Cambridge: MIT Press, 597 p.
- Hardcastle, William J. and Laver, John (1997). *The Handbook of Phonetic Sciences*. Padstow: Blackwell Publishers, 904 p.
- Hargreaves, David (1995). Développement du sens artistique et musical. Dans Deliège et Sloboda, 167-197.
- Hawks, John W. (1994). Difference limens for formant patterns of vowel sounds. In *Journal of Acoustical Society of America*, 95/2, 1074-1084.
- Hepper and Shaidullah, (1994). Developmental of fetal hearing. In *Arch Dis Childhood*, 71, F81-F87.
- Hepper, P. G. (1991). An examination of fetal learning before and after birth. In *Irish Journal of Psychology*, 12, 95-107.
- Hepper, P. G., Scott, D. and Shahidullah, S. (1993). Newborn and fetal response to maternal voice. In *Journal of Reproductive and Infant Psychology*, 11, 147-153.
- Herbinet, Etienne et Busnel, Marie-Claire (1983). *L'Aube des Sens*. Paris : Stock, 414 p.
- Herzog, G. (1945). Drum signalling in a West African Tribe. In *Word*, 1, 217-238.
- Hirsh-Pasek, Kathryn, Kemler, Nelson, Deborah, Jusczyk, Peter W., Wright, Cassidy, K., Druss, B. and Kennedy, L. (1987). Clauses are perceptual units for young infants. In *Cognition*, 26, 269-286.
- Hirst, Daniel (1998). Intonation in British English. In Hirst et Di Cristo, 56-78.
- Hirst, Daniel and Di Cristo, Albert (1984). French intonation: a parametric approach. Dans *Die Neueren Sprachen*, 83/5, 554-569.
- Hirst, Daniel and Di Cristo, Albert (1998). A survey of intonation systems. Dans Hirst et Di Cristo, 1-44.
- Hirst, Daniel and Di Cristo, Albert (1998). *Intonation Systems: a Survey of Twenty Languages*. Cambridge: University Press, 487 p.
- Hugdahl, Kenneth (1988). *Handbook of dichotic listening: Theory, methods and research*. New-York: Wiley & Sons.
- Hulse, S. H., Humpal, J. and Cynx, J. (1984). Discrimination and generalization of rhythmic and arrhythmic sound patterns by European starlings. In *Music Perception*, 1, 442-464.
- Imagier en Anglais (1997). *Père Castor*. Paris : Flammarion, 258 p.
- Imberty, Michel (1969). *L'Acquisition des Structures Tonales chez l'Enfant*. Klincksieck : Paris, 226 p.
- Imberty, Michel (1995). Développement linguistique et musical. Dans Deliège et Sloboda, 223-249.
- Institut National de la Statistique et des Etudes Economiques (1994). *Nomenclature des Professions et Catégories Socioprofessionnelles PCS*. 2e éd. Paris : Imprimerie Nationale, 420 p.

- Institut National de la Statistique et des Etudes Economiques (2001). *Note d'Information 01-18*. Edu050, avril.
- Institut National de la Statistique et des Etudes Economiques (2001). *Annuaire Statistique de la France : résultats de 1999*. Paris : Imprimerie Nationale. 104ème Ed.
- Jackendoff, Ray S. (1991). Musical parsing and musical affect. In *Music Perception*, 9, 199-230.
- Jakobson, Roman et Waught, Linda, R. (1980). *La Charpente Phonique du Langage*. Paris : les Editions de Minuit, 336 p.
- Jespersen, Otto (1922). *Nature, Evolution et Origine du Langage* (Language: It's nature, development and origin). Première édition, Allen and Unwin, Ltd, London. Paris : Payot, 436 p.
- Johnson K., Ladefoged P. and Lindau M. (1993). Individual differences in vowel production. In *Journal of the Acoustical Society of America*, 94, 701-714.
- Johnson, E.L. and Newport, S.J. (1989). Critical period effects in second language learning : the influence of maturational state on the acquisition of english as a second language. In *Cognitive Psychology*, 21, 60-99.
- Johnson, E.L. and Newport, S.J. (1991). Critical period effect on universal properties of language : the status of subadjacency in the acquisition of a second language. In *Cognition*, 39, 215-258.
- Jones, Daniel (1909). *Intonation Curves*.
- Jones, Riess Marie and Boltz, M. (1989). Dynamic attending and responses to time. In *Psychological Review*, 96, 459-491.
- Jones, Riess Marie et Yee, William (1994). L'attention aux événements auditifs : le rôle de l'organisation temporelle. Dans Mc Adams et Deliège, 75-121.
- Journiac, Claire (2002). *La Place de l'Autre dans la Relation Soignante : Contribution à l'Etude des Processus Relationnels à Travers la Voix*. Université de Bourgogne, 360p.
- Jusczyk, Peter W. (1989). Perception of cues to clausal units in native and non-native languages. Paper presented at the *Biennial Meeting of the Society for Research in Child Development*, Kansas City, April.
- Jusczyk, Peter W. (1997). *The Discovery of Spoken Language*. Cambridge : MIT, 314 p.
- Jusczyk, Peter W. and Thompson, E. J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. In *Perception and Psychophysics*, 23, 105-109.
- Jusczyk, Peter W., Hirsh-Pasek, Kathryn, Kemler Nelson, Deborah G., Kennedy, L., Woodward, Amanda and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. In *Cognitive Psychology*, 24, 252-293.
- Jusczyk, Peter W., Friederici, A. D., Wessels, J., Svenkerud, V. Y. and Jusczyk, A. M. (1993a). Infants' sensitivity to the sound patterns of native language words. In *Journal of Memory and Language*, 32, 402-420.
- Jusczyk, Peter W. and Krumhansl, Carol (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. In *Journal of Experimental Psychology: Human Perception and Performance*, 19/3, 627-640.
- Kagan, Jerome and Lewis, Michael (1965). Studies of attention in the human infant. In *Merill-Palmer Quarterly*, 11, 95-122.
- Kail, Michèle (1994). Le Développement du langage et les sciences cognitives. Dans *Psychologie Française*, 39/1, 29-47.
- Kail, Michèle et Fayol, Michel (2000). *L'Acquisition du Langage : le Langage en Emergence de la Naissance à Trois ans*. Vol. I. Paris : PUF, 304 p.
- Kaiser, L. (1957). *Manual of Phonetics*. Amsterdam: North-Holland publishing company, 460 p.
- Katz, G. S., Cohn, J. F. and Moore, F. A. (1996). A combination of vocal fo dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. In *Child Development*, 67, 205-217.
- Kavanagh, James F. and Cutting, James E. (1975). *The Role of Speech in Language*. Cambridge: MIT press, 335 p.
- Kemler Nelson, D. G., Hirsh-Pasek, Kathryn, Jusczyk, Peter W. and Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. In *Journal of Child Language*, 16, 53-68.

- Kern, Sophie et Lange, Jacques (2000). Des premiers gestes aux premiers mots : le développement communicatif chez l'enfant de 8 à 30 mois. *Troisièmes Journées Scientifiques de l'Ecole d'Orthophonie*, Lyon, CD-Rom, 18 pages.
- Kewley-Port, Diane (1990). Thresholds for formant frequency discrimination in isolated vowels. In *Journal of Acoustical Society of America*, 87/1, S159.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. In *Psychological Bulletin*, 112, 24-38.
- Kimura, D. (1961a). Some effects of temporal-lobe damage on auditory perception. In *Canadian Journal of Psychology*, 15, 156-165.
- Kimura, D. (1961b). Cerebral dominance and the perception of verbal stimuli. In *Canadian Journal of Psychology*, 15, 166-171.
- Kirkpatrick, W. C. (1962). *Relationships between the singing ability of prekindergarten children and their home environment*. PhD, Univ. Southern California.
- Klatt, Dennis K. and Cooper, William E. (1975). Perception of segment duration in sentence contexts. In Cohen and Nooteboom, 69-86.
- Kleiber, Michel (1990). *La Sémantique du Prototype : Catégories et Sens Lexical*. Paris : P.U.F., 199 p.
- Kleist, K. (1959). *Sensorische aphasien und amusien auf myeloarchitektonischer grundlage*. Thieme Verlag, Stuttgart.
- Kluender, K. R., Diehl, R. L. and Killeen, P. R. (1987). Japanese quail can learn phonetic categories. In *Science*, 237, 1195-1197.
- Kolinski, M. (1967). Recent trends in ethnomusicology. In *Ethnomusicology*, 11, 1-24.
- Konopczynski, Gabrielle (1977). Jugement linguistique, jugement psycho-physique dans la perception de contours intonatifs : interrelations. Dans *Travaux de L'Institut de Phonétique de Strasbourg*, 9, 205-220.
- Konopczynski, Gabrielle (1978). Un essai original de transcription musicale de la prosodie. Wilhelm Wundt : Völkerpsychologie (1900-1912). Dans *Travaux de L'Institut de Phonétique de Strasbourg*, 10, 82-97.
- Konopczynski, Gabrielle (1981a). Compréhension des structures intonatives et de leur fonction linguistique par des enfants de 6 à 11 ans : un exemple d'expérimentation. Dans *Working Papers of Speech and Voice Society*, Toronto, 2, 50-90.
- Konopczynski, Gabrielle (1981b). Remarques sur la perception des structures rythmiques et mélodiques par les enfants de 6 à 11 ans. Toronto, *Symposium Prosodie*, 61-75.
- Konopczynski, Gabrielle (1990). *Le Langage Emergent : Caractéristiques Rythmiques*. Hamburg : Buske Verlag, 362 p.
- Konopczynski, Gabrielle (1991). *Le Langage Emergent II : Aspects Vocaux et Mélodiques*. Hambourg : Buske Verlag, 425 p.
- Konopczynski, Gabrielle (1991). Acquisition de la proéminence dans le langage émergent. Dans *Actes du XIIIème Congrès International des Sciences Phonétiques*. Aix-en-Provence, Vol I, 333-337.
- Konopczynski, Gabrielle (1999). L'acquisition du système prosodique de la langue maternelle et ses implications pour l'apprentissage d'une L2. Dans Delcloque Philippe, 62-70.
- Konopczynski, Gabrielle (2001). Maturation syntaxique chez le jeune enfant (9-36 mois). Continuité ou rupture ? Neuchâtel : Tranel, *Actes du Colloque " Le Changement Linguistique "*, 34/35, 251-272.
- Konopczynski, Gabrielle (2002). *Linguistique de l'Acquisition : Etat des Lieux*. Cours de D.E.A. de Sciences du Langage, mention F.L.E. à distance, Université de Franche-Comté, 15 pages.
- Konopczynski, Gabrielle et Mac Carthy, Brian (1977). Les effets de la lecture à voix haute sur l'intonation. In *International Review of Applied Linguistic in Language Teaching*, 15/1, 230-238.
- Konopczynski, Gabrielle et Tessier, Sophie (1994). Structuration intonative du langage émergent. Dans Halford et Pilch, 157-192.
- Krashen, Stephen D. (1983). *The Natural Approach : Language Acquisition in the Classroom*. Englewood Cliffs (New-Jersey): Alemany Press, 191 p.

- Krumhansl, Carol L. (1983). Perceptual structures for tonal music. In *Music Perception*, 1, 28-62.
- Krumhansl, Carol L. Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions in music. In *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.
- Krumhansl, Carol L. and Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. In *Memory and Cognition*, 10, 243-251.
- Krumhansl, Carol L. and Jusczyk, Peter W. (1990). Infant's perception of phrase structure in music. In *Psychological Science*, 1, 70-73.
- Kubovy and Pomerantz (1981). *Perceptual Organization*. Hillsdale: Erlbaum Assoc.
- Kuhl, Patricia K. and Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. In *Perception and Psychophysics*, 31, 279-292.
- Kuhl, Patricia K. (1992). Infants' perception and representation of speech: Development of a new theory. In Ohala, Nearey, Derwing, Hodge and Wiebe, 449-456.
- Kuhl, Patricia K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In Boysson-Bardies de, Schonen, Jusczyk, MacNeilage and Morton, 259-274.
- Kuhl, Patricia K. (1994). Speech perception. In Minifie, 118-142.
- Kuhl, Patricia K. and Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. In *Science*, 190, 69-72.
- Kuhl, Patricia K. and Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for voicing feature in macaques. In *Perception and Psychophysics*, 32, 542-550.
- Kuhl, Patricia K., Williams, K. A., Lacerda, F., Stevens, K. N. and Lindblom, Björn (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. In *Science*, 255, 606-608.
- Kuhl, Patricia K., Andruski, J.E., Chistovitch, Inna A., Chistovitch, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U. and Lacerda, F. (1997). Cross-language analysis of phonetic in language addressed to infants. In *Science*, 277, 684-686.
- Kwalwasser, J. and Dykema, P. W. (1930). *Music Tests. Manual of Direction*. New-York: Fisher.
- Lacheret-Dujour Anne et Beaugendre Frédéric (1999). *La Prosodie du Français*. Paris : C.N.R.S. Langage, 353 p.
- Ladd, Robert (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladefoged (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovitch, Inc., 296 p.
- Ladefoged, Peter (1997). Linguistic Phonetic Descriptions. In Hardcastle and Laver, 589-618.
- Lado, Robert (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, Michigan, 141 p.
- Lanchec, Jean-Yvon (1976). *Psycholinguistique et Pédagogie des Langues*. Paris : P.U.F., 164 p.
- Landercy, Albert et Renard, Raymond (1977). *Eléments de Phonétique*. Bruxelles : Didier, 214 p.
- LeBlanc, A. (1991). Effect of maturation/aging on music listening preference: a review of the literature. Communication au 9th National Symposium on Research in Musical Behavior, Cannon Beach, Oregon, USA.
- Lecanuet, Jean-Pierre (1995). L'expérience auditive prénatale. In Deliège et Sloboda, 7-38.
- Lecanuet, Jean-Pierre (1997). Dans tous les sens... bref état des compétences sensorielles fœtales. In Busnel et al., 17-34.
- Lecanuet, Jean-Pierre, Jacquet, A-Y., Hains S.M.J. et Kisilevsky, Barbara S. (2000). Effects of music on behaviour in term fetuses. *International Conference on Infant Studies* ; Brighton , Grande Bretagne, Juillet 3-6.
- Lecanuet, Jean-Pierre, Granier-Deferre, Carolyn, Jacquet, A. Y., Capponi, I. et Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. In *Early Development and Parenting*, 2, 217-228.
- Lecanuet, Jean-Pierre, Jacquet, Anne-Yvonne, Kisilevsky, Barbara, S. and Hains, Sylvia, M. J. (2000). Effects of music on behaviour in term fetuses. Poster, *ICIS 2000* (International Conference on Infant Studies), Brighton.

- LeCours, A.R. (1975). Myelogenetic correlates of the development of speech and language. In Lenneberg and Lenneberg.
- Lehiste Ilse and Peterson Gordon E. (1960). Transitions, glides, and diphthongs. In Baken and Daniloff (1991), 286-295.
- Leipp, Emile (1971). *Acoustique et Musique*. Paris : Masson. 3ème éd. (1980), 352 p.
- Lenneberg, Eric Heinz (1967). *Biological Foundations of Language*. New-York: Wiley & Sons, 489 p.
- Lenneberg, Eric Heinz and Lenneberg, Elizabeth (1975). *Foundations of Language Development of Speech and Language* (Vol. 1). New-York: Academic Press, 350 p.
- Lenz, S. M. (1978). A case study of the musical abilities of three-and four-year-old children. PhD, Univ. Illinois, 1978. Compte rendu de D. R. Tanner in *Bulletin of the Council for Research in Music Education*, 65, 43-46.
- Léon, Pierre (1992). *Phonétisme et Prononciations du Français*. Paris : Nathan Université, 192 p.
- Léon, Pierre et Léon, Monique (1964). *Introduction à la Phonétique Corrective : à l'Usage des Professeurs de Français à l'Etranger*. Paris : Hachette, 98 p.
- Léon, Pierre et Martin, Philippe (1969). *Prolégomènes à l'Etude des Structures Intonatives*. *Studia Phonetica*, n°2. Paris : Didier, 225 p.
- Léon, Pierre et Martin, Philippe (1971). Linguistique appliquée et enseignement de l'intonation. Dans *Etudes de Linguistique Appliquée*, 3, 36-45.
- Lerdahl, Fred (1985). Théorie générative de la musique et composition musicale. Quoi ? Quand ? Comment ? Dans *La Recherche Musicale*. Paris : Bourgeois/IRCAM, 101-120.
- Lerdahl, Fred and Jackendoff, Ray S. (1985). *A Generative Theory of Tonal Music*. Cambridge: M.I.T. Press, 368 p.
- Levarie, S. and Rudolph, N. (1978). Can newborn infants distinguish between tone and noise ? In *Perceptual and Motor Skills*, 47, 1123-1126.
- Lewis, Michael (1936). *Infant Speech: A Study of the Beginnings of Language*. New-York: Harcourt Brace.
- Lhote, Elisabeth (1995). *Enseigner l'Oral en Interaction*. Paris : Hachette, 158 p.
- Lhote, Elisabeth (1987). *A la Découverte des Paysages Sonores des Langues*. Besançon : Annales Littéraires de l'Université, 215 p.
- Liberman, Anatoly M., Cooper, F. S., Shankweiler, D. P. and Studdert-Kennedy, Michael G. (1967). Perception of the speech code. In *Psychological Review*, 74, 431-461.
- Lieberman, Philip (1963). Some effects of semantic and grammatical context on the production and perception of speech. In *Language and Speech*, 6, 172-187.
- Liégeois-Chauvel, C., Peretz, Isabelle, Babai, M and al (1998). Contribution of different cortical areas in the temporal lobes to music processing. In *Brain*, 121, 1853-1867.
- Lindblom, Björn and Sundberg, J. (1970). Towards a generative theory of melody. In *Svensk Tidskrift för Musikforskning, Swedish Journal of Musicology*, 52, 171-881.
- Lindblom, Björn, McNeilage, Peter and Studdert-Kennedy, Michael G. (1984).
- Lipsitt, Lewis P. and Rovee-Collier, Carolyn K. (1981). *Advances in Infancy Research*. Vol I : XXII. Norwood: Ablex, 267 p.
- Llorca, Régine (1987). Le paysage français pour un Français : les bases rythmiques du français. In Lhote, 47-58.
- Llorca, Régine (1992). Le rôle de la mémoire musicale dans la perception d'une langue étrangère. Dans *Revue de Phonétique Appliquée*, 102, 45-67.
- Lonchamp, François (1983). On Vowel Normalisation. *Leeds Experimental Phonetics Symposium*, Leeds.
- Luce, Paul A. and Charles-Luce, Jan (1983). Contextual effects on the consonant/vowel ratio in speech production. Paper presented at the *Meeting of the Acoustical Society of America*, Cincinnati, May.
- Luria, Alexander R. (1974). *L'enfant Retardé Mental*. Paris : Privat, 254 p.

- Lynch, Michael P. (1993). Prototypical representations of music structure in infancy: theoretical exploration and a pilot study. In *Psychomusicology*, 12, 31-40.
- Lynch, Michael P., Eilers, Rebecca E., Oller, Kimbrough D. and Urbano, Richard C. (1990). Innateness experience and music perception. In *American Psychological Society*, 1/4, 272-276.
- Lynch, Michael P. and Eilers, Rebecca, E. (1991). Children's perception of native and nonnative musical scales. In *Music Perception*, 9/1, 121-132.
- Lynch, Michael P., Eilers, Rebecca E., Oller, Kimbrough D. and Urbano, Richard C. (1991). Influences of acculturation and musical sophistication on perception of musical interval patterns. In *Journal of Experimental Psychology: Human Perception and Performance*, 17/4, 967-975.
- Lynch, Michael P. and Eilers, Rebecca, E. (1992). A study of perceptual developement for musical tunings. In *Perception and Psychophysics*, 52/6, 599-608.
- Mac Adams, Stéphane et Deliège, Irène (1989). *La Musique et les Sciences Cognitives*. Liège : Pierre Mardaga, 649 p.
- Mac Adams, Stéphane et Bigand, Emmanuel (1994). *Penser les Sons : Psychologie Cognitive de l'Audition*. Paris : PUF, 402 p.
- Maier, Wolfgang (1996). Théorie et pratique de l'enseignement primaire des langues vivantes. Dans Les Langues à l'Ecole, un Apprentissage ? *Actes du Colloque IUFM*, 13 et 14 mars, Dijon, 57-71.
- Malsheen, Bathsheb (1980). Two hypotheses for phonetic clarification in the speech of mothers-to-children. In Yeni-Komshian, Kavanagh and Ferguson, II, 173-184.
- Manturzevska, Maria (1974). Psychologiczne wyznaczniki powodzenia w studiach muzycznych [Facteurs psychologiques permettant un pronostic de réussite dans les études musicales], Varsovie, *COPSA*.
- Manturzevska, Maria (1990). A biographical study of the life-span development of professional musicians. In *Psychology of Music*, 18, 112-139.
- Manturzevska, Maria (1994). Les facteurs psychologiques dans le développement musical et l'évolution des musiciens professionnels. In Zenatti, 259-290.
- Marler, Peter R. (1969). Animals and Man: Communication and its Development. In Roslansky, 25-62.
- Marler, Peter R. (1977). The evolution of communication. In Sebeok, 45-70.
- Martin, Pierre (1996). *Eléments de Phonétique avec Application au Français*. Saint-Nicolas (Québec) : Presses Universitaires de l'Université de Laval, 253 p.
- Massin, Jean et Massin, Brigitte (1983). *Histoire de la Musique Occidentale*. Messidor-Temps Actuels. Ed. utilisée, 1998. Fayard, 1312 p.
- McAdams, Stephen et Bigand, Emmanuel (1994). *Penser les Sons : Psychologie Cognitive de l'Audition*. Paris : P.U.F., 402 p.
- Mehler, Jacques, Bertoncini, Josiane, Barrière, M., Jassik-Gershenfeld, Dora (1978). Infant recognition of Mother's voice. In *Perception*, 7, 491-497.
- Meigret, Louis (1550). *Le Tretté de la Gramèze Françoeze*. Nach der einzigen Pariser Ausgabe. Herausgegeben von Wendelin Foerster, Heilbronn, 1988.
- Menezes, Flo (1993). *Luciano Berio et la Phonologie : une Approche Jakobsonienne de son Oeuvre*. Berlin : Peter Lang, Publications Universitaires Européennes, 278 p.
- Menyuk, Paula (1977). *Language and Maturation*. Cambridge: MIT Press, 180 p.
- Menyuk, Paula (1994). Word acquisition: the important role of prosody. In *The Study of Sounds*, 23, 218-222.
- Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. In *Journal of Acoustical Society of America*, 85, 2114-2134.
- Meschonnic, Henri (1982). *Critique du Rythme*. Paris : Verdier, 713 p.
- Mettas, Odette (1971). *Les Techniques de la Phonétique Instrumentale et l'Intonation*. Paris : P.U.F., 166 p.
- Michon, John A. (1967). Timing in Temporal Tracking. *Soesterberg: Institute for Perception RVO-TNO*, 127 p.

- Miller J. D. (1989). Auditory-perceptual interpretation of the vowel. In *Journal of Acoustical Society of America*, 85, 2114-2134.
- Miller, Joanne L. and Liberman, Anatoly M. (1979). Some effects of later occurring information on the perception of stop consonant and semivowel. In *Perception & Psychophysics*, 25: 457-465.
- Mingat, A. et Suchaut, B. (1996). Incidences des activités musicales en grande section de maternelle sur les apprentissages précoces au cours préparatoire. Dans *Les Sciences de l'Education*, 29, 3, 49-76.
- Minifie, Fred H. (1994). *Introduction to Communication Sciences and Disorders*. San Diego: Singular Publishing Group.
- Ministère de l'Education Nationale (2001). *Note d'information 01-18*, avril.
- Montreuil, Jean-Pierre (2001). *La Phonologie de l'Anglais*. Rennes : Presses Universitaires de Rennes, 168 p.
- Moog, H. (1976). *The Musical Experience of the Pre-school Child*. London: Scott.
- Moon, C., Cooper, R., and Fifer, W. P. (1993). Two-day-olds prefer their native language. In *Infant Behavior and Development*, 16, 495-500.
- Moore, D. L. (1973). *A study of pitch and rhythm responses of five-year-old children in relation to their early music training*. PhD, Florida, State Univ.
- Moreau, Marie-Louise et Richelle, Marc (1981). *L'Acquisition du Langage*. Bruxelles : Mardaga, 261 p.
- Morgan, James L., Meier, R. P. and Newport, Elissa L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language ? In *Cognitive Psychology*, 19, 498-550.
- Morse, Philip A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. In *Journal of Experimental Child Psychology*, 13, 477-492.
- Mosser, Christine (1990). *Effets physiologiques des Stimulations Sonores chez les Prématurés*. Thèse de Doctorat : Mondrini, Paris XII.
- Mugny, Gabriel, Oberlé, Dominique et Beauvois, Jean-Léon (1995). *La Psychologie Sociale : Relations Humaines, Groupes et Influence Sociale*. Tome I. Grenoble : P.U.G., 373 p.
- Naatanen, Risto, Lehtokoski, Anne, Lennes, Mietta, Cheour, Marie, Huotilainen, Minna, Iivonen, Antti, Vainio, Martti, Alku, Paavo, Ilmoniemi, Risto, Luuk, Aavo, Allik, Juri, Sinkkhonen, Janne and Alho, Kimmo (1997). Language-specific representations revealed by electric and magnetic brain responses. In *Nature*, 385 (6615), 432-434.
- Nakagawa, T., Saito, S. and Yoshino, T. (1982). Tonal difference limens for second formant frequencies of synthesized Japanese vowels. In *Ann. Bull. RILP*, 16, 81-88.
- Nakamura, Satoshi, Sadato, Norihiro, Oohashi, Tsutomu, Nishina, Emi, Fuwamoto, Yoshitaka, Yonekura Yoshiharu (1999). Analysis of music brain interaction with simultaneous measurement of regional cerebral blood flow and electroencephalogram beta rhythm in human subjects. In *Neuroscience Letters*, 275, 222-226.
- Navon, D. (1977). Forest before trees: the precedence of global feature in visual perception. In *Cognitive Psychology*, 9, 353-383.
- Nazzi, Thierry, Kemler Nelson, Deborah, Jusczyk, Peter and Jusczyk, A.M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. In *Infancy*, 1/1, 123-147.
- Neisser, U. (1967). *Cognitive Psychology*. New-York: Appleton Century Crofts.
- Newmeyer, Frederick J. (1988). *Linguistics: The Cambridge Survey*. Cambridge: Cambridge University Press, III, 349 p.
- Newport, Elissa L. (1990). Maturation constraints on language learning. In *Cognitive Science*, 14, 11-28.
- Ninio, A. (1992). The relation of children's world utterances to single word utterance in the input. In *Journal of Child Language*, 118, 87-110.
- Nowik-Stern, Alisa, Marsha G. Clarkson, Mary K. Morris, Roger Bakeman (1998). The effect of premature infants' speech preferences on mother-preterm interaction. Poster, *11th Biennial International Conference on Infant Studies*, Atlanta, Georgia, April 2-5.

- Ohala, J. J., Nearey, T. M., Derwing, B. L., Hodge, M. M. and Wiebe, G. E. *Proceedings of the International Conference on Spoken Language Processing*. Edmonton, Alberta: University of Alberta.
- Oller, Kimbrough D. (1980). The emergence of the sounds of speech in infancy. In Yeni-Komshian, Kavanagh and Ferguson, I, 93-112.
- Osmond-Smith, David (1989). Entre la musique et le langage : vue depuis le pont. La parole et l'exécution de la musique : parallèles et contrastes. Dans Mac Adams et Deliège, 137-148.
- Ostiguy, Luc, Sarasin, Robert et Irons, Glenwood (1996). *Introduction à la Phonétique Comparée : les Sons. Le Français et l'Anglais Nord-Américain*. Laval : Presses de l'Université de Laval, 200 p.
- Pallier, Christophe, Dehaene, Stanislas, Poline, S., LeBihan, D., Argenti, A-M, Dupoux, Emmanuel et Mehler, Jacques (soumis). Brain imaging of language plasticity in adopted adults: can a second language replace the first ? In *Cerebral Cortex*.
- Palmer, C. and Krumhansl, Carol L. (1990). Mental representation for musical meter. In *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.
- Panneton, R. K. (1985). *Prenatal auditory experience with melodies: Effects on postnatal auditory preferences in human newborns*. PhD, University of North Carolina at Greensboro.
- Pantev, Christo, Oostenveld, Robert, Engelien, Almut, Ross, Bernhard, Roberts, Larry, E. and Hoke, Manfred (1998). Increased auditory cortical representation in musicians. In *Nature*, 392/6678, 811-814.
- Papousek, Hanus (1995). Musicalité et petite enfance. Origines biologiques et culturelles de la précocité. In Deliège et Sloboda, 41-62.
- Papousek, Mechthild (1995). Le comportement parental intuitif, source cachée de la stimulation musicale dans la petite enfance. In Deliège et Sloboda, 101-130.
- Papousek, Mechthild and Papousek, Hanus (1981). Musical elements in the infants vocalizations: Their significance for communication, cognition and creativity. In Lipsitt and Rovee-Collier, 163-224.
- Papousek, Mechthild, Bornstein, Marc H., Nuzzo, C., Papousek, Hanus and Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. In *Infant Behavior and Development*, 13, 539-545.
- Papousek, Mechthild, Papousek, Hanus, & Symmes, D. (1991). The meaning of melodies in motherese in tone and stress languages. In *Infant Behavior and Development*, 14, 415-440.
- Pasdeloup Valérie (1991). *Modèles de Règles Rythmiques du Français Appliqué à la Synthèse de la Parole*. Thèse de doctorat, Université de Provence.
- Passy, Paul (1890). *Etudes sur les Changements Phonétiques et leurs Caractères Généraux*. Paris : Firmin-Didot, 270 p.
- Patel, A. and al. (1998) Processing prosodic and musical patterns: A neuropsychological investigation. In *Brain Language*, 61, 123-144.
- Penfield, Wilder and Roberts, Lamar (1959). *Speech and brain mechanisms*. Princeton: Princeton University Press.
- Penfield, Wilder and Roberts, Lamar (1963). *Langage et Mécanismes cérébraux*. Paris : PUF (trad. Fse texte 1959), 311 p.
- Peretz, Isabelle et Morais, José (1988). Determinants of laterality for music: Towards an information processing account. In Hugdahl,
- Peretz, Isabelle, Morais, José and Bertelson, P. (1987). Shifting ear differences in melody recognition through strategy inducement. In *Brain and Cognition*, 6, 202-215.
- Peretz, M. (1990). Processing of local and global musical information by unilaterally brain damaged patients. In *Brain*, 113, 1185-1205.
- Perrot, Jean (1997). *Polyphonie pour Yvan Fonagy*. Paris : L'Harmattan, 530 p.
- Peters, Ann M (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Peterson Gordon E. (1961). Parameters of vowel quality. In *Journal of Speech and Hearing Research*, 4, 10-29.

- Peterson, Gordon E. and Barney, Harold L. (1952). Control Methods used in a study of vowels. In *Journal of the Acoustical Society of America*, 24, 175-184.
- Petit, Jean (1985). *De l'Enseignement des Langues Secondes à l'Apprentissage des Langues Maternelles*. Paris : Champion, 693 p.
- Piaget Jean (1937). *La Construction du Réel chez l'Enfant*. Paris : Delachaux et Niestlé, 399 p.
- Pichon, Edouard (1936). *Le Développement Psychique de l'Enfant et de l'Adolescent : Evolution Normale, Pathologie, Traitement : Manuel d'Etude*. Paris : Masson, éd. 1965, 375 p.
- Pierrehumbert, Janet (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT, Indiana University Linguistic Club.
- Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press, 203 p.
- Pisoni, David B. (1977). Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. In *Journal of the Acoustical Society of America*, 61, 1352-1361.
- Piston (1969). *Harmony*. New-York: Norton.
- Polivanov, E. (1931). La perception des sons d'une langue étrangère. Dans *Travaux du Congrès de Prague*, Den Haag, 4, 31, 79-96.
- Polka, M. and Werker, Janet F. (1994). Developmental changes in perception of nonative vowel contrasts. In *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.
- Porcher, Louis et Groux, Dominique (1998). *L'Apprentissage Précoce des Langues*. Paris : PUF, 127 p.
- Potter R. K. and Steinberg J. C. (1950). Towards the specification of speech. In *Journal of Acoustical Society of America*, 22, 807-820.
- Povel, D. J. (1981). Internal representations of simple temporal patterns. In *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3-18.
- Pritchard, R. M. O. (1985). The teaching of French intonation to native speakers of English. In *International Review of Applied Linguistics*, 23/2, 117-147.
- Proverbio, Alice Mado, Minniti, Alessia and Zani, Alberto (1998). Electrophysiological evidence of perceptual precedence of global vs. Local visual information. In *Cognitive Brain Research*, 6, 321-334.
- Pugin, Maryline (1983). *Les Stimulations Auditives et leur Importance dans la Genèse de la Relation Mère-Enfant*. Université de Besançon: DEA en Neurosciences.
- Pye, Clifton (1983). Mayan telegraphese: intonational determinants of inflectional development in Quiché Mayan. In *Language*, 59/3, 583-604.
- Querleu, Olivier, Renard, Xavier, Versyp, F., Paris-Delrue, L. and Crépin, Gilles (1988). Fetal hearing. In *European Journal of Obstetrics and Reproductive Biology*, 29, 191-212.
- Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. In *Psychology of Music*, 18, 60-72.
- Read, C. and Schreiber, P. (1982). Why short subjects are hard to find ? In Gleitman and Wanner.
- Rivière-Raverlat, Jacquotte (1997). *Développer les Capacités d'Ecoute à l'Ecole. Ecoute Musicale, Ecoute des Langues*. Paris : PUF, 207 p.
- Rittaud-Hutinet, Chantal (2000). Enfance : les jeux de chansons. Communication présentée au *colloque Enseignement précoce du français langue étrangère et musique*, Institut für Didaktik der Romanischen Sprachen und Literaturen, Universität Justus-Liebig, Giessen (Allemagne), 11-13 décembre, Rauischolzhausen, 23 p.
- Rosenthal, Robert, A. and Jacobson, Lenore (1968). *Pygmalion in the Classroom*. New-York : Holt, Rinehart and Winston, Inc. *Pygmalion à l'école* (1971). Casterman, traduit de l'américain par Audebert, Suzanne et Rickards, Yvette, 293 p.
- Roslansky, J. D. (1969). *Communication. A Discussion at the Nobel Conference*. Amsterdam: North Holland Publishing Company.
- Rossi, Mario (1985). L'intonation et l'organisation de l'énoncé. Dans *Phonetica*, 42, 135-153.
- Rossi, Mario (1999). *L'Intonation, le Système du Français : Description et Modélisation*. Paris : Ophrys, 237 p.

- Rostand, Jean (1952). *L'Hérédité Humaine*. Paris : PUF (10ème ed. 1975), 128 p.
- Roudet, Léonce (1899). Méthode expérimentale pour l'étude de l'accent. Dans *La Parole*, 5, 321-344.
- Roudet, Léonce (1910). *Eléments de Phonétique Générale*. Paris : Welter, 363 p.
- Rousseau, Jean-Jacques (1763). *Essai sur l'Origine des Langues où il est parlé de la Mélodie et de l'Imitation Musicale*. Ed. de 1990. Paris : Gallimard, 284 p.
- Rousselot, l'abbé (1901-1908). *Principes de Phonétique Expérimentale*. Paris : Welter, 2ème vol., 1252 p.
- Rovee-Collier, Carolyn K. and Lipsitt, Lewis P. (1993). *Advances in Infancy Research*. Stanford: Ablex Publishing.
- Rovee-Collier, Carolyn K., Lipsitt, Lewis P. and Hayne, (1998). *Advances in Infancy Research*. Stanford: Ablex Publishing.
- Ruwet, Nicolas (1972). *Langage, Musique, Poésie*. Paris : Seuil, 247 p.
- Sansavini, A., Bertoncini, Josiane and Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. In *Developmental Psychology*, 33, 3-11.
- Sauvanet, Pierre (2002). Le rythme. Dans *Encyclopedia Universalis en ligne*.
- Schaeffer, Pierre (1966). *Traité des Objets Musicaux*. Paris : Seuil, 712 p.
- Schaffer, L. H. (1982). Rhythm and timing in skill. In *Psychological Review*, 83, 109-122.
- Schellenberg, E. G. & Trehub, Sandra E. (1996). Natural musical intervals. In *Psychological Science*, 7 (5), 272-277.
- Schenker, H. (1906). *Harmony*. (ed. Jones, trans. Borgese from 1906 German ed.). Cambridge: MIT.
- Schlaug, Gottfried, Jaencke, Lutz, Huang, Yanxiong and Steinmetz, Helmuth (1995). In vivo evidence of structural brain asymmetry in musicians. In *Science*, 267 (5198), 699-701.
- Schön D., Magne, C., Schrooten, M. and Besson, Mireille (2002). The music of speech: electrophysiological approach. In Bel and Marlien, *Actes de Speech Prosody*, 635-638.
- Schuyesmans (2002). *Le Chant Grégorien*. Site Internet disponible à l'adresse suivante : <http://www.schuyesmans.be/gregoriaans/>
- Schwartz J. L. (1987). *Représentation Auditive des Spectres Vocaliques*. Thèse de Doctorat d'Etat, Université de Grenoble.
- Scott, D. (1982). Duration as a cue to the perception of a phrase boundary. In *Journal of Acoustical Society*, 71, 996-1007.
- Scott, Thomas (1987). *Le français face au paysage anglais*. Dans Lhote, 131-142.
- Scripture, Edward, W. (1902). *The Elements of Experimental Phonetics*. Londre: Edward Arnold. Présente édition (1973), New-York, AMS Press Inc, 627 p.
- Sergeant, D. C. and Thatcher, G. (1974). Intelligence, social status and musical abilities. In *Psychology of Music*, 2, 32-57.
- Sebeok, Thomas (1977). *How Animals Communicate*. Bloomington: Indiana University Press.
- Selinker, Larry (1969). Language Transfer. In *General Linguistics*, 9, 67-92.
- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge : MIT Press.
- Sgard, Jean (1982). *Condillac et les Problèmes du Langage*. Paris : Slatkine, 462 p.
- Shadle, Christine H. (1997). The aerodynamics of speech. In Hardcastle and Laver, 33-64.
- Shahidullah, S. and Hepper, P. G. (1992). Hearing in the Fetus: Prenatal Detection of Deafness. In *International Journal of Prenatal and Perinatal Studies*, 4 (3/4), 235-240.
- Shankweiler (1966). Effects of temporal-lobe damage on perception of dichotically presented melodies. In *Journal of Comparative and Physiological Psychology*, 62, 115-119.
- Shelton, J. S. (1965). *The influence of home musical environment upon musical response of first-grade children*. PhD, Nashvill, Peasbody, College for Teachers.

- Shuter-Dyson, Rosamund (1994). Le problème des interactions entre hérédité et milieu dans la formation des aptitudes musicales. In Zenatti, 205-231.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992). ToBI: A standard for labelling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, 867-870.
- Sloboda, J. A. (1976). The effect of item position on the likelihood of identification by influence in prose reading and music reading. In *Canadian Journal of Psychology*, 30, 228-236.
- Smith, Laura, Hains, Sylvia, M.J., Lecanuet Jean-Pierre et Kisilevsky, Barbara S. (2000). Maturation of behaviours elicited by music in 32-37 weeks GA fetuses. Poster, *ICIS 2000* (International Conference on Infant Studies), Brighton.
- Snow, Catherine E. (1977). The development of conversation between mothers and babies. In *Journal of Child Language*, 4, 1-22.
- Snow, Catherine E. and Ferguson, Charles A. (1977) *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press, 219-235.
- Soulez, Antonia (1997). L'essence de la phonè (à propos du Cratyle 426C-427E). Dans Perrot, 411-423.
- Spencer, M. J. and DeCasper, A. D. (1987). Prenatal experience with low-frequency maternal-voice sounds influence neonatal perception of maternal voice samples. In *Infant Behavior and Development*, 10, 133-142.
- Spring, D. R. and Dale, P. S. (1977). Discrimination of linguistic stress in early infancy. In *Journal of Speech and Hearing Research*, 20, 224-232.
- Steele, Joshua (1775). *Prosodia Rationalis or an Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols*. London: Nichols, J.
- Stern, Daniel N., Spiker, Susan and MacKain, Kristine (1982). Intonation as signals in maternal speech to pre-linguistic infants. In *Developmental Psychology*, 18, 727-735.
- Stern, Daniel N., Spieker, Susan, Barnett, R. K. and Mac Kain, Kristine (1983). The prosody of maternal speech: infant age and context related changes. In *Journal of Child Language*, 10, 1-15.
- Studdert-Kennedy, Michael G. (1975). From continuous signal to discrete message: Syllable to phoneme. In Kavanagh and Cutting, 113-125.
- Tohkura, Vatikiotis-Bateson and Sagisaka (1992). *Speech Perception, Production and Linguistic Structure*. Tokyo: IOS Press.
- Thorpe, Leigh A. (1985). *Auditory-temporal organization: Developmental perspectives*. PhD, Université de Toronto.
- Thorpe, Leigh A., Trehub, Sandra E., Morrongiello, Barbara A. and Bull, Dale (1988). Perceptual grouping by infants and preschool children. In *Developmental Psychology*, 24, 484-491.
- Thorpe, Leigh A. and Trehub, Sandra E. (1989). Duration illusion and auditory grouping in infancy. In *Developmental Psychology*, 25, 122-127.
- Todd, N. (1985). A model of expressive timing in tonal music. In *Music Perception*, 3, 33-57.
- Trainor, Laurel (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. In *Infant Behavior and Development*, 19, 83-92.
- Trainor, Laurel and Trehub, Sandra (1992). A comparison of infants' and adults' sensitivity to Western musical structure. In *Journal of Experimental Psychology: Human Perception and Performance*, ??.
- Trainor, Laurel and Rock, Adrienne, M.L. (1997). Distinctive messages in Infant-directed lullabies and playsongs. Poster, *SRCD*, Washington DC, April.
- Trainor, Laurel, Clark, E.D., Huntley, A. and Adams, B.A. (1997). The acoustic basis of preferences for infant-directory singing. In *Infant Behavior and Development*, 20, 383-396.
- Trainor, Laurel and Zacharias, C.A. (1999). Infants prefer higher-pitched singing. In *Infant Behaviour and Development*.
- Tramo, Mark Jude (2001). Music of the hemispheres. In *Science*, 291 (5501), 54-56.

- Trehub, Sandra E. (1973). Infant's sensitivity to vowel and tonal contrasts. In *Developmental Psychology*, 9, 91-96.
- Trehub, Sandra E. (1989). Infants' perception of musical sequences: Implications for language acquisition. In *Speech-Language Pathology and Audiology*, 13, 3-11.
- Trehub, Sandra E., Bull, Dale and Thorpe, Leigh A. (1984). Infants' perception of melodies: The role of melodic contour. In *Child Development*, 55, 821-830.
- Trehub, Sandra E., Morongiello, B. A. and Thorpe, Leigh A. (1985). Childrens' perception of familiar melodies: The role of intervals, contour, and key. In *Psychomusicology*, 5, 39-48.
- Trehub, Sandra E. and Thorpe, Leigh A. (1989). Infants' perception of rhythm. Categorization of auditory sequences by temporal structure. In *Canadian Journal of Psychology*, 43, 217-229.
- Trehub, Sandra, Thorpe, Leigh A. and Trainor, Laurel J. (1990). Infants' perception of good and bad melodies. In *Psychomusicology*, 9, 5-19.
- Trehub, Sandra, Trainor, Laurel, J. and Unyk, A. M. (1993a). Music and speech processing in the first year of life. In *Advances in Child Behavior and Development*, 24, 1-35.
- Trehub, Sandra, Unyk, A. M. and Trainor, Laurel J. (1993b). Material singing in cross cultural perspective. In *Infant Behavior and Development*, 16, 285-295.
- Trehub, Sandra E. et Trainor, Laurel, J. (1994). Les stratégies d'écoute chez le bébé : origines du développement de la musique et de la parole. In McAdams et Bigand, 299-347.
- Trehub, Sandra E. and Henderson, J. (1994). Caregivers' songs and their effect on infant listeners. In Deliège (Ed.). *Proceedings of the 3rd International Conference for Music Perception and Cognition*, Liège, Belgium, ESCOM, 47-48.
- Trehub, Sandra E., Unyk, A. M., Kamanetsky, S.B., Hill, D.S., Trainor, Laurel J., Henderson, J.L. and Saraza, M. (1997). Mothers' and fathers' singing to infants. In *Developmental Psychology*, 33, 500-507.
- Trehub, Sandra E. and Trainor, Laurel, J. (1998). Singing to infants: lullabies and play songs. In Rovee-Collier, Lipsitt and Hayne, 43-77.
- Troubetzkoy, Nicolas S. (1949). *Principes de Phonologie*. Paris : Klincksieck, 396 p. Traduit par Cantineau, J. (1986).
- Umiker, Jean and Sebeok, Thomas A. (1976). *Speech Surrogates*. Mouton. The Hague.
- Unyk, A. M., Trehub, Sandra, Trainor, Laurel, J. and Schellenberg, E. G. (1992). Lullabies and simplicity: a cross-cultural perspective. In *Psychology of Music*, ??.
- Valette, Rebecca, M. (1975). *Le Test en Langue Etrangères : Guide Pratique*. Paris : Hachette, 111 p.
- Van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences*. PhD. The Netherlands, Eindhoven University of Technology.
- Van Waesberghe, J. Smits (1957). Phonetics in its relation to musicology. In Kaiser, 372-384.
- Verdeau-Pailles (1981). *Le Bilan Psychomusical*. Ed. J. M. Fuzeau S.A.
- Vihman, Marylin M., Macken, Marlys A., Miller, R., Simmons, H. & Miller, J. (1985). *From babbling to speech: A reassessment of the continuity issue*.
- Vinter, Shirley (1992). *Mise en Place des Eléments Prosodiques dans le Langage Emergent de l'Enfant Sourd : Rôle des Stimulations Acoustiques et des Interactions Sociales*. Thèse de Doctorat, Besançon.
- Vinter, Shirley (1994). *L'Emergence du Langage de l'Enfant Déficient Auditif : des Premiers Sons aux Premiers Mots*. Paris : Masson, 147 p.
- Viret, Jacques (1987). *La Modalité Grégorienne : un Langage pour quel Message ?* Lyon : A Cœur Joie, 167 p.
- Vos, P. and Troost, J. (1989) Ascending and descending melodic intervals: statistical findings and their perceptual relevance. In *Music Perception*, 6, 383-396.
- Weber, E. W., Spychiger, M. und Patry, J. L. (1993). *Musik macht Schule. Biografie und Ergebnisse eines Schulversuchs mit erweitem Musikunterricht*. Essen.
- Wenk, and Wioland, François (1982). Is French really syllable-timed ? In *Journal of Phonetics*, 19.

- Werker, Janet F. and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross language speech perception. In *Journal of the Acoustic Society of America*, 75, 1866-1878.
- Wertheimer, M. (1923c) Untersuchung zur lehre von der gestalt II. In *Psychologische Forschung*, 4, 301-350.
- Whitaker, Harry A. (1976). Neurobiology of Language. In Caterandte and Friedman.
- Williams, Linda V. (1986). *Deux Cerveaux pour Apprendre : le Gauche et le Droit*. Paris : les Editions d'Organisation, 204 p.
- Wing, A. M. and Kristofferson, A. B. (1973a). The timing of inter-response intervals. In *Perception and Psychophysics*, 13, 455-460.
- Wing, A. M. et Kristofferson, A. B. (1973b). Response delays and the timing of discrete motor responses. In *Perception and Psychophysics*, 14, 5-12.
- Wing, H. D. (1941). *Musical Ability and Appreciation*. PhD Thesis, London University.
- Wundt, Wilhelm (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Leipzig : Engelmann, W., 3ème édition, 1912.
- Yaeger, Malcah (1979). *Context Determined Variation in Montreal French Vowels*. University of Pennsylvania, Ph.D, 353 p.
- Yeni-Komshian, G., Kavanagh, J. F. and Ferguson, Charles A. (1980). *Child Phonology: Production*, tome I. New-York: Academic Press.
- Yeni-Komshian, G., Kavanagh, James F. and Ferguson, Charles A. (1980). *Child Phonology: Perception*, tome II. New-York: Academic Press, 255 p.
- Zenatti, Arlette (1969). *Le développement génétique de la perception musicale*. Monographies Françaises de Psychologie. Paris : CNRS (n°17).
- Zenatti, Arlette (1980). *Tests Musicaux pour Jeunes Enfants, avec Applications en Psychopathologie de l'Enfant et de l'Adulte*. Issy Les Moulineaux : Editions Scientifiques et Psychologiques, 53 p.
- Zenatti, Arlette (1981). *L'Enfant et son Environnement Musical. Etude Expérimentale des Mécanismes Psychologiques d'Assimilation Musicale*. Issy Les Moulineaux : E.A.P., 275 p.
- Zenatti, Arlette (1994). *Psychologie de la Musique*. Paris : PUF, 391 p.
- Zwanenburg, Wiecher (1965). *Recherches sur la Prosodie de la Phrase Française*. Leiden : Universitaires Pers Leiden, 136 p.

Discographie

(par ordre d'apparition des extraits musicaux dans le texte)

The Jimi Hendrix Experience : Electric Ladyland, CD MCA LC 01056, 1997 (Première publication : , 1968).

Piste 13 (4 mn 24) : Still Raining, Still Dreaming (1968).

Piste 13 (4 mn 24) : Still Raining, Still Dreaming

Bobby McFerrin : The Voice, CD Elektra Musician 7559-60366-2, 1984.

Piste 2 (4 mn 49) : The Jump.

Moines de l'Abbaye de Notre-Dame de Ganagobie : *Chant Grégorien, Chant de Toujours*, (1992). Editions Jade (74321-39406-2).

Piste 9 (1 mn 39) : Kyrie XIII

Piste 12 (0 mn 51) : Sanctus XIII

Henry Purcell (1682). *Didon et Énée*. Taverner Choir and players, direction Andrew Parrott (1999). Enregistré à St Giles's Cripplegate, 9-15/09/1994. Sony Classical (SK 62993).

Piste 20 (1 mn 02) : Ruin'd ere the set of Sun ? (Enchantress I, Enchantress II, Sorceress). Acte II, scene, the Cave.

Jean-Sébastien Bach (1729). *Matthäus Passion* (BWV 244). Harmonia Mundi (1990), RD77848.

Piste 7 (disque I, 0 mn 39). Rezitativ (Evangelistas, Judas) : Da ging him der Zwölfen einer.

Wolfgang Amadeus Mozart (1787). *Don Giovanni*. Chor der Niederländischen Oper. Royal Concertgebouw Orchestra Amsterdam. Nikolaus Harnoncourt (1989). Teldec Classics (2292-441184-2).

Piste 8 (4 mn 54, disque III) : Récitativo Don Giovanni - Leporello : Ah ah ah !ah, questa è buona.

Ludwig Van Beethoven (?). *Symphonie N° 9 d-moll* opus 125. Deutsche Grammophon (1977, Polydor), 429089-2. Berliner Philharmoniker, Herbert Von Karajan.

Disque III, Piste 5 (6 mn 26). Presto, allegro assai.

Claude Debussy (1902). *Pelléas et Mélisande*, dirigé par Inghelbrecht. Naïve : Montaigne Archives (V4854).

CD1, piste III (acte I, scène III), 7 mn 23. Devant le Château. Mélisande, Geneviève et Pelléas.

Schoenberg, Arnold (1912). *Ode to Napoleon Buonaparte*, opus 41 (16 mn 15). David Pittman-Jennings (baryton), Pierre Boulez. Polygram (457 630-2). Texte : Lord Byron (traduction Stadelmann, H.).

Luciano Berio (1998). *Sequenzas*. Deutsche Grammophon (457 038-2). Ensemble Intercontemporain. Premier disque

Piste 3 " Sequenza III pour voix de femme (UE 13723 mi) " (8 mn 21), 1966. Luisa Castellani, soprano.

Luca Francesconi (1994) *Etymo*. Pour soprano, ensemble et dispositif électronique. Luisa Castellani, soprano. Ensemble Intervontemporain, direction : Pascal Rophé. Durée : 25 mn.

Robert Wyatt (1974). *Rock Bottom*. Rykodisc (HNCD1426), réédité en 1998.

Piste 5 (6 mn 31) : Alife

Hekstatic

Piste 3 (2 mn 59) : Ninja Tune.

Piste 4 (5 mn 03) : Kids can dance.

The Mothers of Invention (1966). *Freak out !* Rykodisc (10501). 1987/1995.

Piste 14 "It can't happen here" (3 mn 55).

Partitions

Wolfgang Amadeus Mozart (1787). *Don Giovanni* (KV 527). Opéra en deux actes. Budapest (1995) : Könemann Music, 385 p.

Acte II, scène XII : récitatif, Don Giovanni, Leporello : extrait de “ Ah ah ah ! Questa ”, p : 297.

Jean-Sébastien Bach (1729). *St Matthew Passion* (BWV 244). Londres : Novello, 248 p.

Rezitativ (Evangelistas, Judas) : Da ging him der Zwölfen einer (première partie), p : 36.

Henry Purcell (1682). *Dido and Aeneas*. Edited by Margaret Laurie and Thurston Dart. Novello, Borough Green, Sevenoaks, Kent (revised june 1966, revision april 1974), 112 p.

Ruin'd ere the set of Sun ? (acte II), 32-33.