



**HAL**  
open science

# Impact de l'intelligibilité dans les troubles de la production de la parole pathologique

Virginie Woisard

► **To cite this version:**

Virginie Woisard. Impact de l'intelligibilité dans les troubles de la production de la parole pathologique. Sciences du Vivant [q-bio]. Aix-Marseille 1 (Aix-en-Provence), 2011. Français. NNT : . tel-04508342

**HAL Id: tel-04508342**

**<https://hal.science/tel-04508342>**

Submitted on 18 Mar 2024

**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.



UMR 6057

Laboratoire Parole & Langage

Aix-en-Provence, France



*Ecole doctorale Cognition, Langage, Education*

**THESE**

# *Impact de l'intelligibilité dans les troubles de la production de la parole pathologique*

***Doctorat Aix-Marseille Université, délivré par l'Université d'Aix en Provence***

Présentée et soutenue le 14 juin 2011 par

**Virginie Woisard-Bassols**

***Composition du jury***

***Rapporteurs : Lorraine Baqué : Université autonome de Barcelone***

***Philippe Dejonckere : Institut de Phoniatrie, Université d'Utrecht***

***Examineurs : Danielle Duez : Laboratoire Parole et Langage, Aix en Provence***

***Rudolphe Sock : Institut de phonétique de Strasbourg***

***Directeur de thèse : Danielle Duez***



## RESUME

---

L'intelligibilité de la parole est un véritable enjeu dans la prise en charge des personnes présentant un trouble de la production de la parole pathologique (TPPP). L'usage de ce terme, avec un sens commun recouvrant le domaine de la compréhension, et un sens restreint répondant à la mesure du transfert d'information linguistique dans un contexte de test, entretient une ambiguïté ; ambiguïté qui entrave le développement des concepts de prise en charge.

L'objectif de ce travail de thèse est, à travers trois groupes d'expériences autour de la notion d'intelligibilité, de proposer un cadre conceptuel pouvant servir à organiser la prise en charge des TPPP et à identifier les questions de recherche fondamentale impliquées dans la problématique de l'intelligibilité dans les TPPP.

L'objectif principal du premier groupe d'expériences est d'étudier la place de l'intelligibilité dans les stratégies cognitives sous-jacentes à la perception de la parole pathologique. La méthode utilisée est une catégorisation libre. Le Corpus est constitué de phrases lues par 33 locuteurs comprenant 20 sujets présentant un TPPP (origine neurologique, séquelles de malformation et de cancers) et 13 témoins. Trois groupes de 15 auditeurs ont participé à cette expérience : 1 groupe d'experts (orthophonistes et de phoniatres), 1 groupe de non experts (étudiants en début de cursus), 1 groupe « entourage » (famille ou aidants). Cette première série d'expériences a été complétée, sur le même principe, par l'étude d'un corpus produit par des locuteurs dysphoniques. Les résultats sont en faveur d'une classification déterminée à une extrémité par une comparaison à un « prototype de normalité » et à l'autre extrémité par le degré d'altération de la production de parole. Le prototype de normalité apparaît ici comme l'expression d'une norme par rapport à une dégradation dans le cadre d'une maladie. Il soulève la question de la spécificité des altérations d'un signal de parole lié à un TPPP. L'intelligibilité en tant qu'indice de la sévérité est un concept retrouvé pour les TPPP aussi bien dans la construction des arbres que dans les commentaires caractérisant les catégories. Intelligibilité et compréhensibilité sont alors interchangeables, ces termes renvoyant à la difficulté ressentie à accéder globalement au sens. La notion de gravité est clairement identifiable sur les arbres pour les dysphonies mais n'est pas explicitée dans les commentaires. En fonction de l'entrée suggérant un prototype de voix ou de parole nous n'utiliserions pas les mêmes critères. Ces critères sont partagés par tous les groupes d'auditeurs.

La deuxième série d'expériences est consacrée à la manière dont un groupe d'experts réalise une tâche de jugement de sévérité comparée à une tâche de jugement d'intelligibilité. Au cours d'une expérience off line, le corpus de l'étude précédente, complété par un corpus de phrases différentes pour chaque locuteur, a été analysé par 5 juges (orthophonistes et phoniatres). Pour chaque stimulus, était d'abord demandée l'attribution d'un jugement en termes de qualité de la voix, de résonance, de prosodie et de qualité de l'articulation des phonèmes. Une échelle visuelle analogique (d'une étendue de 1 à 30 unités) permettait ensuite de scorer chaque stimulus pour la sévérité et le degré d'altération



de l'intelligibilité ; l'intelligibilité étant définie comme la compréhensibilité du message véhiculé par le signal, alors que la sévérité était définie comme le degré d'altération globale du signal. Les scores moyens de sévérité et de détérioration de l'intelligibilité sont respectivement de 9,8/30 (+/- 1,72) et 8,2/30 (+/- 1,59). Une différence dans la distribution des scores est mise en évidence pour 3 juges alors que la distribution est superposable pour 2 juges. Quand une différence existe, la tendance est à l'attribution dans la partie médiane de la sévérité, d'un score moins élevé pour l'altération de l'intelligibilité. Ceci est associé à une différence dans la manière d'analyser les caractéristiques des stimuli lors de la tâche réalisée juste avant. Ces résultats révèlent un comportement potentiellement différent parmi des juges expérimentés. Ils apportent des arguments pour mesurer l'intelligibilité au niveau du code de surface avec une tâche de reconnaissance et pour réserver les mesures d'amplitude directes à l'évaluation de la sévérité. La standardisation des consignes paraît fondamentale pour ce type d'évaluation.

Le troisième groupe d'expériences cherche à déterminer les cibles les plus efficaces dans la prise en charge des TPPP. Elle repose sur l'hypothèse suivante : la conservation d'une perception juste d'un phonème ou d'une syllabe malgré sa distorsion pathologique permet de déterminer sa contribution à l'intelligibilité et participe à la définition de la notion de robustesse perceptive. Pour tester cette hypothèse, une méthode d'analyse du taux de perception obtenu par un jury d'auditeurs, en fonction du degré de vraisemblance issu d'un système automatique de reconnaissance a été formalisée. Pour cela, un jury de 11 auditeurs a été réuni. Le corpus composé de 16 consonnes combinées à 5 voyelles du français constituant une combinaison de 80 syllabes de type CV a été produit par 12 locuteurs : 10 patients opérés et irradiés pour un cancer de la cavité buccale ou du pharynx et 2 témoins. Les résultats obtenus sont à pondérer en raison des mauvaises performances de la reconnaissance vocale pour les consonnes /g/,/t/ et à un moindre degré /z/,/b /,/p/ et /k/. Les consonnes robustes, dans le sens d'une résistance avec une bonne perception malgré une altération de la production (robustesse à la perception) sont le /n/ et /v/. Certaines consonnes sont bien identifiées alors qu'elles sont produites avec un indice de vraisemblance éloigné du générique : /ch/ et /d/. Les consonnes /p/ et /g/ ont un profil similaire mais la mauvaise performance du système à leur égard discrédite ce résultat. La consonne repérée comme fragile est le /s/. Ce travail est en faveur d'une contribution inégale des différentes consonnes pour l'intelligibilité de la parole. La méthode utilisée paraît pertinente et faisable. Mais une analyse plus approfondie des résultats obtenus par le système automatique de reconnaissance est indispensable avant de continuer à caractériser les consonnes par cette méthode. Les résultats de ces expériences, interprétées sous l'éclairage croisé des théories sur la perception et de la production de la parole et des modèles de santé, nous ont permis de proposer une définition des différents concepts dans la perspective de la classification internationale du fonctionnement.

## **SUMMARY**

---

The intelligibility of speech is a real challenge for the management of speech disorders.

The use of this word, with a common meaning of comprehension and, a specific meaning in a context of test, of measurement of the information transfer, carries some ambiguity which could hinder the concept of management.

The purpose of this work is to propose a conceptual framework for the management of pathological speech production (PSP) through three kinds of experiments.

The aim of the first set of experiments is to study the place of the intelligibility within the underlying cognitive strategies of the perception of pathological speech.

A free classification method is used with a corpus of sentences read by 33 speakers: 20 patients with speech disorders (a pathological speech production due to neurological diseases or sequellae of cancer or malformative lesions) and 13 normal subjects. Three groups of 15 people attended the test : 1/ group of experts (logopedists and phoniaticians), 2/group of non experts (students at the beginning of their cursus), 3/group of natural aids (family or caregivers). This first set of experiments was filled, following the same principle, by a study of a corpus produced by speakers with chronic dysphonia. The results show a preference for a classification determined at an extremity by a comparison with a "prototype of normality" and at the opposite by the degree of alteration of the signal. Here, the prototype of normality is relative to the degradation in the frame of pathology. In speech disorders, intelligibility as a hint of severity is a construct confirmed by this study. Intelligibility and comprehensibility are then mixed and used to express the difficulty in making sense. The concept of severity is also important in chronic dysphonia but is not explicit. Depending on the input suggesting a prototype of voice or a prototype of speech, we would not use the same criteria. This interpretation of the results is available whoever the listeners are.

The second set of experiments studies the way an expert group performs evaluation of severity compared to evaluation of intelligibility. During an off line experiment, the corpus of the previous study filled by a corpus of different sentences by speaker, was presented to 5 judges (logopedists and phoniaticians). For each stimulus, a first task consists in a judgment in terms of quality of voice, resonance, prosody and articulation of phonemes. After that, a visual analogical scale, range from 1 to 30, allowed to score each stimulus for severity and the degree of degradation of intelligibility.

Intelligibility was defined as comprehensibility of the message supported by the signal, severity being defined as the global degradation of the signal. The average scores of severity and alteration of intelligibility are respectively 9,8/30 (+/- 1,72) and 8,2/30 (+/- 1,59). A difference in the distribution of the scores is pointed out concerning 3 judges while the distribution is similar for the 2 others. When there is a difference, the trend is to give a lower score for the alteration of intelligibility in the middle of the scale of severity. This is associated with a difference in the way to analyze the characteristics of the stimuli during the task performed previously. These results show a potential difference of behavior among experienced judges. They bring some arguments for measuring intelligibility at the surface code

with an identification task and for measuring the severity with a direct magnitude measurement. The standardization of the instructions appears fundamental for this kind of assessment.

The third group of experiments looks for the most efficient targets for managing speech disorders. It's based on the following assumption: the preservation of a correct perception of a phoneme or a syllable, despite its pathological distortion, allows to determine its contribution to intelligibility and participates to the notion of perceptive robustness. For testing this assumption, was proposed a method of analysis by the identification score (obtained by a jury of listeners) function of the degree of' plausibility (obtained by an automatic system for speech recognition). A jury of 11 listeners was organized. The corpus of 80 syllables is composed by 16 French consonants combined with 5 vowels. The speakers were 10 patients treated for a carcinoma of the oral cavity or the pharynx and control subjects. The results need to be carefully interpreted because of low performances of the automatic system for /g/,/t/ and at a lesser degree for /z/,/b /,/p/ et /k/. The robust consonants are /n/ and /v/. Some consonants are well identified despite a very low index of plausibility (interpreted as produced very far from the generic phoneme). That's the case for /ch/ and /d/. The consonants /p/ and /g/ have a similar profile, but the low confidence of the automatic system for these phonemes justifies caution. The consonant identified as fragile is the /s/. This work demonstrates a non similar contribution between the different consonants. The method used here seems to be pertinent and reliable. But the verification of the results obtained by the automatic system is an obligation previous characterize consonants.

The outcome of these experiments, adopting the points of view of some perception theories of SD mixed with health models allow us to precise the definition of the main concepts in the prospect of the international classification of functioning.

***Mots clés : intelligibilité, troubles de la production de la parole, dysarthries, dysphonies, catégorisation libre,cancer de la tête et du cou, percpetion, robustesse***

***Key words : intelligibility,speech disorders, dysarthria, dysphonia, free classification, head and neck cancer, perception, robustness***

## REMERCIEMENTS

---

*« O dieux, donnez moi la sérénité  
d'accepter ce que je ne puis changer, le  
courage de changer ce que je puis, et la  
sagesse d'en connaître la différence »*

*Marc Aurèle ou Saint François d'Assise*

*A ma famille....*

*A tous ceux, sans qui ce travail n'aurait pas été possible...*

*Danielle, Alain, Robert, Pascal...*

*A tous ceux qui m'ont accompagnée, soutenue...*

*Michèle, Nicole, Stéphanie, Pascale, Sabine et...*



## TABLE DES MATIERES

Résumé	3
Remerciements	7
1.Introduction	11
1.1. Motivation	13
1.2. Réflexion	13
1.3. Le cadre pathologique	14
1.4. La conception de la prise en charge des patients	15
1.5. Questionnement	17
2.De l'intelligibilité aux troubles de la production de la parole pathologique	21
2.1.Qu'est ce que l'intelligibilité ?	23
2.1.1.Définition	23
2.1.2. Usage actuel	23
2.2.Qu'est ce que l'intelligibilité de la parole ?	24
2.2.1. Définition	24
2.2.1. Les différentes applications	25
2.3. Les différents cadres théoriques de références	34
2.3.1. Les différents modèles de communication	35
2.3.2. De la perception de la parole à la compréhension du discours en conversation	40
2.4. Les principes de l'évaluation de l'intelligibilité de la parole	66
2.4.1. Les grands courants	66
2.4.2. Les différentes manières de mesurer l'intelligibilité.	70
2.5. .... Quels sont les facteurs déterminant l'intelligibilité dans les troubles de la production de la parole	75
2.5.1. Travaux issus du domaine des TPPP	75
2.5.2. Travaux issus de la perception dans le bruit	77
2.5.3. Travaux issus de la manipulation de signaux hors contexte des TPPP	78
2.5.4. Le caractère naturel de la parole	82
2.5.5. En synthèse	83
3. Catégorisation libre dans les troubles de la production pathologique de la parole	85
3.1. Introduction	89
3.2. Première étude	93
3.2.1. Matériel et méthode	93
3.2.2. Description de la procédure des tests	98
3.2.3. Résultats	100
3.2.4. En résumé	125
3.3. Seconde étude	126
3.3.1. Matériel et méthode	127
3.3.2. Description de la procédure des tests	131
3.3.3. Résultats	132
3.3.4. Résumé-discussion	152
3.4. Comparaison résultat étude 1 et étude 2	153
3.4.2. Comparaison des résultats des classes	153
3.4.2. Comparaison entre l'analyse des commentaires du corpus comportant des pathologies vocales et du corpus comportant des pathologies de la production de la parole.	154
3.5. Discussion générale de la catégorisation	158
3.5.1. Proposition d'une hypothèse	158
3.5.2. La normalité et la déviance pathologique par rapport à la « normale »	161
3.5.3. Contribution à la définition de l'intelligibilité	163
3.5.3. La notion d'expert	164
3.6. Conclusion	166
4. Intelligibilité et sévérité dans les troubles de la production de la parole pathologique	167
4.1. Introduction	169
4.1.1. La sévérité d'un trouble, une notion mal définie en français.	169
4.1.2. Réflexions à partie du modèle de la surdit� et �bauche d'une d�finition	170
4.1.3. Qu'en est-il dans les autres domaines ?	171
4.1.4. Les notions de s�v�rit� et d'intelligibilit� dans les TPPP	173
4.1.5. Objectifs de ce travail	174
4.2. Population, Mat�riel et M�thode	174
4.2.1. Les locuteurs	174
4.2.2. Le corpus	175
4.2.3. Les auditeurs	176
4.2.4. La proc�dure	176
4.2.5. M�thodologie	178

4.3. Résultats .....	179
4.3.1. Expérience 1 .....	179
4.3.2. Expérience 2 .....	188
4.3.3. Expérience 3 .....	190
4.4. Discussion.....	194
4.4.1. Les faits établis par cette expérience .....	194
4.4.2. Quelles peuvent être les raisons de cette hétérogénéité ? .....	195
4.4.3. En ce qui concerne la différence de score entre les deux notions, quelles peuvent être les explications ? .....	196
4.5. Conclusion .....	198
5. Robustesse et contribution a l'intelligibilité .....	199
5.1. Introduction .....	203
5.1.1. La notion de robustesse .....	203
5.1.2. Hypothèses de recherche .....	205
5.2. choix méthodologiques .....	212
5.2.1. Les locuteurs .....	212
5.2.2. Le Corpus .....	212
5.2.3. La Procédure de la tache de perception .....	218
5.2.3. La méthode de la reconnaissance automatique de la parole. ....	221
5.2.4. Le Plan d'analyse .....	226
5.3. Résultats .....	227
5.3.1. Description des locuteurs .....	227
5.3.2. Résultats du point de vue des auditeurs .....	232
5.3.3. Résultats du système automatique de reconnaissance vocale .....	253
5.3.4. Test des hypothèses .....	263
5.4. Discussion.....	275
5.4.1. Synthèse des résultats .....	275
5.4.2. La robustesse aux distorsions .....	276
5.4.2. Les méthodes pour évaluer l'impact du déficit de production sur la perception de la parole .....	280
5.4.3. La robustesse à la perception .....	283
5.5. Conclusion .....	286
6. Conclusions et perspectives .....	287
6.1. Implication sur la définition des différents concepts .....	289
6.1.1. Intelligibilité et sévérité .....	289
6.1.2. Essai de définition des notions de sévérité, intelligibilité et compréhensibilité dans le cadre conceptuel de la CIF .....	290
6.1.3. Implications pour la mesure de l'intelligibilité et perspectives pour l'évaluation dans les TPPP .....	294
6.2. Implications sur la détermination des unités linguistiques minimales pertinentes pour les TPPP.....	298
6.2.1. L'intrication des différentes informations comprises dans le signal .....	299
6.2.2. Les compensations observées dans les TPPP .....	299
6.2.3. La notion de robustesse et les unités minimales perceptives .....	300
7. Annexes .....	303
7.1. Annexes en rapport avec le chapitre 3 (catégorisation libre).....	305
7.1.1. Les textes utilisés pour les stimuli .....	305
7.1.2. Annexe Logiciel TCL .....	305
7.1.3. Annexe EVA .....	307
7.1.4. Annexes en rapport avec la section 3.2.3. : l'étude 1 .....	310
7.1.5. Annexes en rapport avec la section 3.3.3. : Etude 2 .....	316
7.2. Annexes du chapitre 4 (Intelligibilité et sévérité).....	323
7.2.1. Les Phrases du Hint utilisées pour le corpus de stimuli différents .....	323
7.2.2. Tableaux avec les valeurs des corrélations entre les paramètres des différentes tâches. ....	324
7.3 Annexes en rapport le chapitre 5 (Robustesse et contribution a l'intelligibilité).....	324
7.3.1. Annexes en rapport avec la section.5.3. ....	324
7.3.2 Annexes en rapport avec la section 5.4 : la fréquence moyenne des consonnes .....	335
8. Références .....	337
8.1. Bibliographie .....	339
8.2. Logiciels.....	359
8.3. Adresse des Sites web .....	360

# 1. INTRODUCTION

---

Ou

Pourquoi s'engager dans une thèse en sciences du langage quand on est un phoniatre expérimenté et  
notoirement reconnu dans le domaine des troubles de la déglutition ?

1.	Introduction	11
1.1. Motivation .....		13
1.2. Réflexion .....		13
1.3. Le cadre pathologique .....		14
1.4. La conception de la prise en charge des patients .....		15
1.5. Questionnement .....		17





## 1.1. MOTIVATION

---

Depuis 1992, date à laquelle m'a été confiée par le Pr Pessey la responsabilité de l'unité de la voix et de la déglutition dans un service d'otorhinolaryngologie du centre hospitalo-universitaire de Toulouse, mon quotidien consiste à résoudre des problèmes relevant des dysfonctionnements du carrefour aérodigestif. Certains vont répondre à des traitements dits « curatifs », faisant disparaître la cause du dysfonctionnement et le dysfonctionnement lui-même. Beaucoup de personnes garderont des dysfonctionnements, soit par régression partielle de la cause et du dysfonctionnement qui en découle, soit parce que le traitement induit par lui-même des séquelles, soit parce qu'il n'existe pas de traitement étiologique et que la maladie peut évoluer.

Mon cursus initial d'otorhinolaryngologiste (ORL) complété par ma pratique dans les domaines de la pathologie vocale et de la cancérologie des voies aérodigestives, a dû, sous la pression du profil de la patientèle ORL puis neurologique, s'enrichir de connaissances permettant la prise en charge en priorité de la déglutition, du fait de son caractère vital. Cependant, la majeure partie des patients qui souffrent de troubles de la déglutition présentent aussi des troubles de la production de la parole. D'ailleurs, en pratique, ce sont souvent, pour la réhabilitation fonctionnelle, les mêmes acteurs qui assurent la prise en charge rééducative de ces deux troubles : les orthophonistes. Mais, bien qu'il s'agisse des mêmes structures anatomiques et des mêmes processus étiopathogéniques, les conséquences sur les différentes fonctions ne se distribuent pas de la même façon. Cette observation explique que la démarche clinique guidant la prise en charge est commune d'un point de vue étiologique puis qu'elle diverge au fur et à mesure du cheminement vers un objectif d'amélioration de la qualité de vie malgré la persistance de dysfonctionnements. L'absence de satisfaction dans les propositions de prise en charge actuelle pour améliorer la qualité de vie des patients présentant des troubles pathologiques de la production de la parole, quelques soient les résultats de la démarche étiologique, est à l'origine de l'engagement dans ce travail.

## 1.2. REFLEXION

---

Dans le cheminement des réflexions qui ont conduit au choix du sujet de cette thèse, je voudrais rapporter ici plusieurs points clés ou étapes du « détricotage »...Je veux dire par là du processus qui, je l'espère, m'a permis d'avoir un point de vue « nouveau » sur la base de la remarque de Goldberg Elkhonon (2007) : « Et comment la nouveauté apparaît-elle ? Elle est créée en configurant de manière

originale, selon des assemblages inconnus jusqu'alors, des bribes de l'ancien savoir. Les éléments sont anciens, mais leur agencement est nouveau, sans précédent historique, et donc sans correspondance exacte avec aucune des représentations mentales déjà connues dans la tête de l'auteur ». Ces étapes ont consisté en :

- un « retour en arrière » pour arriver à se libérer des filtres apposés par les champs disciplinaires de la culture médicale, masquant une vision plus large du questionnement,
- une recherche de concepts susceptibles d'accueillir la démarche,
- suivi d'un « retour » en avant pour se recentrer à nouveau sur l'objet de la préoccupation.

Une métaphore pour illustrer ce processus pourrait être celle du microscope réglé sur le grossissement adapté à la vision de l'objet mais trop fort pour voir toutes les ramifications de l'objet et donc apprécier sa globalité. En passant à un grossissement plus faible, la nature de l'objet s'enrichit des différents rapports avec le « monde » et peut être « vu », et éventuellement orienté d'une manière différente lors du retour au grossissement initial. Ainsi ai-je été passionnée par les implications de théories sous-tendant l'approche systémique, la théorie de la complexité, la théorie de l'auto-organisation... Ceci je l'espère transparaîtra dans la partie théorique qui précédera l'exposé des travaux expérimentaux. Je m'attacherai ici à préciser deux écueils dont il faut s'affranchir avant d'aller plus loin.

### **1.3. LE CADRE PATHOLOGIQUE**

---

Le premier est déterminant dans le titre de ce travail qui aurait pu s'appeler « impact des troubles de l'intelligibilité sur les dysarthries ». Or, si le préfixe « dys » emprunté au grec  $\delta\upsilon\sigma$ - signifiant négation, malformation, mauvais, erroné, difficile, fait évoquer un mauvais état, un mauvais fonctionnement, une anomalie, une difficulté s'appliquant au signifiant du morphème qu'il précède, il n'est supposé en aucun cas induire de restriction étiologique au concept associé au morphème. Ainsi, il n'y a pas de différence significative entre un trouble de la déglutition et une dysphagie, entre un trouble de la voix et une dysphonie. Une analyse plus exhaustive des définitions de la liste des termes médicaux l'utilisant le confirme. Mais il n'y a pas qu'en français que l'exception confirme la règle et la dysarthrie est devenue une « propriété » de la neurologie avec comme définition éditée à la première ligne de la préface de l'ouvrage « les dysarthries » coordonné par P Auzou, V. Rolland-Monnoury, S. Pinto et C. Özanscak (2007) : « Le terme de dysarthrie désigne un trouble de l'exécution motrice de la parole consécutif à une atteinte du système nerveux central et/ou périphérique ». Les détails sur cette appropriation survenue à partir des travaux de Darley et al (1969) sont évoqués, pour ceux que cela intéresse, par

Darley et al dans leur ouvrage de 1975 : « Motor speech disorders » et par Duffy JR dans le chapitre « History, current practice and future trends and goal » du livre « Motor speech disorders, Essays for Ray Kent » édité par Weismer (2006). Les discussions avec ma directrice de thèse, Danielle Duez, à ce sujet, ont conduit à fixer notre choix sur la terminologie de « troubles de la production de la parole pathologique » (TPPP) qui correspond le mieux au paradigme qui nous intéresse et au sens « primitif » du mot dysarthrie. D'une part parce que cela n'exclut aucun type de trouble de la production notamment les troubles liés à des causes structurelles, comme les séquelles de cancers des voies aérodigestives qui, contrairement aux séquelles de fentes et à la parole des malentendants, sont systématiquement exclues des ouvrages anglo-saxons consacrés aux « motor speech disorders » (Yorston, Beukelman, Strand, Bell (1999), Duffy (2005)) ou sensorimotor speech disorders (MR McNeil 2009). D'autre part, cela élimine les troubles secondaires à des phénomènes non pathologiques comme les émotions, les situations langagières complexes favorisant les erreurs. L'équivalent anglais le plus proche serait donc tout simplement « speech disorders », sauf que celui-ci ne se restreint pas explicitement à la production.

#### **1.4. LA CONCEPTION DE LA PRISE EN CHARGE DES PATIENTS**

---

Le second concerne le cadre théorique dans lequel s'inscrit l'approche médicale, intégrée ici aux TPPP. Les différents modèles de santé issus de la recherche de ces 30 dernières années ont conduit à l'émergence d'un modèle biomédical centré sur la maladie et l'organe en souffrance et d'un modèle psycho-social centré sur les facteurs psychologiques et sociaux, personnels et environnementaux. Cette évolution a donné naissance à plusieurs classifications publiées par l'Organisation Mondiale de la Santé (OMS) (figure n°1) : la classification internationale des handicaps (CIH, 1988) traduction française de la « international classification of impairments, disabilities and disadvantages (ICDIH, 1980) puis la Classification internationale du fonctionnement, du handicap et de la santé (CIF, 2001) et à plusieurs méthodes d'évaluation (Fougeyrollas P 2002). Cependant, l'imbrication entre l'approche biomédicale et l'approche psychosociale reste à améliorer. La transdisciplinarité qu'elle implique reste difficile du fait des préoccupations pratiques différentes, avec une échelle d'action individuelle pour la première et un objectif de population pour l'autre. Et même au sein des disciplines médicales, cette approche reste méconnue bien que nous ayons tous ressenti ces effets sous la forme de la loi du handicap de 2005. C'est ce cadre qui oriente la démarche médicale implicite de ce travail et ceci dans la continuité de

l'approche proposée par Yorston, Beukelman, Strand, Bell KR dans la seconde édition de l'ouvrage intitulé « Management of motor speech disorders » de 1999.

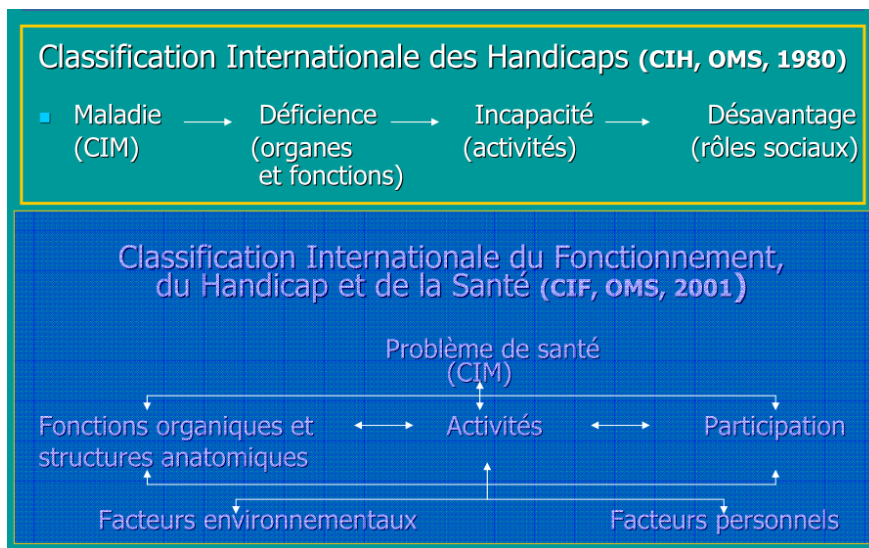


Figure n°1 : Les schémas conceptuels des classifications des handicaps de l'OMS extraits des documents édités par le Centre Technique National d'Etudes et de Recherches sur les Handicaps et les Inadaptations : CTNERHI [http://www.ctnerhi.com.fr/accueil\\_ctnerhi3.php](http://www.ctnerhi.com.fr/accueil_ctnerhi3.php)

Mais pour revenir à notre objet, « la prise en charge des personnes présentant un trouble de la production de la parole pathologique », comment trouver une méthode d'analyse du trouble de la production qui permet d'organiser et de générer de nouvelles idées dans le domaine thérapeutique ? Et par quel bout prendre le problème ? En effet, toutes les méthodes que j'ai pu expérimenter sont longues, fastidieuses, pour le patient et pour l'examinateur. Les approches thérapeutiques qui en découlent sont peu satisfaisantes dans la mesure où il n'y a pas de moyens pour prioriser les objectifs thérapeutiques en fonction de leur efficacité à pouvoir améliorer la qualité de vie de la personne. En effet, un défaut de production de la parole a très souvent plusieurs niveaux lésionnels, il induit des compensations qui, si elles ont des effets souvent positifs, peuvent être négatifs. Et finalement, c'est l'impact relatif sur la qualité de la communication du patient qui permettra d'adopter la bonne stratégie de prise en charge, d'en prédire les limites et de la ré-ajuster au fur et à mesure de l'évolution. Cette idée a été développée par Yorston KM, Beukelman DR, Strand EA, Bell KR avec la participation de Hustad KC dans un chapitre intitulé « Optimizing communicative effectiveness : bringing it together » de l'ouvrage cité ci-dessus.

## 1.5. QUESTIONNEMENT

Ces réflexions ont conduit à une liste de questions pour lesquelles différents axes de recherche pouvaient être développés. Elles sont regroupées dans le tableau n°1.

1. Pourquoi ne comprenons nous pas les malades qui articulent mal ?
1.1. Y a t-il une différence entre les erreurs de la parole normale et celle de la parole pathologique ? →Comparaison parole normale/parole pathologique
1.2. A quel niveau se situe le trouble de l'intelligibilité du dysarthrique ? indices acoustiques, traits distinctifs, phonèmes, syllabes ?
1.3. Est-ce que le fait d'écouter de la parole pathologique induit des mécanismes de reconnaissance particuliers ? (est-ce les mêmes processus que pour la parole non-pathologique)
1.4. Est ce que les auditeurs « pathologiques » c'est à dire produisant une parole pathologique, ont une perception différente ?
2. Quels sont les facteurs qui impactent le plus l'intelligibilité ?
2.1. Le nombre d'erreurs différentes par rapport au nombre absolu d'erreur ?
2.2. Y-a t-il des structures acoustiques-phonétiques plus résistantes ou robustes ? Les phonèmes les plus robustes ? des traits plus robustes ? des indices acoustiques ?
2.3. Quelle est l'importance des facteurs temporels par rapport aux facteurs acoustiques ?
2.4. S'agit-il d'une entité relative à un nombre de distorsions rapportée à une unité sémantique (plus le sens de l'unité est facile à trouver, plus le nombre d'indices nécessaire est faible).
3. Peut-on trouver les éléments fondamentaux du geste articulatoire minimum ? (à l'opposé de la production prototypique)
3.1. Quelles sont les limites de vitesse d'articulation (min et max)? Quels sont les rapports entre la vitesse des différentes structures articulatoires et la vitesse du signal au final) ?
3.2. Y a t-il des organes articulatoires plus importants ? (en fonction du type de langue) ?
3.3. Comment faire la part entre les compensations et les déficits ? (ex : problème du ralentissement pour prendre le temps de bien réaliser le mouvement et du ralentissement car la personne ne peut pas aller plus vite).
4. Comment construire un test qui réponde aux principales questions ?
4.1. Le choix d'un corpus pertinent sous une forme pertinente
4.2. La variabilité de la parole étant une caractéristique intrinsèque, rechercher si des normes prenant en compte cette nécessaire variabilité ne seraient pas pertinentes
4.3. Quelle est la méthodologie à adopter pour construire ce test ? (la détection de cible paraît difficile vu la fréquence des distorsions, l'accès au sens paraît plus écologique).
5. Quelles sont les pistes qui en découlent pour améliorer l'intelligibilité de la parole dysarthrique ?
5.1. Améliorer la qualité de la production du locuteur : renforcer ce qui est pertinent (Cibler la correction du déficit par la chirurgie, la rééducation ... Optimiser les stratégies de compensation (monémiques, sémantiques...)

## 5.2. Pour améliorer la perception de l'auditeur : La correction du signal de parole ? Comment ?

En partant comme Yorkston et al (1996) du principe fondamental que le trouble de la parole induit un handicap de communication, le paramètre fondamental paraît être l'intelligibilité du message émis. Et c'est sur l'intelligibilité de la parole que nous allons centrer notre questionnement.

L'impact des TPPP sur l'intelligibilité de la parole est ainsi apparu déterminant sur la communication et la vie de relation de la personne. Elle conditionne les stratégies de compensation qu'utilisera le locuteur pour rester compréhensible pour son ou ses auditeurs. Son étude peut conduire à la détermination des indices acoustiques minimaux pour la perception de la parole. De plus, les applications de l'étude de l'impact sur l'intelligibilité des troubles de la production de la parole pathologique dans le domaine médicale sont l'élaboration d'une méthodologie adaptée pour l'analyse de la parole pathologique et le développement des moyens de réhabilitation. Nous sommes ainsi arrivés à cibler les questions qui ont guidé les 3 grands groupes d'expérience de cette thèse.

Question 1 : L'ensemble de ce travail est fondé sur le fait que l'intelligibilité est un élément fondamental pour la gestion en clinique des TPPP. Cependant, dans la mesure où aucune étude perceptive n'a validé ce concept et que cela n'a pas toujours été mis systématiquement en avant dans la conception des bilans cliniques, une vérification préalable par une expérimentation adaptée a été réalisée.

Question 2 : Dans les TPPP, l'intelligibilité de la parole est souvent assimilée à la notion de sévérité du TPPP. Ceci est basé sur l'hypothèse que plus l'intelligibilité de la parole est altérée, plus le trouble est sévère. Dans la littérature, cet amalgame conduit, dans de nombreuses publications, à assimiler les deux notions. Les expériences menées pour répondre à la question 1 nous ont conduit à analyser le comportement d'un jury d'écoute pour expliciter ces deux notions.

Question 3 : Les TPPP sont caractérisés par la fréquence et le fort degré des distorsions de la production de la parole. La correction des distorsions sur un modèle de « normalisation » du schéma moteur de la production est une des bases du travail de réhabilitation. Mais est-ce que toutes les distorsions doivent être corrigées au même niveau ? Les distorsions en fonction des phonèmes ont-elles le même impact sur l'intelligibilité ?

En vérité, cette question a été la première posée chronologiquement, mais a demandé plusieurs années de réflexions, en raison des difficultés à trouver les moyens d'y répondre. Elle a finalement donné naissance à des hypothèses que nous avons essayé de valider et qui l'ont fait glisser à la dernière place.





*Intelligibility is part of a more general concern that we may call communicative competence [...]. But even if the best attempts are made to distinguish intelligibility from other aspects of communicative competence, it quickly becomes clear from a study of the literature that intelligibility is not a monolithic phenomenon [...]. Therefore, although two persons may share some common idea of what intelligibility is, they may use very different methods to measure it and understands its correlates in the act of speaking.*

*Raymond D. Kent 1992 in Intelligibility in speech disorders*

## **2. DE L'INTELLIGIBILITE AUX TROUBLES DE LA PRODUCTION DE LA PAROLE PATHOLOGIQUE**

---

2.	De l'intelligibilité aux troubles de la production de la parole pathologique	21
2.1.	Qu'est ce que l'intelligibilité ?	23
2.1.1.	Définition	23
2.1.2.	Usage actuel	23
2.2.	Qu'est ce que l'intelligibilité de la parole ?	24
2.2.1.	Définition	24
2.2.1.	Les différentes applications	25
2.2.1.1.	Du domaine des télécommunications à l'installation permanente d'équipements de sonorisation	26
2.2.1.2.	Des troubles de la perception de la parole aux troubles de la production de la parole	30
2.3.	Les différents cadres théoriques de références	34
2.3.1.	Les différents modèles de communication	35
2.3.2.	De la perception de la parole à la compréhension du discours en conversation	40
2.3.2.1.	La perception de la parole	42
2.3.2.1.1.	La perception des phonèmes est-elle catégorielle ?	43
2.3.2.1.2.	Existe t-il ou non une unité minimale de perception ?	46
2.3.2.1.3.	L'identification de mots isolés ou comment s'effectue l'accès au lexique mental ?	49
2.3.2.1.4.	La perception de la parole continue ou comment segmentons-nous le signal de parole ?	52
2.3.2.1.5.	Comment perçoit-on la parole dégradée ?	54
2.3.2.2.	La compréhension du discours oral	57
2.3.2.2.1.	La compréhension de texte et le modèle de Kintsch et Van Dijk (1978, 1983)	57
2.3.2.2.2.	De la compréhension d'un discours oral au discours en interaction	60
2.3.2.2.3.	La compréhension de la parole dégradée	63
2.4.	Les principes de l'évaluation de l'intelligibilité de la parole	66
2.4.1.	Les grands courants	66
2.4.2.	Les différentes manières de mesurer l'intelligibilité.	70
2.4.2.1.	Les mesures subjectives	71
2.4.2.1.1.	Les tests par estimation directe d'amplitude	71
2.4.2.1.2.	Les tests « explicatifs »	72
2.4.2.1.3.	Les tests de reconnaissances d'unités linguistiques	73
2.4.2.2.	Les mesures prédictives	74
2.5.	Quels sont les facteurs déterminant l'intelligibilité dans les troubles de la production de la parole	75
2.5.1.	Travaux issus du domaine des TPPP	75
2.5.1.1.	Informations provenant de l'exploration des TPPP	75
2.5.1.2.	Travaux issus de la manipulation de signal dans le cadre des TPPP	76
2.5.2.	Travaux issus de la perception dans le bruit	77
2.5.3.	Travaux issus de la manipulation de signaux hors contexte des TPPP	78
2.5.3.1.	La parole compressée	79
2.5.3.2.	Dans le domaine fréquentiel	79
2.5.3.3.	L'enveloppe temporelle	80
2.5.3.4.	La restauration phonémique	81
2.5.4.	Le caractère naturel de la parole	82
2.5.5.	En synthèse	83

## 2.1. QU'EST CE QUE L'INTELLIGIBILITÉ ?

---

### 2.1.1. DÉFINITION

---

Dans le dictionnaire historique de la langue française dirigé par Alain Rey (Rey, 1998), le nom « intelligibilité » (1712, Fénelon) est défini à partir de l'adjectif « intelligible ». Emprunté vers 1265 au latin *intelligibilis* : « qu'on peut comprendre », il s'applique d'abord, en philosophie, à ce qui ne peut être connu que par l'intelligence. Il prend ensuite dans l'usage courant le sens (1521) de « qui est plus ou moins aisé à comprendre », d'où, en parlant d'une personne « qui se fait comprendre », en général par le langage, et (1538), « qui peut être distinctement perçu par l'ouïe », donc « compris ».

L'utilisation actuelle du terme « intelligibilité » reflète son historique.

### 2.1.2. USAGE ACTUEL

---

Dans le sens général de « compréhensibilité », ce mot est largement employé dans le domaine juridique. Ainsi, P. Malaurie, dans un article intitulé « l'intelligibilité des lois » (Malaurie, 2005) écrit : « une loi inintelligible est une mascarade juridique. Seule la loi claire, simple, limpide, transparente, compréhensible de tous peut être respectée, devenir efficace et assurer ce que l'on peut attendre du droit : la justice, l'ordre, la prévisibilité, la sécurité, le bien-être, la paix et peut-être le bonheur. Pourquoi d'ailleurs se limiter aux lois ? L'intelligibilité est nécessaire à tout le droit : constitutions, traités internationaux, décisions judiciaires, actes administratifs, actes publics et privés. Le droit n'a de sens que s'il est compris de tous et qu'il est ainsi populaire. »

Ce terme s'applique ainsi à tout type de texte. « Clarté, lisibilité, intelligibilité des textes : un état de la question et une proposition pédagogique » est le titre d'un article rédigé par C. Baudet (Beaudet, 2001). « L'enseignement de la rédaction exige l'adoption du postulat qu'un texte intelligible, compris par son destinataire, possède des qualités, des attributs qu'on peut décrire, en dehors des aptitudes cognitives des lecteurs.

Sur le plan philosophique, c'est la possibilité pour un phénomène d'être saisissable par l'intellect, l'esprit, la raison. D'actualité, la question « *De quoi parlons-nous lorsque nous parlons d'intelligibilité ?* » a fait l'objet d'un colloque en mai 2009 au Canada animé par le Dr Louis Perron de la Faculté de philosophie de l'Université Saint-Paul à Ottawa : « Sur quelle(s) base(s) peut-on revendiquer l'intelligibilité d'un savoir, d'une pratique, d'un domaine particulier ou encore de l'ensemble du réel ? Quelles sont les conditions de possibilité de l'intelligibilité ? ». La question de la nature et de l'origine de l'intelligibilité, commune aux différents domaines de l'activité scientifique ainsi qu'à la philosophie et à la théologie, est celle que cette conférence se proposait d'explorer.

Dans le domaine de la parole, l'intelligibilité désigne la qualité de compréhension d'un message linguistique contenu dans un signal de parole et relève des domaines linguistique, acoustique et/ou médicale.

Ainsi, ce terme est difficile à différencier de la compréhension qui vient, lui du latin « *comprehensio* », dérivé de « *comprehendere* » formé de *cum* (avec) et *prehendere* (prendre) action de saisir ensemble, par l'intelligence, la pensée, saisir implicitement le sens.

Nous pouvons retenir que l'intelligibilité est la capacité à être compris alors que la compréhension est la capacité de comprendre. L'hypothèse selon laquelle « intelligibilité » existe en l'absence de développement du mot « compréhensibilité » est probable. Mais cette interprétation s'applique-t-elle au domaine de la parole ?

## **2.2. QU'EST CE QUE L'INTELLIGIBILITE DE LA PAROLE ?**

---

### **2.2.1. DEFINITION**

---

L'intelligibilité de la parole n'est pas une notion simple. Le but de la communication parlée est la transmission de sens et non la transmission de mots ou de phonèmes. La communication se fait dans un certain environnement et dans une situation, dont les interlocuteurs ont conscience, en même temps qu'ils sont censés utiliser les mêmes codes linguistiques. De plus la parole transmet des informations qui contribuent à l'élaboration du sens, mais qui ne sont pas de nature linguistique.

La notion même d'intelligibilité doit être considérée différemment selon qu'il s'agisse de caractériser la production de parole par un locuteur, la transmission de parole par un appareillage électroacoustique ou par une salle, ou la perception de parole par un auditeur.

Ainsi, l'intelligibilité de la parole peut être abordée de différents points de vue, s'adresser à différents types de la communication par rapport aux locuteurs : dialogue, information de masse, cibler différents niveaux de traitement du signal de parole : décodage phonétique, accès au lexique, accès au sens...Il en découle de nombreuses définitions laissant transparaître les préoccupations liées aux domaines d'applications dont elles sont issues. Nous citerons ici :

- « L'intelligibilité de la parole peut être définie comme la concordance entre l'intention du locuteur et la réponse de l'auditeur à la parole transférée à travers le système de transmission » (Schiavetti, 1992)

- « L'intelligibilité de la parole est une expression décrivant la qualité d'écoute et de compréhension d'un échange dans une pièce » (Lambour 2002)

- « L'intelligibilité est le degré de précision avec lequel le message est compris par l'auditeur » (Yorkston 1992)

---

### 2.2.1. LES DIFFÉRENTES APPLICATIONS

---

Historiquement, les besoins de mesure sont apparus dans le domaine de la télécommunication et ont été modélisés par la théorie de l'information de Shannon (1948), ingénieur à la Compagnie des Téléphones Bell. Elle reste jusqu'à nos jours la théorie princeps dite scientifique d'information à la base de la vision « classique » de la communication linguistique par un processus d'encodage – décodage. Le cadre originel de la théorie est celui d'un système de communication dans lequel un émetteur transmet un message à un récepteur à travers un canal matériel/énergétique donné. Emetteur et récepteur ont par hypothèse un répertoire commun, un code qui contient les catégories de signaux utilisables. Ainsi le message codé est transmis, de l'émetteur au récepteur à travers le canal, sous forme de signes ou signaux portés par de la matière/énergie.

Les applications de l'évaluation de l'intelligibilité se sont ensuite développées dans des domaines aussi différents que l'industrie de la sonorisation des salles ou des lieux publics, l'apprentissage des langues, les troubles de la perception ou de la production pathologique de la parole.

## 2.2.1.1. DU DOMAINE DES TELECOMMUNICATIONS A L'INSTALLATION PERMANENTE D'EQUIPEMENTS DE SONORISATION

A l'origine, les systèmes de télécommunication cherchent à transmettre les messages de façon à la fois la plus économique et la plus fiable et donc à ce que la signification du message de départ subisse le moins de déformations possibles lors de la transmission et de la réception.

La qualité du transfert se mesure à un niveau acoustique, la performance étant la restitution la plus parfaite possible du signal acoustique.

Le niveau linguistique ciblé ici est de l'ordre du code acoustico-phonétique.

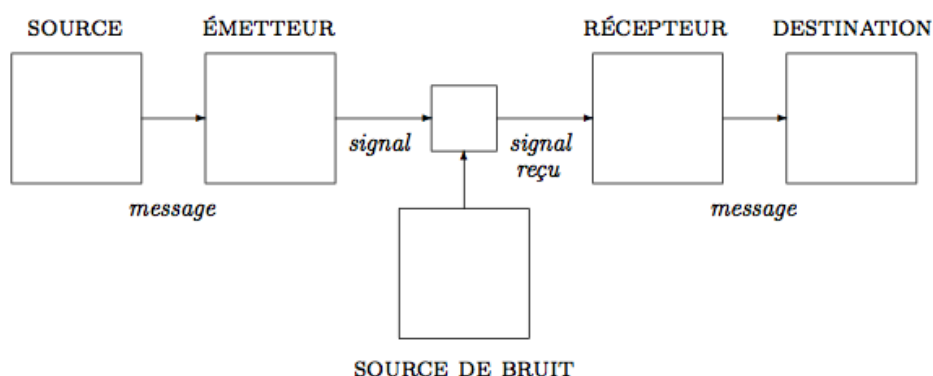


Figure n°1.2.1. Modèle linéaire, dit du télégraphe, d'après Shannon 1948

Ce modèle (figure n°1.2.1) reste la référence pour les équipements électroacoustiques. Il a conduit au développement de mesures de prédiction de l'intelligibilité et répond à des normes industrielles. L'intelligibilité de la parole devient l'intelligibilité de la sonorisation. Dans les gares, les aéroports, les stations de métro où les annonces sont indispensables à l'information et à la sécurité, des études d'intelligibilité de la parole et leur influence sur la perception du confort sont nécessaires. Afin de spécifier les systèmes électroacoustiques pour les services de secours, une norme européenne traduite en une norme Française (NF EN 60849 - traduction CEI 60849 :1998) a été promulguée puis modifiée en 2004 (NF EN 60268-16). Elle est publiée par l'association française de la normalisation (AFNOR 2004). Cette norme internationale s'applique aux systèmes de sonorisation et de diffusion utilisés pour provoquer la mise en alerte rapide et disciplinée des occupants d'une zone intérieure et/ou extérieure en situation d'urgence, utilisant des signaux à tonalité ou des annonces vocales. Elle définit les

caractéristiques et les méthodes d'essai qui sont nécessaires à la spécification du système. Les définitions suivantes sont applicables.

**Audibilité** : propriété des sons leur permettant d'être entendus parmi d'autres sons.

**Clarté** : propriété d'un son permettant à ses composantes porteuses d'informations d'être distinguées par un auditeur. Cette propriété est liée à l'absence de distorsion dans le son.

**Intelligibilité** : mesure de la quantité d'information contenue d'un message vocal pouvant être compris.

Une intelligibilité suffisante nécessite une audibilité et une clarté adéquate.

Dans cette norme, les méthodes mesurant la qualité d'une transmission vocale en fonction de l'intelligibilité des paroles sont classées en deux types (Lambour 2002) :

- méthodes subjectives ou directes nécessitant un jury d'auditeur et fondées sur le décompte de mots ou d'unités linguistiques correctement perçues.
- méthodes objectives ou indirectes reposant sur des mesures physiques.

Cette norme fournit des indications sur la procédure d'essai mais pas de procédures complètes et précises.

Parmi les mesures objectives, l'Indice de transmission de la parole (STI, Speech Transmission Index) mesure l'effet combiné du bruit de fond et de l'acoustique d'une salle sur l'intelligibilité des paroles, c'est à dire comment le bruit de fond et l'acoustique d'une salle réduisent les fluctuations naturelles d'amplitude des sons de la voix, par la mise en œuvre d'un signal test le plus proche d'un signal de parole (Hugon 2008). Cet indice s'appuie sur le fait que la parole est un signal modulé en amplitude. L'intelligibilité est d'autant plus importante que les modulations sont bien perçues. Ce type d'indicateur est aussi bien utile aux opérateurs de télécommunications, qu'au développement de systèmes de reconnaissance vocale. Il en existe plusieurs variantes dont :

- le RASTI (Room Acoustic STI) mesurée en émettant un signal sinusoïdal de fréquence  $f$  modulé en amplitude. Au point de réception est enregistré un signal dont la modulation est plus faible. On calcule la réduction de modulation.
- le STI-3, qui ne prend en compte que trois bandes d'octaves ;



- le STITEL (STI for TELEcommunication channels) qui vise à évaluer les systèmes de télécommunication ;

- le STIPA (STI for Public Address systems), qui a été développé pour les systèmes de sonorisation.

Parmi les méthodes subjectives, deux classes de tests se distinguent par les tâches que le jury doit réaliser et le type de matériel phonétique utilisé :

- Test sur grand corpus avec choix libre

Les éléments d'une liste de mots ou de logatomes sont soumis un par un à un jury, qui note ce qu'il a compris. Le résultat final est exprimé par le pourcentage de mots ou de phonèmes correctement identifiés. La norme précise que chaque liste doit contenir un minimum de 50 mots ou logatomes choisis dans un grand corpus initial (entre 500 et 1 000 éléments). Pour ce type de test, les auditeurs ne doivent pas avoir de connaissance a priori des choix possibles (corpus initial). Pour réduire la dispersion des résultats, les listes doivent être d'égale difficulté et «phonétiquement équilibrées», c'est-à-dire contenir le même nombre de types de phonèmes, et si possible dans des proportions représentatives de la parole courante considérée.

- Test sur des petits corpus fermés avec choix forcé.

Les éléments du test sont classés dans des sous-ensembles de 2 à 10 éléments (mots ou logatomes) qui se distinguent uniquement par une consonne, toujours à la même place. On parle souvent de « tests de rime ». Le nombre de sous-ensembles doit être suffisant pour pouvoir tester de manière systématique toutes les consonnes, associées avec au moins plusieurs voyelles différentes. Dans ce type de test, le jury doit identifier parmi la liste des éléments de chaque sous-ensemble, le mot tiré au hasard qu'il vient d'écouter. Les réponses sont notées sur un questionnaire à choix multiples, ce qui permet d'automatiser facilement l'étape de dépouillement des résultats.

Le nombre de choix possibles (nombre d'éléments dans chaque sous-ensemble) conditionne le pourcentage de réponses correctes. Sur 2 alternatives possibles, un auditeur a par exemple 50 % de chances de trouver la bonne réponse par hasard même si l'intelligibilité est nulle. Pour pouvoir comparer les résultats des tests par choix forcé avec ceux obtenus par test sur grand corpus, une correction doit donc être apportée.

Quel que soit le type de test, la norme conseille de présenter au moins 3 listes au jury pour chaque difficulté. Elle insiste sur l'importance de faire précéder chaque élément par une phrase porteuse, destinée en particulier à focaliser l'attention de l'auditeur et à exciter la réverbération le cas échéant.

Pour un matériel phonétique, un jury et une difficulté donnés, l'évolution statistique des résultats au cours de tests successifs peut être décomposée en trois parties : la première correspond à la durée d'apprentissage de l'auditeur, pendant laquelle les résultats obtenus s'améliorent de test en test. Suit ensuite la période utile du test, au cours de laquelle les résultats sont statistiquement stationnaires. Enfin, pour les listes « pseudo-ouvertes », les auditeurs ont tendance à connaître les éléments après un certain nombre de présentations, ce qui entraîne une nouvelle phase d'amélioration des scores.

Le tableau 1.2.1. permet de comparer les durées d'apprentissage et de validité typiques pour les deux types de tests. Selon ces critères, on note un net intérêt pour les tests sur petits corpus avec choix forcé.

	Essais sur des grands corpus (basés sur environ 20 listes de 50 éléments de test, le contenu de chaque liste étant redistribué)		Essais sur des petits corpus fermés	
	Logatomes	Mots	Logatomes	Mots
<b>Durée typique d'apprentissage de l'auditeur</b>	24 heures	12 heures	2 heures	5 minutes
<b>Durée typique de test autorisée après apprentissage de l'auditeur</b>	200 heures	120 heures	Pas de limite	Pas de limite

Tableau 1.2.1. Comparaison des deux types de test extraits de la norme ISO TR 4870 d'après Lambour (2002).

La période d'apprentissage est très courte, en raison de la simplicité de la tâche à réaliser (remplir un questionnaire à choix multiples). Il n'y a d'autre part aucune limitation sur la durée de validité du matériel phonétique utilisé.

Ce n'est pas le cas en ce qui concerne les essais sur listes « pseudo-ouvertes » qui doivent être renouvelées régulièrement pour un jury donné, et pour lesquelles une période d'apprentissage de plusieurs heures est nécessaire. De plus, la meilleure précision des tests par choix forcé permet de réduire considérablement le nombre de personnes à interroger par rapport aux tests sur listes «pseudo-

ouvertes » : pour l'étalonnage, la norme préconise un effectif minimum de 5 auditeurs dans le premier cas, contre 10 auditeurs dans le second cas.

De nombreuses études ont porté sur la comparaison des critères objectifs et des résultats de tests d'intelligibilité. Il est ainsi déterminé que la méthode de l'indice de transmission de la parole (STI) a la plus grande discrimination pour les valeurs élevées de l'intelligibilité, alors que le décompte de mots phonétiquement équilibrés parmi 256 mots a la plus grande discrimination pour les valeurs faibles. Il existe également des courbes d'équivalence entre les différentes mesures d'intelligibilité décrites dans la norme ISO avec des discordances significatives dans certaines zones de l'étendue de l'intelligibilité.

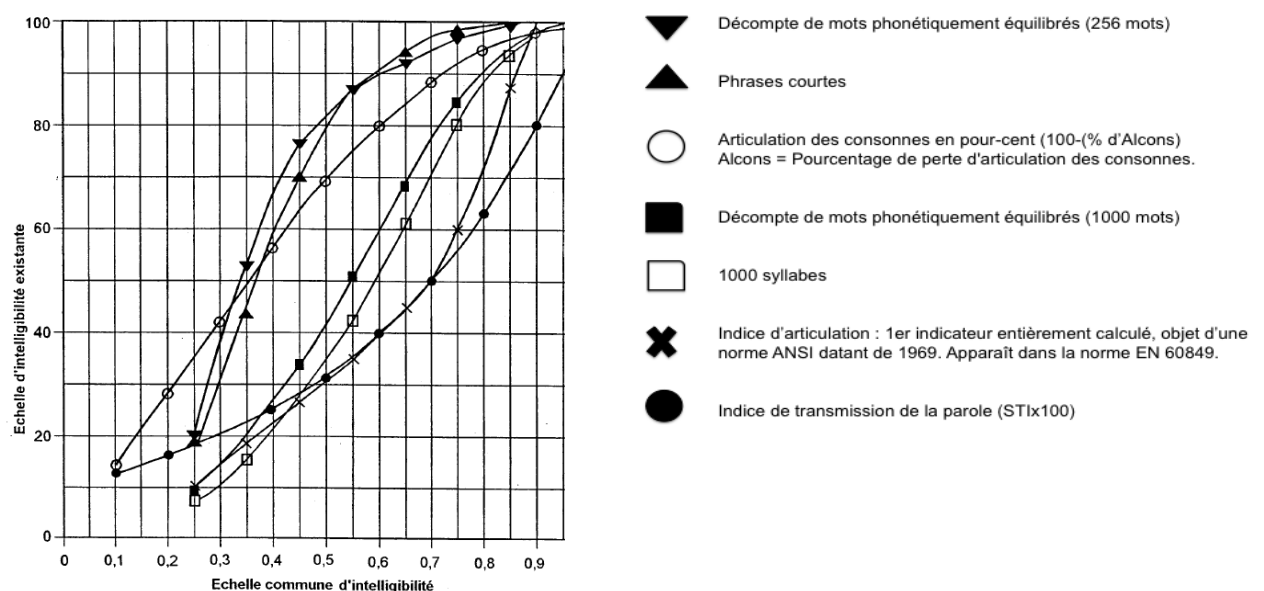


Figure n°1.2.2. La CIS ou échelle commune d'intelligibilité et ses corrélations avec différents tests de mesure de l'intelligibilité extrait de la norme EN NF 60849.

Lambour (2002) conclue que «la lourdeur de mise en œuvre des tests basés sur le décompte de phrases, de mots ou de phonèmes correctement reconnus par un jury, qui représentent la majorité des méthodes proposées par la norme, les rendent quasiment inapplicables dans la plupart des cas d'étude, à moins d'effectuer les tests en laboratoire à partir de signaux enregistrés ou simulés par convolution ». Ceci explique probablement la pauvreté des références quand aux sources validées de listes de mots ou autres unités linguistiques pour ces tests (Combescure 1981, Zialani 1988).

## 2.2.1.2. DES TROUBLES DE LA PERCEPTION DE LA PAROLE AUX TROUBLES DE LA PRODUCTION DE LA PAROLE

Les champs d'applications s'étendent de la surdité et des troubles de la perception dans le bruit, aux troubles de la production pathologique de la parole (TPPP) et à l'apprentissage des langues.

L'impact de la perte auditive de l'auditeur sur sa perception de la parole définit le cadre de l'intelligibilité dans le domaine de la surdité. L'impact du bruit sur la perception de parole associé à une perte auditive est devenu avec l'évolution du statut du bruit dans nos sociétés un enjeu majeur (Jilliot et al, 2008). Ici aussi, les aspects industriels liés à la prothèse auditive et maintenant aux implants cochléaires ont favorisé la normalisation des évaluations. L'audiomètre doit satisfaire à des normes de construction établies de façon consensuelle par la Commission Électrotechnique Internationale. Le cadre théorique reste celui du modèle de Shannon mais ce ne sont plus les caractéristiques du canal de transmission qui sont à tester, c'est la dégradation de la discrimination perceptive de la parole (Fournier 1951).

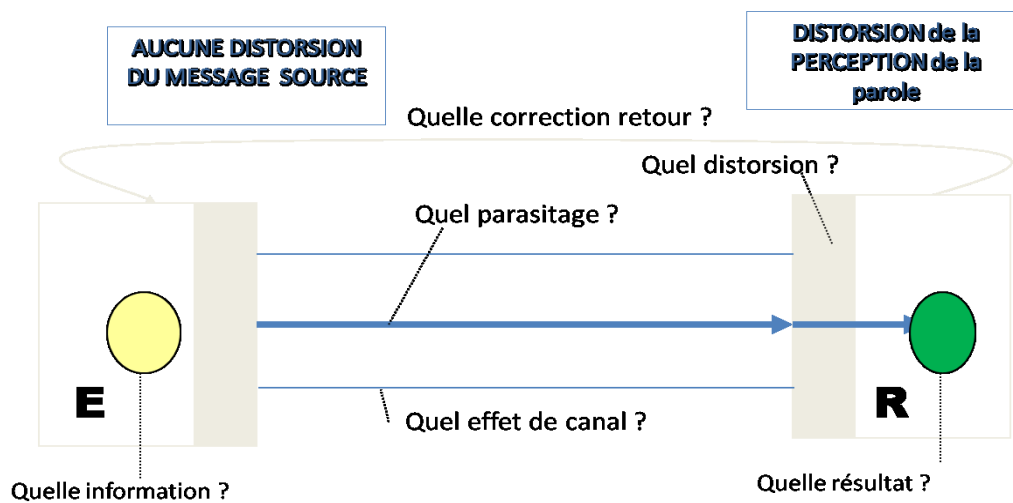


Figure n°1.2.3. Modèle de Shannon spécifié pour les troubles de l'audition

En pratique, les tests utilisés sont par essence subjectifs dans la mesure où l'objet de la mesure est la perception du sujet. L'audiométrie vocale est une mesure globale de la fonction auditive chargée d'évaluer l'intelligibilité de la parole. La norme française et européenne NF EN ISO (389-1-5, 1994a-F) et (8253-3 1998) définit le mode opératoire ainsi que les messages vocaux pour la pratique de cette technique qui assure les mesures de la performance auditive par une méthode de détection de seuil. Le matériel vocal doit être enregistré ce qui a pour avantage de garantir le niveau (pression acoustique) et la qualité sonore du message vocal, indépendamment de l'opérateur. L'enregistrement doit cependant obéir à des règles précises, et comporter notamment un signal d'étalonnage et des signaux permettant de contrôler la distorsion harmonique de l'audiomètre vocal. Les listes de mots

sont davantage utilisées que les listes de phrases. En effet, ces dernières sont soumises à l'influence sémantique, facilitant la compréhension de la phrase du fait de la suppléance mentale. Les listes les plus utilisées en français sont celles de Fournier (1951) et de Lafon (1964). Ce dernier a particulièrement travaillé pour établir des listes adaptées aux différents types de surdité. Elles évaluent les distorsions du message acoustique selon le niveau d'atteinte du système auditif et sont très employées pour l'adaptation prothétique.

Les listes de Fournier sont les plus utilisées en première intention. L'unité d'erreur est le mot (dissyllabique). L'influence sémantique est forte, la lecture labiale peut être mesurée. Cependant, ces listes sont mal adaptées à l'analyse des distorsions d'origine cochléaire. Elles ne sont pas phonétiquement équilibrées: la proportion des différents phonèmes n'est pas identique d'une liste à l'autre et ce n'est pas non plus celle de la langue française parlée. En outre, elles sont de difficultés inégales.

Lafon (1964) a proposé des listes qui ont l'avantage d'être mieux équilibrées phonétiquement. De plus, il a décrit un test de mesure des distorsions cochléaires qui se pratique en supraliminaire, dans le champ auditif résiduel. L'unité d'erreur est le phonème, la suppléance mentale est d'autant plus réduite que le patient doit répondre instantanément (en écholalie).

Il s'agit de vingt listes phonétiquement équilibrées comportant 17 mots de trois phonèmes chacun. On assimile le nombre de phonèmes par liste à 50. On compte le nombre de phonèmes non perçus ou erronés. Le fait de répéter le mot avec un phonème surnuméraire n'est pas comptabilisé comme une erreur. La non répétition d'un mot nécessite de le proposer à nouveau en fin de liste.

La courbe d'intelligibilité vocale est tracée à partir des scores de reconnaissance vocale pour différents niveaux de parole. Le score de reconnaissance vocale est le pourcentage d'items correctement répétés dans une liste. L'épreuve débute à un niveau situé 20dB au-dessus de la moyenne du seuil tonal des fréquences conversationnelles (500, 1 000 et 2 000 Hz). Le score de reconnaissance vocale est déterminé en diminuant le niveau sonore à chaque liste de 5 à 10 dB selon la réponse. En modifiant le niveau vocal, on obtient une série de points qui dessinent la courbe d'intelligibilité vocale. Il faut en général 5 à 10 points pour tracer une courbe complète sur laquelle on peut identifier le seuil d'intelligibilité vocale.

A partir de la perception de sons purs (audiométrie tonale) ont été proposés des indices pronostiques ou prédictifs de l'intelligibilité : l'indice d'articulation ou du Speech Intelligibility Index (SII). Il s'agit de mesures objectives physiques dont l'origine est le laboratoire Bell Téléphone. Comme pour leurs applications dans les environnements réverbérant, ils sont décrits dans l'American National Standard Methods for the Calculation of the speech Intelligibility Index (ANSI. 1997). Mais s'il existe une corrélation entre ce type d'indice et l'intelligibilité mesurée, les scores mesurés sont inférieurs à ceux prédits et le désaccord est dépendant du degré de perte auditive.

En ce qui concerne le bruit, les travaux de recherche sont focalisés sur l'impact du bruit sur la perception des unités phonétiques dans le but de faire progresser les aides auditives et de préciser les modalités de réglages des implants cochléaires.

L'apprentissage des langues est un domaine d'application où l'intelligibilité est communément considérée comme un objectif approprié pour l'enseignement de la prononciation en langue seconde. Cependant, il n'existe pas non plus dans ce domaine de définition « universelle » de l'intelligibilité, ni de consensus sur la meilleure façon de la mesurer. En outre, il existe peu d'évidence empirique concernant les caractéristiques de la prononciation qui sont essentielles à l'intelligibilité, ce qui limite les professeurs dans leurs choix pédagogiques (Isaacs T 2008).

De nombreuses études ont montré que l'intelligibilité au niveau purement phonétique (testé avec des syllabes sans signification par exemple) est plus faible pour les apprenants de langue seconde (L2) et que les erreurs sont fortement influencées par la relation entre les systèmes phonémiques de L1 (langue première) et de L2 (Van Dommelen & Hazan V 2010). Les sujets sont aussi beaucoup plus affectés par la dégradation du signal de L2 (addition de bruit ou réverbération) que les sujets dont L2 est la langue maternelle, et cela même si la maîtrise de L2 est excellente. Compte tenu des difficultés éprouvées par les locuteurs de L2 à percevoir la langue au niveau acoustico-phonétique, il semble que la compréhension de L2 requiert une charge cognitive plus forte que pour la langue maternelle. Il semble aussi que l'information sémantique acquiert une plus grande importance.

À ce propos, Lado (1964) distinguait le niveau de communication et le niveau linguistique. En termes de communication, la substitution d'un autre système de sons produit un bruit qui fait chuter l'intelligibilité du message transmis. En termes linguistiques, les distinctions phonologiques qui servent à identifier les mots et les phrases de la langue-cible peuvent ne pas être identifiées, l'auditeur de L2

utilisant son système phonologique natif, créant une ambiguïté (Zielinski 2008). La langue présentant des redondances considérables, le message total sera encore transmis, mais les chances de mauvaise compréhension ou de perte d'information augmentent avec chaque signal linguistique non différencié. Cependant, si la notion d'intelligibilité prégnante dans l'efficacité de l'apprentissage des langues, elle n'a pas conduit au développement de test spécifique (Hayes-Harbs et al 2008).

L'apprentissage des langues ou la production d'une langue étrangère est la situation qui se rapproche le plus des TPPP.

En utilisant toujours le modèle de Shannon (figure n°1.4), le locuteur produit un signal de parole dont le contenu en terme de traits acoustiques est distordu soit par une méconnaissance des caractéristiques phonétiques de la langue soit par l'altération de la production du signal acoustique. Le canal de transmission est considéré comme non bruité et l'auditeur est supposé avoir une parfaite capacité de perception auditive.

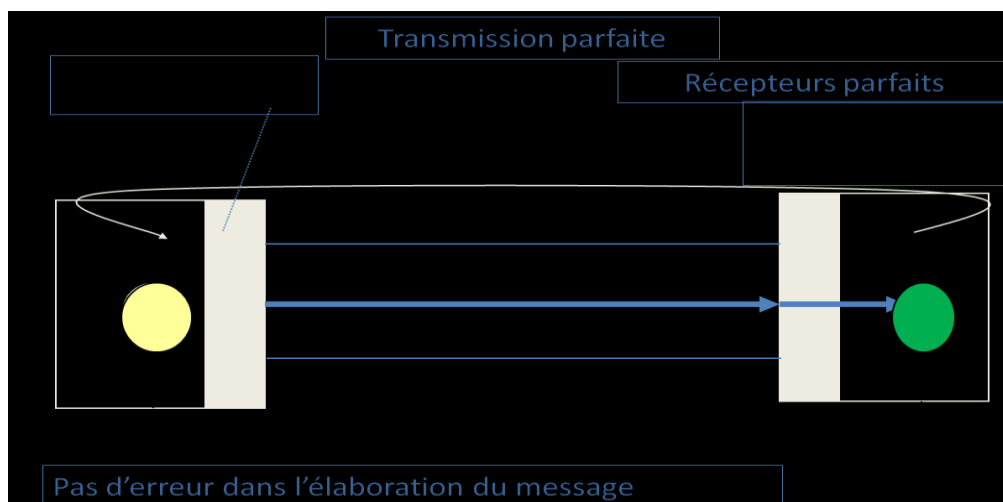


Figure n°1.2.4. Modèle de Shannon spécifié pour les troubles de la production de la parole

Si ce cadre théorique a dominé la conceptualisation de l'intelligibilité pendant tout le XX<sup>ème</sup> siècle, différents auteurs et particulièrement KC Hustadt (2002, 2003, 2008 a et b) en se référant aux théories de la compréhension du discours ont ouvert de nouveaux horizons. Nous allons donc rechercher les différents cadres théoriques qui pourraient nous aider à appréhender la notion d'intelligibilité en vue de l'application aux TPPP : les situations de communication, la perception de la parole et la compréhension du discours en interaction.

### 2.3. LES DIFFERENTS CADRES THEORIQUES DE REFERENCES

Dans ce chapitre l'objectif est de recenser de manière non éclectique les théories et les modèles qui peuvent contribuer à positionner la problématique de l'intelligibilité dans la prise en charge des TPPP (du diagnostic à la remédiation). En effet, la parole comme le langage s'inscrivant dans la fonction de communication, les TPPP sont en pratique sources de troubles de la communication invalidants pour les personnes qui en sont atteintes.

---

### 2.3.1. LES DIFFERENTS MODELES DE COMMUNICATION

---

Mucchielli (2000) introduit ainsi le chapitre sur les modèles de la communication dans l'ouvrage « La communication : état des savoirs » :

« Les chercheurs ont produits divers modèles pour expliquer la communication. Chaque modèle – lié à un contexte, à une époque et à un projet scientifique différent - agit comme un mécanisme perceptif et cognitif qui transforme la réalité en représentation. Il permet ainsi d'en voir certains aspects mais, obligatoirement, en occulte d'autres ».

En adoptant un point de vue relativiste et constructiviste, il aborde les modèles de la communication dans leur ordre d'apparition historique, expliquant les assemblages théoriques sous jacents et les préoccupations justifiant l'élaboration du modèle.

Ainsi, le premier modèle (positiviste) est celui au quel nous nous référons depuis le début de notre rédaction, le modèle « émetteur – récepteur » mis au point par Claude Shannon et Warren Weaver en 1945 et qui a donné lieu à la théorie mathématique de l'information (1948). Linéaire, il renvoie à la métaphore du télégraphe. Il est justifié par le fait que la signification du message de départ subit le moins de déformations possibles lors de la transmission et de la réception, le message étant un signal acoustique. Cependant cette définition de l'information ne pourrait s'appuyer ni sur la forme matérielle/énergétique, ni sur le contenu cognitif des messages émis : leur contenu sémantique est laissé de côté, de même que leur contenant physique, pour ne s'intéresser qu'aux aspects mathématiques.

En fait, un modèle antérieur doit être cité ici, illustrant parfaitement les propos introductifs de Mucchielli. Il s'agit du modèle d'information proposé par Ferdinand de Saussure dans le cours de



linguistique générale (1916,1995), modèle qui réduit la communication à un transfert d'ondes sonores dans la conception du signe linguistique (figure n°1.3.1).

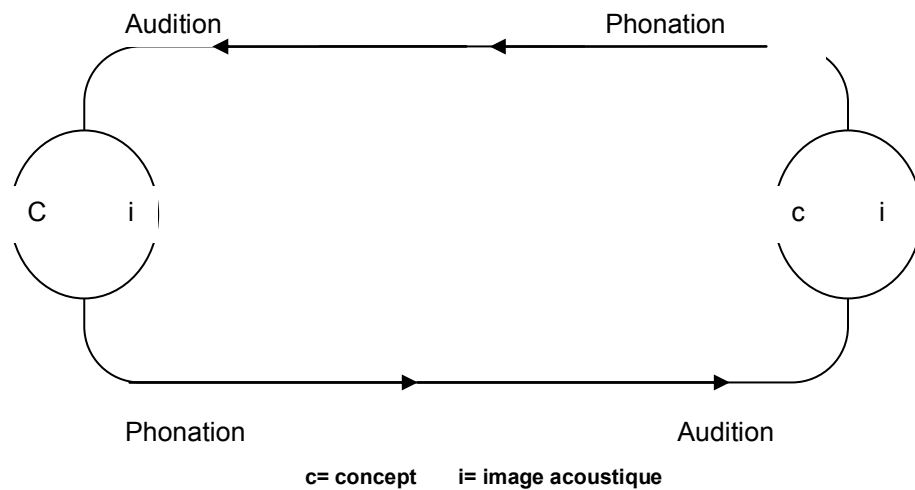


Figure n°1.3.1 : Concept d'information selon la linguistique moderne, cours de linguistique générale de F. de Saussure

Nous retiendrons ensuite le modèle sociométrique (1954) (figure n°1.3.2). Il présente graphiquement le réseau dessiné par les relations « informelles » dans un groupe. Pour la première fois émerge une « structure » où apparaît la notion de relation. Il est également le premier à être partiellement systémique. La communication y est définie comme une relation d'affinités (positive ou négative, sympathie, antipathie) mais en même temps, comme un canal, support privilégié du transport de l'information. La question principale qui sous tend ce modèle est : « Quel changement opérer pour améliorer la forme et la structure du réseau d'échanges ? ».

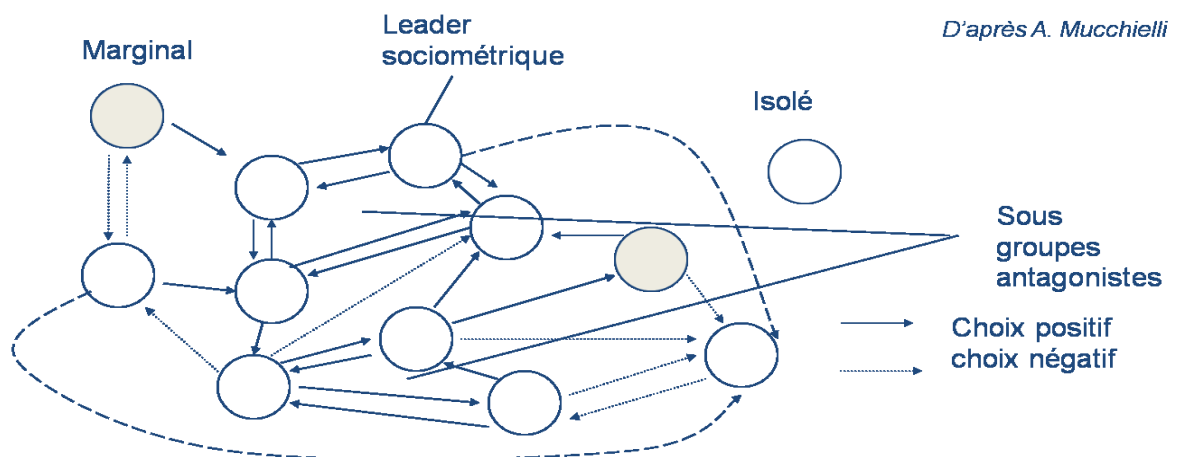


Figure n°1.3.2. Modèle sociométrique d'après A. Mucchielli (2000)

Pendant la même période Jakobson (1963) propose un modèle linguistique intitulé schéma de la communication en six pôles (figure n°1.3.3)

Le destinataire est le locuteur qui, lorsqu'il adresse un message qui trahit les pensées ou les sentiments, présente une fonction émotive ou expressive. Le destinataire est l'allocataire s'il n'intervient pas, l'interlocuteur s'il y a échange verbal. Il lui correspond la fonction conative. Est conatif ce qui est destiné à produire un certain effet sur le récepteur. Cette fonction est fondamentale : parler est un acte dirigé vers l'autre pour produire des effets variés.

Le contexte est à prendre au sens large. La fonction référentielle d'un message a trait à l'aspect purement informatif de l'énoncé. Le message est l'énoncé. Si l'on s'intéresse au message lui-même, aux signifiants plus qu'aux signifiés, c'est la fonction poétique de l'énoncé qui est prise en compte. La fonction phatique est ce qui permet d'établir ou de maintenir le contact avec l'interlocuteur, de vérifier qu'il est à l'écoute. Le code est la langue que nous utilisons pour communiquer avec l'autre. La fonction métalinguistique est la fonction réflexive du langage

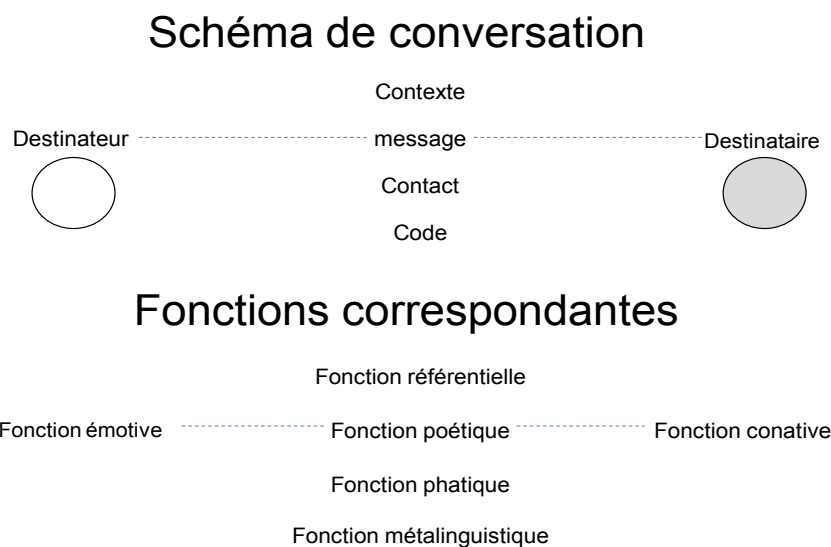


Figure n°1.3.3 : schéma de la communication en six pôles de Roman Jakobson

Le modèle en vogue dans l'année 1970 fut celui de l'analyse transactionnelle. Systémique, il rompt avec les autres modèles par le fait que l'on ne s'occupe plus du message mais de sa forme générale. Puis apparaît le modèle interactionniste (Figure n°1.3.4) et systémique principalement issu de l'école de Palo Alto. La communication est envisagée comme la participation d'un individu à un système

d'interactions qui le relie aux autres en s'intéressant au système des échanges et aux communications paradoxales.

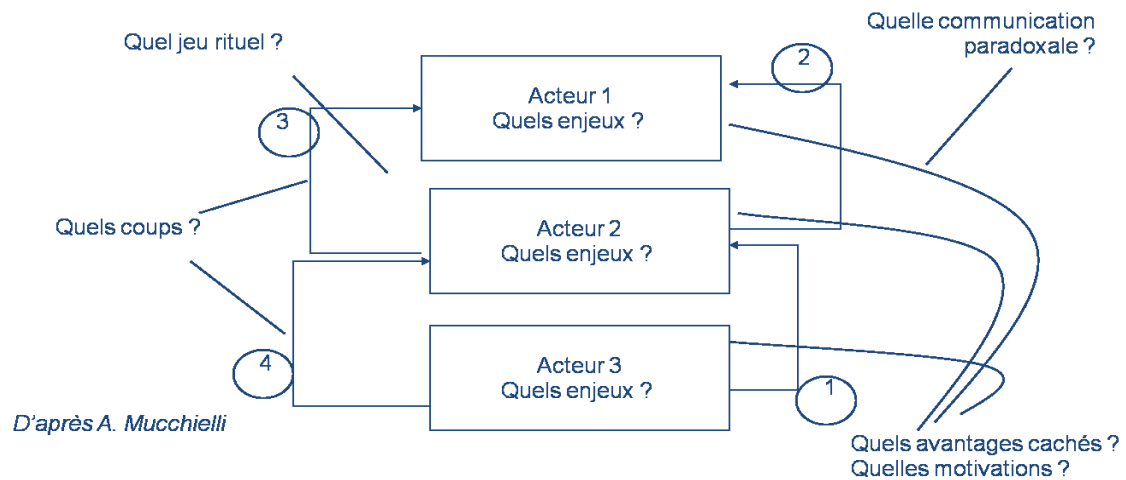


Figure n°1.3.4. Modèle interactionniste d'après A. Mucchielli (2000)

En 1980, un autre modèle de type interactionniste est publié par C. Kerbrat-Orrechioni (1980) dans une conception plus linguistique (figure n°1.3.5). Fondé sur les limites du modèle de Jakobson, il s'inscrit dans le cadre théorique de l'énonciation. Ainsi, tout échange s'inscrit dans un cadre spatio-temporel, avec deux locuteurs ayant chacun des compétences relatives et différentes. La langue n'est plus considérée comme un système figé. Il est soumis aux locuteurs (acte de parole) qui en ont nécessairement une maîtrise limitée et imparfaite.

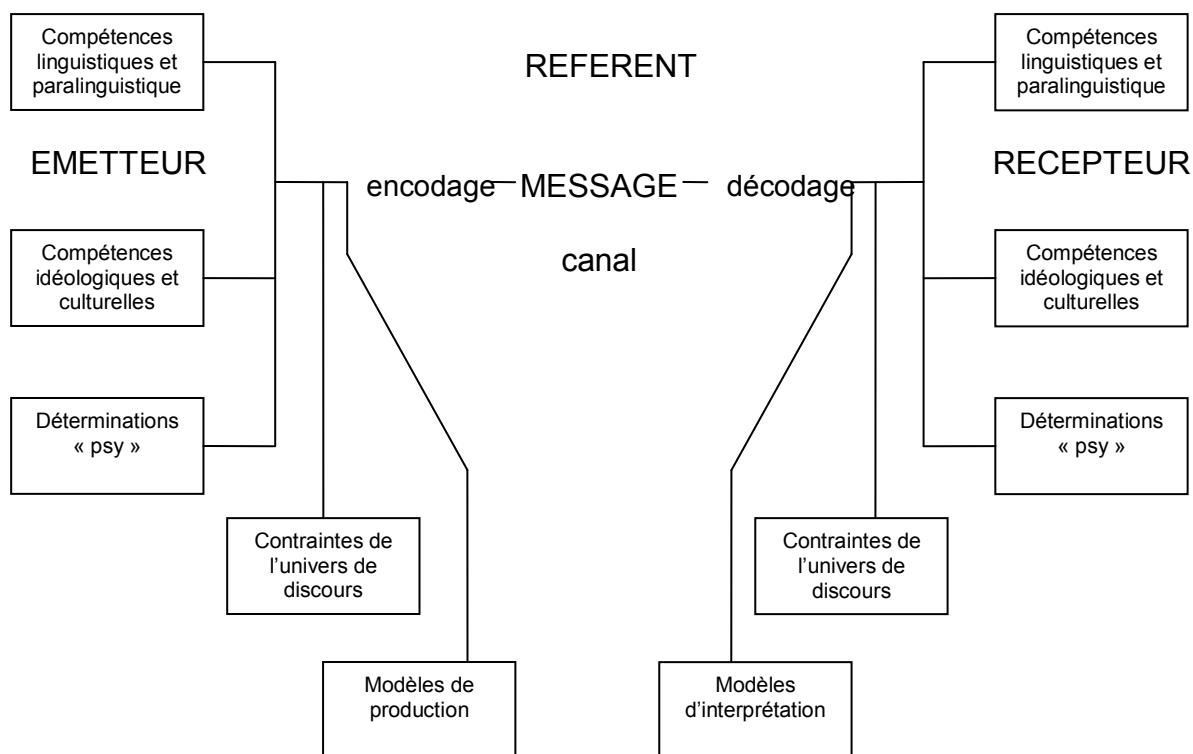


Figure n°1.3.5 : Modèle de l'énonciation d'après C. Kerbrat-Orechioni (1980)

Enfin, les modèles constructivistes sont les derniers apparus avec proposé par A. Mucchielli lui-même, le modèle situationnel (Figure n°1.3.6) qui envisage la communication en terme de « processus ». Il fait apparaître différents contextes (ou dimensions de toute situation) qui fonctionnent nécessairement pour toutes communications. L'interprétation du phénomène observé a des racines dans les processus de contextualisations différents. Les situations de référence étant plurielles, les significations de l'échange le seront également et le « sens partagé » sera à découvrir.

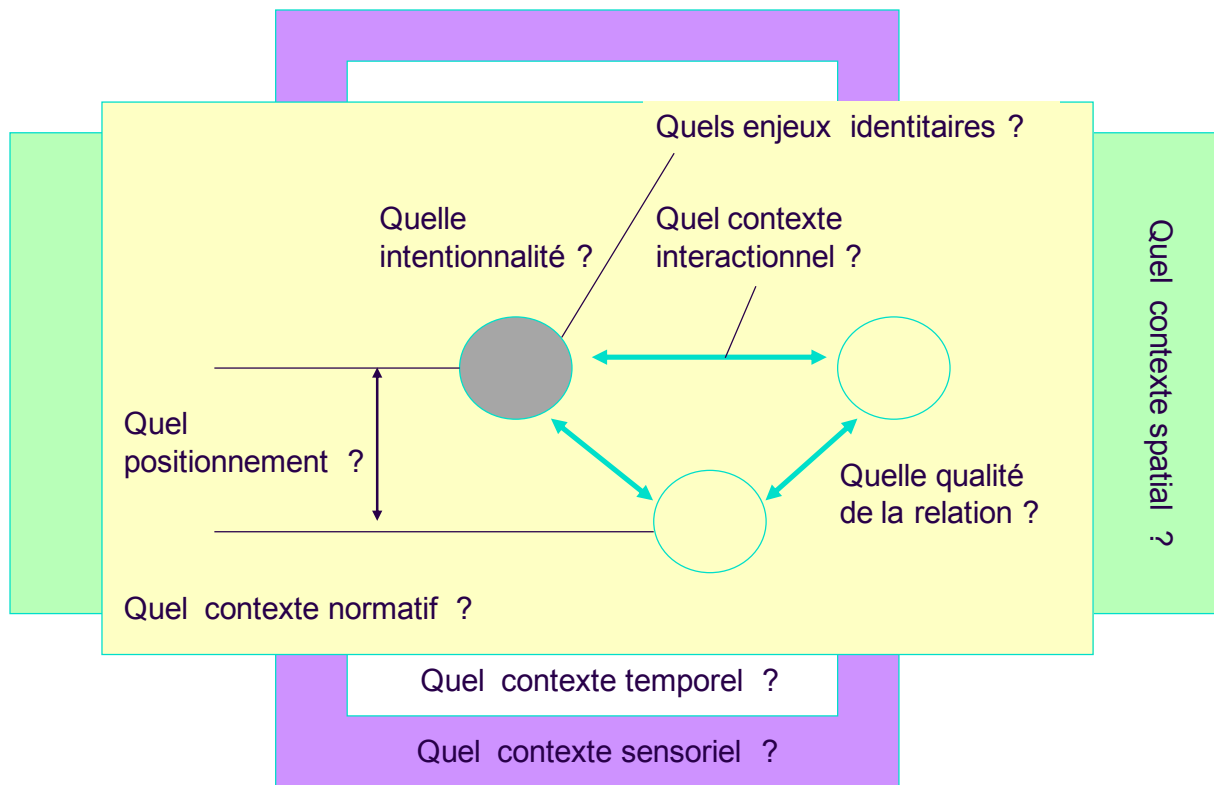


Figure n°1.3.6. Modèle situationnel d'après A. Mucchielli (2000)

Ces modèles sont liés à des préoccupations historiques et sociales différentes et donc à des enjeux de communication qui ont varié eux aussi au cours du temps. Ils ont ainsi évolué du positivisme (l'enjeu est l'analyse des effets de la communication et de son efficacité), à la systémique (l'analyse intéresse la permanence et le changement des systèmes de communication) et au constructivisme (la construction du sens partagé par les acteurs).

Ces différents schémas de la communication font échos aux différentes étapes de la prise en charge des TPPP. Le modèle de Shannon a permis d'établir la base de l'évaluation de l'intelligibilité de la parole au niveau du code acoustico-phonétique. Nous souhaitons présenter ensuite un aperçu des théories sous-tendant l'accès au sens pour ouvrir les perspectives offertes par ses différents modèles de communication.

---

### 2.3.2. DE LA PERCEPTION DE LA PAROLE A LA COMPREHENSION DU DISCOURS EN CONVERSATION

---

La recherche des différentes théories et modèles intéressant ce champ de connaissance est rendue difficile par son étendue relevant de nombreuses disciplines, chacune ayant une histoire, des

méthodes et des objectifs différents. Cette complexité n'a d'égale que la simplicité avec laquelle nous réalisons une conversation !

La figure n°1.3.7 propose un schéma de l'ensemble du processus en adoptant les approches psycholinguistiques (Caron 2001) et pragmatiques (Garic & Calas 2007, Moeschler 2009).

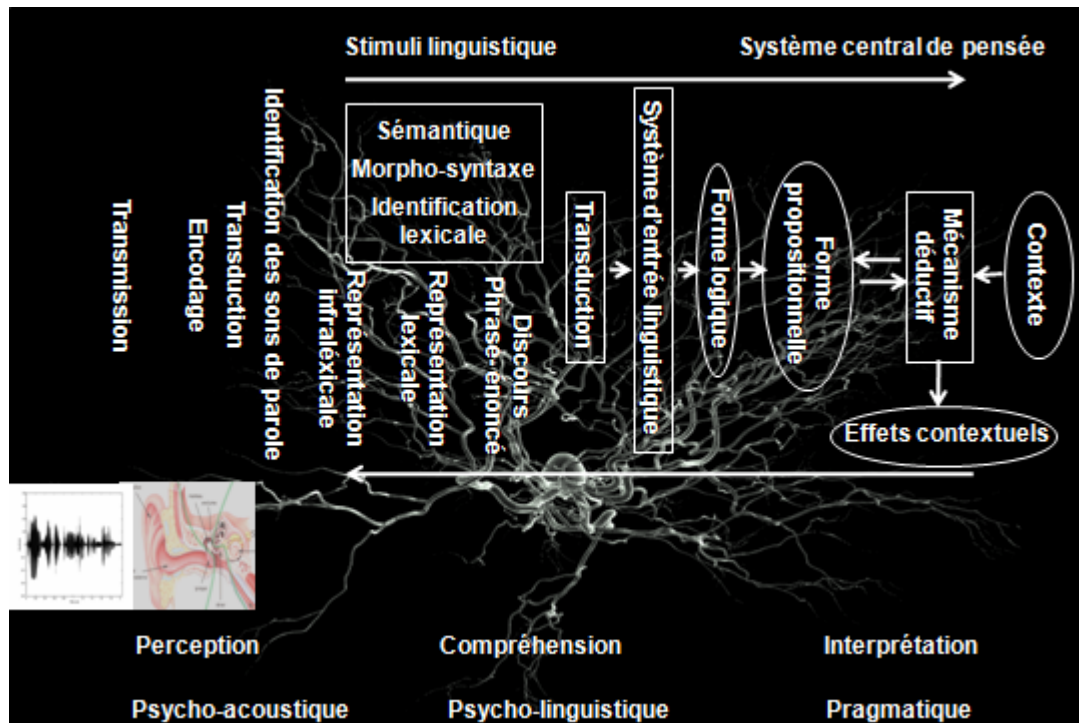


Figure n°1.3.7. Schéma des différentes étapes de la perception du son à l'interprétation des énoncés.

Dans la partie inférieure de ce schéma figure les trois disciplines les plus impliquées : la psychoacoustique, la psycholinguistique et la pragmatique. Elles utilisent en raison des processus qu'elles décrivent les termes de perception, compréhension et interprétation avec des chevauchements issus de leurs interactions. Le terme d'intelligibilité est plutôt réservé à l'étape de reconnaissance acoustico-phonétique du phénomène et donc à la reconnaissance des représentations des unités infra-lexicales ou lexicales.

Dans la partie supérieure du schéma figurent les différentes étapes du processus en présentant l'accès au sens linguistique, comme le stimuli du système central de pensée tel que le conçoit la pragmatique dans la théorie de la pertinence (Reboul & Moeschler 2008), théorie s'appuyant sur l'approche cognitive de Fodor (1986).

Si nous considérons un auditeur « idéal » (même si c'est exceptionnellement la réalité), voulant discuter avec une personne présentant un TPPP, il est légitime de faire l'hypothèse que la perception du son dégradé sera complète et que ses capacités cognitives sont intactes. C'est pour identifier les mots et obtenir une représentation mentale correcte des unités sous jacentes que la dégradation du signal aura une répercussion sur l'ensemble du processus. A partir de là, tout le système devra s'adapter à l'altération de ce stimuli. La question « comment maintenir une communication efficace en intégrant la dégradation ? » impose une réflexion sur la manière dont cela peut impacter les différents niveaux du processus. Pour cette raison, après un aperçu des différentes théories, nous ferons le point sur les connaissances actuelles des processus de l'accès au sens, à la compréhension du discours en interaction.

#### 2.3.2.1. LA PERCEPTION DE LA PAROLE

---

« Exposé à un flux de parole continu, l'auditeur doit être capable d'établir une relation entre deux univers hétérogènes que sont l'univers physique des sons du langage et l'univers symbolique des énoncés (mots, phrase, discours). La question de savoir comment ces unités linguistiques abstraites peuvent être recouvrées en mémoire à partir de l'information acoustique disponible dans le signal reste, à l'heure actuelle, sans réponse consensuelle ». Cette introduction de l'article « Alignement lexical et segmentation de la parole » (Bagou & Frauenfelder 2002) témoigne d'une des questions fondamentale restant en suspend dans le domaine de la perception de la parole.

La perception est l'opération complexe par laquelle l'esprit, en organisant les données sensorielles, se forme une représentation des objets extérieurs et prend connaissance du réel. La perception de la parole constitue la première étape psycholinguistique de traitement du langage oral. Pour appréhender la question, les chercheurs ont décomposé le processus en différentes étapes telle qu'il est représenté sur le schéma proposé figure n°1.3.8. Les multiples flèches signalent toutes les hypothèses explorées. La flèche en pointillé symbolise le processus top-down.

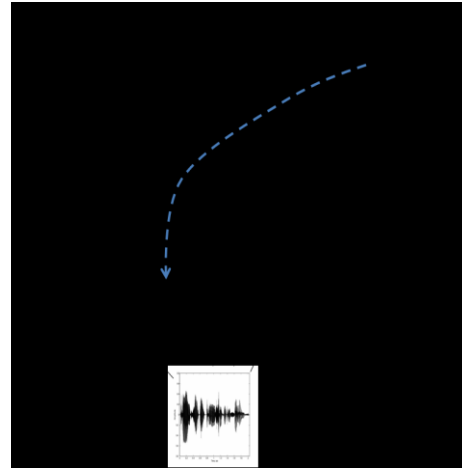


Figure n°1.3.8 : Du signal à l'identification d'un mot.

Le processus ascendant dit « bottom-up » correspond à la description du traitement commençant par les unités de bas niveau, puis successivement les unités de plus haut niveau. L'existence d'un processus descendant dit « top-down » sous-entendant que les unités de bas niveau peuvent être « déduites » après l'accès au niveau supérieur est maintenant admise (Ohala 1986).

Pour un aperçu sur « La perception de la parole », le chapitre de Noel Nguyen dans l'ouvrage intitulé « Phonétique et phonologie : forme et substance » (Nguyen 2005) apporte une analyse actualisée, à la fois critique et synthétique complétant la présentation faite dans les ouvrages cités jusqu'à présent.

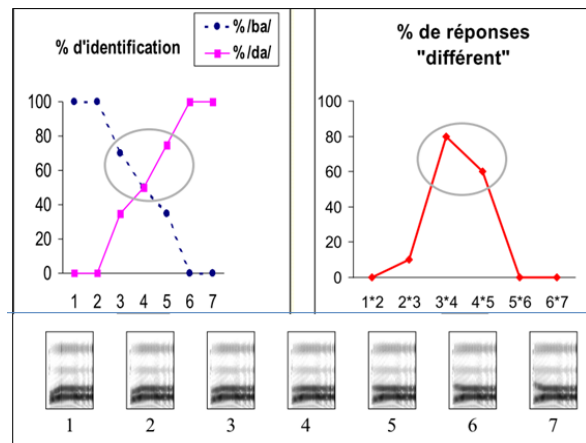
#### 2.3.2.1.1. LA PERCEPTION DES PHONEMES EST-ELLE CATEGORIELLE ?

Précisons tout d'abord que cette question nécessite des expérimentations basées sur des tâches d'identification de phonèmes alors que ce processus n'est peut-être pas à une étape nécessaire pour la reconnaissance d'un mot. Bien qu'elle ait été un des arguments contribuant à l'élaboration de la théorie de la perception motrice de la parole (Libermann 1967, 1985), différentes études ont ouvert d'autres perspectives sur lesquelles nous reviendrons après avoir présenté les conclusions des principaux travaux.

Les études fondamentales ont été menées dans les années 50 dans le but d'identifier les indices acoustiques servant de support aux oppositions de phonèmes (Liberman et al 1952). Elles ont abouti à mettre en évidence qu'il est, dans certaines circonstances, plus facile à l'auditeur de percevoir des différences entre deux sons lorsque ces derniers se rangent dans deux catégories phonémiques différentes, plutôt que dans une même catégorie. Ainsi, un continuum de stimuli (ex /ba/ vers /da/ est perçu de manière « discontinue », catégorielle : soit /ba/ soit /da/, sans intermédiaire (figure n°1.3.9).



Figure n°1.3.9 : Réponses type dans un test d'identification (à gauche) et dans un test de discrimination à droite sur un continuum /ba/ vers /da/ (A noter la variation de direction dans l'orientation de la transition du second formant)



Mais contrairement à ces résultats établis sur les occlusives, il a été montré que la perception des voyelles n'est pas catégorielle et ce même si ces voyelles sont associées à une consonne de même catégorie phonémique (Fry et al 1962). Puis des investigations réalisées sur de nouvelles classes de consonnes (fricatives, glides, liquides) ont révélé que la perception de ces stimuli n'était pas aussi catégorielle que celle des consonnes occlusives (Fujisaki & Kawashima 1971). Ces mêmes auteurs ont également montré que même les voyelles pouvaient être perçues sur un mode catégoriel dans des conditions défavorables et notamment lorsqu'elles étaient de courte durée. Ces résultats ont soulevé la question de savoir dans quelle mesure la perception catégorielle était conditionnée par la nature de la réponse demandée au sujet. Elle était peut être induite par un choix effectué parmi un petit nombre de réponses prédéfinies. De nouvelles expériences ont apporté une confirmation à cette idée en conduisant les chercheurs à établir une distinction entre perception catégorielle et réponse catégorielle. Ainsi selon Massaro (1974, 1983), l'auditeur peut dans certaines circonstances répondre sur un mode catégoriel alors que les informations qu'il extrait conservent dans tous les cas un caractère graduel et continu.

Ainsi, les résultats de Fujisaka et Kawashima (1971) entraînent en contradiction avec la théorie motrice de la perception de la parole où l'auditeur interprète le signal de parole en référence aux mouvements articulatoires dans l'espace des consonnes ce qui expliquerait que les consonnes soient perçues sur un mode catégoriel alors que les voyelles le seraient sur un mode continu. L'interprétation des résultats de Fujisaka et Kawashima (1971) a fait émerger un modèle qui attribue un rôle central à la mémoire auditive à court terme (Pisoni 1975). Pour Massaro (1983), il serait désavantageux pour l'auditeur de traiter les sons de la parole de manière catégorielle en passant par le filtre d'un ensemble d'étiquettes phonétiques. En faisant abstraction des informations acoustiques détaillées sur le signal

de parole, l'auditeur perdrait la possibilité d'exploiter ces informations ultérieurement, pour corriger de possibles erreurs d'indentification relatives au phonème ou mot précédent par exemple. La stratégie la plus efficace consisterait au contraire à conserver autant d'information que possible sur le signal le plus longtemps possible.

La théorie motrice de la perception de la parole a été cependant renforcée ces dernières années par les travaux sur l'effet de l'adaptation perceptive à l'apprentissage moteur (Nasir & Ostry 2009) et l'effet de la recalibration perceptive (Miller J 1994, Shiller DM et al 2009). Une synthèse de ces contributions est présentée par Houde (2009). Les résultats sont concordants avec l'idée que la production de la parole crée des associations entre les représentations motrices, somato-sensorielles et auditives. Ainsi dans le cadre théorique d'un possible couplage fonctionnel entre systèmes de perception et de production de la parole, les variations de catégorisations en fonction du contexte sont justifiées par la variabilité de la production de la parole. Cette grande variabilité de la production est en partie due aux contraintes de la production et notamment au fait que le changement de position des articulateurs prend du temps et qu'il atteint rarement sa cible dans la parole continue. L'hypothèse que l'auditeur utilise alors sa propre expérience passée dans cette production contrainte quand il perçoit la parole des autres, peut alors être présentée comme un phénomène naturel.

Ainsi, au-delà du débat de la nature catégorielle de la perception, Nguyen (2005) suggère qu'il « n'est pas exclu de penser que l'auditeur puisse procéder à un double codage de l'information ». La perception catégorielle ne serait pas un artéfact expérimental. Elle serait à mettre en relation avec ce processus de codage symbolique et discret où l'information pourrait être codée de manière continue sous la forme de spectrogramme auditif par exemple et discrète sous la forme d'un ensemble d'étiquettes phonétiques.

Cette vision soulève la question de la capacité de la mémoire à court terme à mémoriser des signaux dégradés de parole mais aussi de la manière dont les patients présentant un TPPP perçoivent la parole non dégradée.

#### 2.3.2.1.2. EXISTE T-IL OU NON UNE UNITE MINIMALE DE PERCEPTION ?

Pour poser le problème de l'existence d'une unité infra lexicale ou d'une unité minimale de perception, nous citerons encore une fois Bagou & Frauenfelder (2002) (Figure 1.3.10) :

« Quoique l'existence d'un niveau de représentation intermédiaire infra-lexical continue d'alimenter les débats, nous admettons, [...], que le signal physique est préalablement analysé par le système auditif, puis transformé en une ou plusieurs représentation(s) dites « infra-lexicales » qui constituent une interface entre la réalité acoustique continue, non séquentielle et variable du signal, et les unités discrètes du lexique. Cependant, compte tenu des caractéristiques du signal de parole, cette conversion n'est pas triviale. La psycholinguistique s'est beaucoup interrogée sur la nature des unités infra-lexicales. Dans la plupart des modèles actuels du traitement de la parole, les représentations sont décomposables en une séquence d'unités élémentaires, le plus souvent définies en termes linguistiques : le diphone (Klatt 1980) ; la semi-syllabe (Fujimura 1976) et la syllabe (Massaro 1974 ; 1987 ; Mehler et al, 1981). L'approche la plus commune suggère que la syllabe constitue l'unité perceptive de base (Segui et al, 1990). Ainsi, la chaîne de parole serait segmentée puis classifiée afin de faire émerger une représentation intermédiaire syllabique, alors utilisée pour accéder au lexique mental. Cependant, malgré un grand nombre de travaux confirmant cette hypothèse, certains résultats montrent que la syllabe n'est pas une unité de perception dans certaines langues romanes, en l'occurrence en français (Frauenfelder & Kearns, 1996 ; Kolinsky, 1998 ; Pallier, 1994). Il apparaît donc qu'elle ne constitue pas l'unité de classification permettant l'accès au lexique dans toutes les langues du monde (Tabossi et al, 2000). La question de savoir quel rôle peut être attribué à cette unité en français reste donc en suspens. »

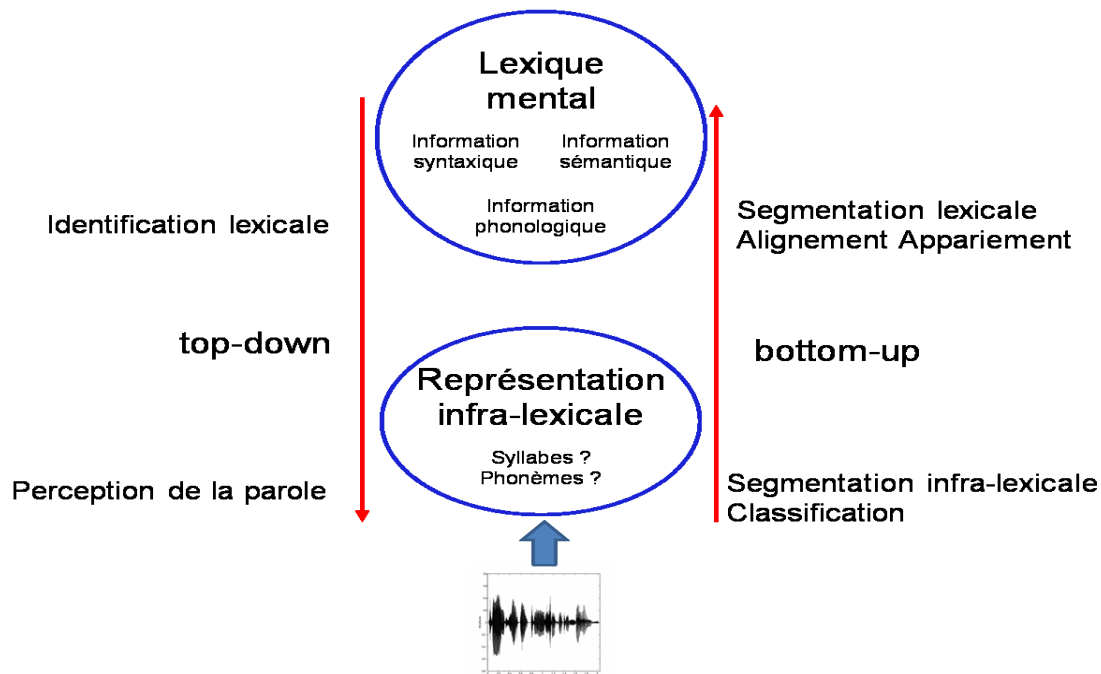


Figure n°1.3.10 : Du signal d'entrée aux représentations lexicales d'après Bagou et al 2002

Ainsi, bien que Savin et Bever (1970) aient montré qu'une syllabe entière était détectée plus rapidement que la consonne initiale de cette syllabe en anglais, d'autres études ont suggéré que cet effet syllabique était peut être spécifique à certaines langues (Segui 2000) dont le français (Cutler et al 1986). De plus, une série d'expériences récentes (Content & Frauenfelder 2002) n'a pu retrouver cet effet syllabique. Leurs résultats donnent cependant à penser que certains constituants syllabiques, et notamment ce que les auteurs désignent sous le terme d'attaque, ont un rôle important dans la perception de la parole en servant de point d'ancrage pour accéder au lexique (Dumay et al 2002).

D'autres chercheurs soutiennent une hypothèse concurrente où l'unité primaire de traitement dans la parole serait le trait c'est-à-dire qu'il y aurait une correspondance directe entre une représentation en trait du signal et le lexique, sans passer au niveau phonémique ou syllabique. Ceci est le cas dans le modèle Cohort révisé en 1994 (Marslen-Wilson 1994), le modèle LAFF (Lexical access from features) de Stevens (Stevens 2002) et dans le modèle FUL (Featuraly underspecified lexcon) de Lahari & Reetz (2002, 2010).

Enfin, certains auteurs ont récemment suggéré que les auditeurs pouvaient être simultanément sensibles à des unités de taille différente dans le signal de parole. Dans la théorie de la résonance adaptative proposée par Grossberg (2003) les unités de petite et grande taille sont simultanément activées avec un biais naturel en des unités plus larges. Les discussions sur le sujet ont conduit

plusieurs auteurs dont Hawkins (2003) à envisager une compétition parmi les différentes unités dont le domaine temporel est conditionné par des facteurs phonologiques, lexicaux et grammaticaux, mais aussi par la dynamique de l'interaction conversationnelle et par les contraintes de la situation expérimentale dans laquelle l'auditeur est placé.

Ainsi, bien que les travaux se poursuivent pour étudier comment l'auditeur identifie des phonèmes lorsque la situation expérimentale l'amène à le faire, ils ne se placent pas en position de déterminer si les unités phonémiques jouent un véritable rôle dans la compréhension du langage oral par l'auditeur. Il en est ainsi de la théorie des aimants perceptifs de Kuhl (1991) qui à partir de la perspective prototypique de la catégorisation des phonèmes (Rosch E. 1976), décrit un espace perceptif contracté autour du prototype. Cette théorie permet d'interpréter les frontières catégorielles comme le simple fait que les sons situés à mi chemin entre deux prototypes échappent à l'attraction de ses prototypes. Cette observation permet d'introduire la contribution potentielle du concept d'auto-organisation.

L'auto-organisation caractérise les systèmes dont les propriétés macroscopiques résultent de celles de leurs constituants tout en étant qualitativement différentes. On parle d'auto-organisation quand ces structures globales organisées apparaissent dans des systèmes qui ne possèdent au départ, ni plan, ni modèle réduit pour les faire émerger. Ce concept issu des travaux du biomathématicien D'Arcy Thompson (1917, édition revisitée 2000) sous-tend à un certain degré les théories sur l'origine de la parole. Les travaux de Oudeyer (2005) qui en utilisant un système artificiel, conceptualisent la manière dont une société d'agents (entités intelligentes en intelligence artificielle), dotés de conduits vocaux et d'oreilles reliés par des réseaux neuronaux peut former par auto organisation un code de la parole discret (figure n°1.3.11) combinatorial et partagé par tous les agents, sans que l'on présume de capacité linguistique ou de capacité de coordination sociale. Ces résultats apportent un éclairage complémentaire à la théorie de la dispersion de Lindblom (1986) sur le fait que les langues humaines s'orientent préférentiellement vers les voyelles les plus distinctes les une des autres pour des raisons fonctionnelles et à la théorie dispersion-focalisation des systèmes vocaliques de Schwartz (1997), qui à la contrainte de dispersion dans l'espace vocalique, ajoute une focalisation dans l'espace des voyelles qui ont deux formants adjacents proches l'un de l'autre..

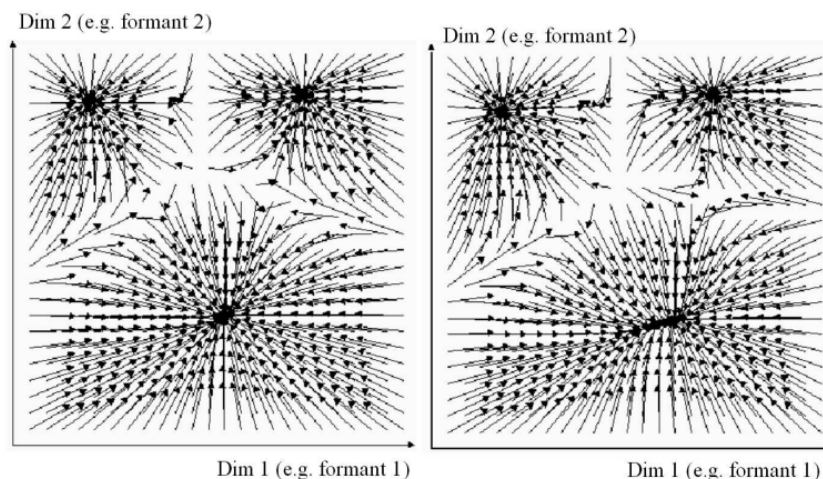


Figure n° 1.3.9. Représentation du paysage des attracteurs associés à deux cartes neurales perceptuelles. On observe la présence de trois bassins d'attractions, correspondant à trois catégories phonémiques.

Ainsi, le débat sur l'unité minimal et les modalités de perception de ces unités pourrait s'enrichir, en adoptant d'un point de vue cognitiviste, un questionnement sur le rôle du potentiel de l'auto-organisation du système dans la part « automatique » du processus par rapport à sa composante « contrôlée ».

#### 2.3.2.1.3. L'IDENTIFICATION DE MOTS ISOLÉS OU COMMENT S'EFFECTUE L'ACCÈS AU LEXIQUE MENTAL ?

Comment cette masse d'informations acoustiques séquentielles qui varie de manière non triviale d'un locuteur à l'autre et de discours en discours converge vers la représentation mentale d'un seul mot ?

Pour répondre à cette question, plusieurs modèles ont été décrits répondant globalement à deux hypothèses. Nguyen (2005) oppose une approche « abstractionniste » et une approche « exemplariste ». « Dans l'approche abstractionniste l'auditeur se représente les mots sous une forme minimaliste réduite à un petit nombre de traits distinctifs (figure n°1.3.12). La perception de la parole s'apparente à un processus de réduction de l'information et elle conduit à convertir le signal de parole en une séquence de représentations symboliques dont le degré d'abstraction augmente à chaque étape du traitement. Les modèles à exemplaires attribuent à l'auditeur la capacité de stocker en mémoire les différentes formes de surface associées à un mot. Lorsque l'auditeur doit reconnaître un mot, les propriétés de ce mot sont comparées à chaque exemplaire et l'exemplaire est activé proportionnellement à son degré de similarité avec le mot d'entrée. Par opposition à l'approche

abstractionniste, une relation naturelle d’analogie vient s’établir entre le signal d’entrée et les exemplaires stockés dans le lexique ».

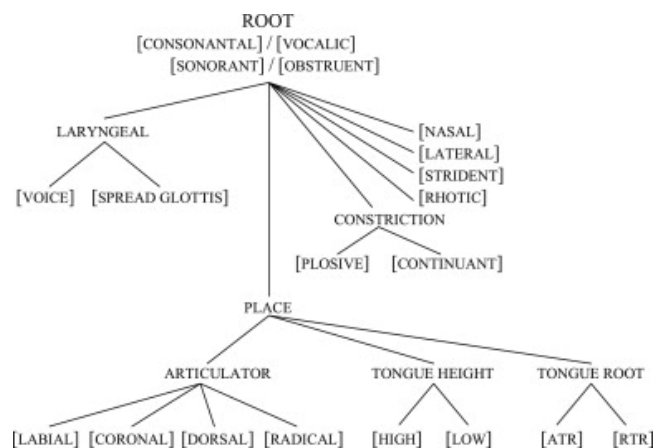


Figure n°1.3.10 : Organisation des traits dans le modèle FUL 'd'après Hawkins (2010) (ATR : palato alvéolaire, RTR rétroflexe)

Le modèle FUL (Featuraly underspecified lexcon) de Lahari (2002, 2010) est un des plus représentatif des modèles abstractionnistes. Chaque morphème est associé à une représentation phonologique unique, abstraite et sous-spécifiée indépendamment des variations dont le morphème peut être le siège en surface. Le caractère sous-spécifié des représentations phonologiques lexicales conduit à supposer que la relation entre le signal de parole et le lexique peut revêtir 3 formes : correspondance, non correspondance et absence de non correspondance. Ce dispositif permet d'expliquer qu'un mot soit correctement reconnu par l'auditeur malgré les déformations dont il est parfois le siège, notamment en vertu des phénomènes d'assimilations. Pour Nguyen (2005) « Ce type de modèle peut être qualifié d'abstractionniste dans la mesure où il fait intervenir une procédure de normalisation, mise en application de manière automatique à une étape précoce du traitement, permettant à l'auditeur de faire abstraction de la variabilité inter-individuelle. Ils s'inscrivent dans une perspective dualiste postulant l'existence de deux niveaux de représentation : une représentation de surface, physique et variable, et une représentation phonologique sous jacente, abstraite et symbolique».

Liss JM (2007) dans son approche perceptive des TPPP adopte l'hypothèse « exemplariste ». Elle décrit les deux processus fondamentaux sous tendant la reconnaissance des mots : l'activation lexicale et la compétition lexicale, en faisant la synthèse d'une série de modèles listés ci-dessous :

- le modèle logogène de Morton (1969),
- le modèle cohort (Marslen-Wilson & Welsh 1978)

- le modèle TRACE (McClelland & Elman 1986),
- le modèle shortlist (Norris 1994,2008),
- le modèle Neighborhood activation (Luce & Pisoni 1998) ,
- le modèle PARSYN (Luce et al 2000).

Quand un auditeur rencontre un mot parlé, certaines caractéristiques du signal acoustique servent à activer une série de candidats lexicaux possibles. Ceci est suivi par une compétition entre ces candidats pour s'ajuster au mieux à l'entrée, le gagnant de la compétition étant le mot reconnu.

L'activation lexicale commence dès que le premier phonème du mot est produit, de telle manière que nous n'avons pas à attendre la fin du mot pour poursuivre la reconnaissance du mot. Les débuts de mots jouent ici un rôle important dans l'activation des possibilités lexicales de la cohorte. Des segments de mots survenant après peuvent intervenir pour discriminer des mots divergeant tardivement. La forme globale du mot cible participe également à l'activation des items du lexique.

Avec l'activation de la cohorte lexicale, commence la compétition entre les différents candidats pour sélectionner le mot qui concorde le mieux avec l'entrée. Cette compétition peut être caractérisée par la facilitation ou l'inhibition. Ainsi les candidats lexicaux les plus proches sont facilités et les plus éloignés inhibés. La compétition est plus qu'un simple processus de concordance entre la composition de l'entrée et la composition du candidat, d'autres facteurs de la macrostructure du lexique affectant le processus. La densité lexicale de voisinage renvoie au groupe de mots voisins similaires à la cible (Luce & Pisoni 1998, Luce et al 2000) :

- cela prend plus de temps et d'effort pour distinguer la cible des voisins quand la forme de la cible est relativement unique.
- à l'intérieur des mots de voisinage, certains sont plus fréquents et communs que d'autres. Les auditeurs prennent en compte cet effet «fréquence de mots » durant la phase de compétition facilitant la reconnaissance des mots les plus communs, les autres facteurs étant égaux par ailleurs (Luce 2000).

De plus, à un niveau de densité de voisinage égal, la compétition est affectée par l'information sous lexicale comme la structure acoustique. Nous citerons :



- les indices acoustiques attendus pour la coarticulation entre les segments adjacents
- la probabilité de l'occurrence des séquences de phonèmes ou probabilité phonotactique.

Ainsi, les items du lexique activés sont des candidats potentiels en compétition. Celle-ci est guidée par la similarité segmentale avec l'entrée et influencée par d'autres facteurs lexicaux et sous-lexicaux qui renforcent ou réduisent la vraisemblance du candidat à sélectionner ou conserver.

Un point critique de ce processus est le déroulement temporel. Parce que ce processus immédiat se déploie « on line », de manière sériale dans le cours du temps, la reconnaissance peut parfois survenir avant que le mot entier soit prononcé. C'est possible quand un mot contient un point d'unicité (point qui avant la fin du mot ne laisse qu'un seul candidat possible dans la compétition). Cependant peu de mots sont uniques dans le lexique avant leur fin, ainsi dans la plupart des cas les auditeurs sont obligés d'attendre la fin du mot ou le mot suivant dans la parole continue avant de décider définitivement de l'identité du mot. Sinon il risque de sélectionner le mauvais candidat. Cette stratégie de « concomitant retardé » décrite par McQueen et al (1995) impose que l'information soit stockée dès le début du mot et gardée en mémoire à court terme jusqu'à la fin de la production du mot. Tous les aspects du mot sont alors considérés simultanément pour prendre la décision. Ceci est utile pour éviter une décision prématurée et inadaptée qui risque de survenir particulièrement si le signal est lent ou s'il est dégradé.

Toutes ces théories d'identification des mots traduisent la place fondamentale attribuée au lexique interne dans le processus de compréhension. Cependant, l'accès au mot isolé est un cas de figure expérimental dans le contexte de la communication.

#### 2.3.2.1.4. LA PERCEPTION DE LA PAROLE CONTINUE OU COMMENT SEGMENTONS-NOUS LE SIGNAL DE PAROLE ?

De passer par le mot comme unité de perception dans la parole continue nous permet de tolérer les ambiguïtés et distorsions segmentales. Le phénomène perceptuel le plus connu à cet égard est celui de la restauration phonémique où les illusions auditives (Warren 1970) intéressent le signal acoustique. Il y a ainsi des arguments pour retenir que les auditeurs s'appuient sur les mots plus que sur les phonèmes dans la parole continue, et des arguments pour que les mots à l'intérieur de la phrase exercent une influence les uns sur les autres.

Mais comment décomposons-nous la phrase en mots ? La segmentation lexicale est considérée comme aussi fondamentale pour comprendre la parole continue que l'activation et la compétition lexicale, ces processus étant intimement liés dans cette tâche. Le débat reste ouvert, mais une possibilité est que l'activation et la compétition lexicale dépendent de la détermination des frontières de mots. Quand une frontière de mot est déterminée, une cohorte possible de candidats est activée et la compétition entre les candidats peut se produire.

Deux approches tendent à expliquer la segmentation lexicale :

Pour certains la segmentation est le produit de la reconnaissance des mots et le processus est guidé par le lexique (Mc Clelland Elman 1986, Norris 94 et 2001). Ainsi, dès que le mot est reconnu dans la phrase, il permet de segmenter les autres mots l'entourant dans la phrase. Ce processus est grandement facilité par la connaissance de la structure de la langue (possibilités sémantiques et règles syntaxiques) et cela aide à activer les cohortes lexicales. Les auditeurs exploitent les régularités de la langue pour anticiper et activer les séries les plus probables de candidats lexicaux. La perception est facilitée par les amorçages sémantiques et syntaxiques.

Ainsi, la segmentation guidée par le lexique fait référence à l'activation des formes de mots qui correspondent à diverses portions du signal que l'information contextuelle rend hautement prédictibles. La reconnaissance d'un mot dans la phrase facilite la reconnaissance et la segmentation des autres mots de cette phrase.

D'autres ont argumenté que la segmentation est guidée par un processus infralexical. Les auditeurs « balayent » le courant acoustique pour détecter les indices pertinents de fin et de début de mot. Ces indices infralexicaux comprennent l'information prosodique et accentuelle. Une longue liste d'indices est susceptible d'être utilisée par un auditeur pour assurer la segmentation, variant en fonction des langues, comprenant :

- des indices phonotactiques, allophoniques et autres indices acoustico phonétiques,
- les silence et les indices métriques basés sur l'entrée rythmique de la structure de la langue.

Les procédures de segmentation semblent fortement contraintes par l'organisation rythmique des langues (Ségui 2000) : rythme syllabique pour le français, l'espagnol le portugais et la catalan (Ségui

et al 1990), rythme basé sur la mora en japonais (McQueen et al 2001), l'occurrence des syllabes fortes en anglais (Cutler & Norris 1988) et la probabilité phonotactique (McQueen & Cutler 2001).

Savoir si certains de ces indices ont plus de poids que d'autres dans la segmentation sera une étape importante dans la connaissance de la perception de la parole.

#### 2.3.2.1.5. COMMENT PERÇOIT-ON LA PAROLE DEGRADÉE ?

Les sources d'information concernant la perception de la parole dégradée dans le cadre des TPPP sont très pauvres.

De nombreuses études ont porté sur la perception de la parole altérée dans le bruit ou bien ayant subi des manipulations expérimentales spectrales, fréquentielles ou temporelles. Ceci nous a appris que malgré ces dégradations et grâce à la redondance des indices acoustiques du signal de parole et du contexte (Grosjean 1985, Bard et al 1988, Duez D 2001), l'auditeur est capable de récupérer des informations manquantes (voir section 1.5).

Liss (2007) dans un chapitre intitulé « le rôle de la perception dans les troubles moteurs de la parole » présente une réflexion synthétique sur comment la nature de la parole dégradée des TPPP peut intervenir sur les processus de reconnaissance des mots.

Cet auteur a pour cela sélectionné les différents processus potentiellement stratégiques pour la perception d'une parole dégradée (tableau n°1.3.1). Elle liste ainsi :

- la redondance par les relations entre les différents indices acoustiques,
- le phénomène de normalisation,
- l'importance de la mémoire de travail,
- la connaissance des régularités linguistiques avec le phénomène d'amorçage
- les phénomènes de restauration induisant un biais lexical.

Processus / Stratégies	Description
« Trading relations »  Libermann & al 1967, Matthys & al 2005, Repp 1982, Weismer & Martin 1992	Le signal de parole est intrinsèquement très redondant, produisant de multiples indices acoustiques pour chaque trait de parole.  Les auditeurs ont une aptitude très développée pour prendre des décisions perceptives à partir de n'importe quelle combinaison d'information acoustique saillante, fiable et relativement reproductible. Cela signifie que la présence exhaustive des indices acoustiques n'est pas nécessaire à la juste perception d'un phonème.
Normalisation  Carré & Divenyi 2000, Miller & Liberman 1979, Ladefoged & Broadbent 1957	Il est remarquable que nous comprenions une phrase qu'elle soit prononcé rapidement ou lentement, par un homme, une femme ou un enfant, ou encore avec un fort accent étranger.  Ceci est possible car les auditeurs ont la capacité de réaliser un « processus de normalisation » du signal acoustique pour s'adapter aux sources de variations de la production. Ainsi, la manifestation acoustique d'un mot ou d'une phrase peut varier et être toujours comprise car l'auditeur a la capacité de le « changer » en un « patron » mental existant
Mémoire épisodique pour la parole  Bradlow & al 1999, Goldinger 1996, Goldinger & al 1999	En sus du processus de normalisation, existe des arguments pour un stockage par les auditeurs de traces de voix spécifiques dans la mémoire épisodique. Ces souvenirs détaillés aident les auditeurs à percevoir la parole produite par la personne dans les interactions suivantes. Il est possible que ce type d'informations soit particulièrement utile quand la parole a quelque chose d'inhabituel ou est dégradée.
Connaissance des règles linguistiques et des régularités  Borsky & al 2000, Dupoux & al 2001 Samuel & Pitt 2003 Samuel 1997	Les langues et la manière dont elles sont produites sont contraintes par des règles et des régularités. Les auditeurs utilisent leur connaissance de ces règles pour percevoir la parole. Cela comprend la connaissance de l'ordre habituelle et réglementaire des sons dans un mot (phonotactique), de quel son peut être assimilé, combiné, ou éliminé dans un processus phonologique, des contraintes sémantiques, syntaxiques, de la structure prosodique et autres. Cette connaissance nous permet de nous adapter à des réductions significatives et des distorsions du signal de parole.
Biais lexical"  Warren 1970, Samuel 1981, Ganong 1980	Les auditeurs ont tendance à percevoir les mots dans leur entièreté, même quand des phonèmes qui les composent sont absents ou altérés pour devenir ambigus dans leur identité. Cela signifie que les auditeurs peuvent tolérer d'importantes et significatives variations segmentales dans le signal acoustique.

Tableau n° 1.3.1 : Sélection de processus cognitivo-perceptifs et stratégies en rapport avec un signal dégradé d'après Liss JM 2007

Ensuite sont présentées les conséquences possibles sur le processus de reconnaissance des mots en tenant compte du fait que les erreurs segmentales produites par les locuteurs atteints d'un TPPP sont principalement des distorsions, mais également des omissions et des substitutions (tableau n°1.3.2).

Traits segmentaux		
	Omissions	L'absence de phonème peut entraîner la non activation du mot cible
	Distorsions	La distorsion d'un phonème peut entraîner une incertitude ou une ambiguïté. La cohorte activée peut être plus grande que nécessaire pour autoriser toutes les interprétations des phonèmes ambigus (Wurm et al 2004). Cela peut provoquer une dépense très excessive de ressources pendant la phase de compétition et augmente le risque de choisir une mauvaise cible comme mot perçu (Andruski et al 1994).
	Substitutions	Une substitution peut conduire à la non activation du mot cible.
Traits supra-segmentaux		
	Durée inappropriée des syllabes	Les auditeurs peuvent avoir des difficultés à appliquer une stratégie de segmentation métrique en raison de difficultés à discerner les syllabes fortes des syllabes faibles. (Liss et al 2000, 2002)
	Des variations d'intensité intersyllabiques inappropriées	Les auditeurs peuvent avoir des difficultés à discerner les syllabes fortes des syllabes faibles.
	Des variations de fréquence fondamentales intersyllabiques inappropriées	Un aplatissement de la fréquence fondamentale (monotone) est associé avec une réduction de l'intelligibilité (Bradlow et al 1996, Laures & Bunton 2003)
	Réduction de l'espace vocalique (en particulier la réduction des voyelles fortes)	La réduction de l'espace vocalique est associée à une réduction de l'intelligibilité dans les mots simples et les phrases (bunton & weismer 2001, Weismer et al 2001)

Tableau n°1.3.2. Traits acoustiques de la parole dysarthrique qui peut affecter l'activation la compétition ou la segmentation lexicale.

Peuvent être rajouter :

- que la longueur du message (Yorkston & Beukelman 1981) est également reconnue comme altérant l'intelligibilité, interférant avec les capacités de la mémoire de travail si elle est associée à un ralentissement du débit articulatoire
- qu'un indicage alphabétique soit à l'aide d'un tableau, soit gestuel, améliore de 10 à 15% la compréhension des productions d'un locuteur infirme moteur cérébral (Hustadt 2005). Cet indicage a en plus une effet sur le rythme et la vitesse de production de la parole contribuant au résultat en plus de l'incrémentation des informations apportées par l'amorçage.

L'impact sur l'intelligibilité de la parole est ici analysé en référence à l'identification d'un mot, reconnu comme l'unité de la compréhension de la phrase, de l'énoncé ou du discours.

### 2.3.2.2. LA COMPRÉHENSION DU DISCOURS ORAL

---

Hustadt (2008) dans un article sur l'intelligibilité et la compréhensibilité dans les TPPP fait référence au cadre théorique de la compréhension des énoncés défini par Kintsch (1983).

La recherche d'informations complémentaires ouvre la porte à un questionnement partagé par de nombreuses disciplines avec des préoccupations et des applications potentielles très différentes.

Nous ne rentrerons pas ici dans la description des débats entre la syntaxe et la sémantique autour du sens de la phrase. Mais, si les sciences du langage, les sciences computationnelles, les sciences cognitives et la philosophie abordent la question de la compréhension du discours, c'est finalement la pragmatique qui s'attache à la description des processus de compréhension liés à l'usage du langage dans la communication et dans l'interaction. Sa récente différenciation au sein de la linguistique traduit bien une problématique complexe nécessitant la convergence des connaissances de toutes ces disciplines. Cette recherche est fortement stimulée par les besoins dans le domaine du traitement automatique des langues qu'elle que soit sa modalité, orale ou écrite.

#### 2.3.2.2.1. LA COMPREHENSION DE TEXTE ET LE MODELE DE KINTSCH ET VAN DIJK (1978, 1983)

Mais revenons d'abord aux travaux de Kintsch qui renvoient à la psycholinguistique cognitive de la compréhension de texte (Marin & Legros 1992) et à la construction d'une représentation mentale qui dépasse l'information fournie dans le texte.

Plusieurs générations de modèles caractérisent les recherches conduites sur ce sujet depuis les années 1970. Le fait que la compréhension du texte explicatif soit beaucoup plus difficile que celle des textes narratifs, le traitement des inférences étant beaucoup plus complexe si le lecteur n'est pas familier avec le monde évoqué par le texte (Graesser et al, 2003), a été un moteur dans cette évolution.

Kintsh et Van Dijk ont proposé en 1978 un modèle où plusieurs processus complexes opèrent en parallèle et de manière interactive. Ce premier modèle a ensuite été complété dans la mesure où il ne prenait pas en compte ni la langue, ni les variables culturelles identifiant les spécificités du lecteur

qui favorisent l'activité d'inférence. Pour dépasser ces limites Van Dijk et Kintsch ont introduit la notion de « modèle de situation » dans le modèle de 1983 (tableau n°1.3.3)

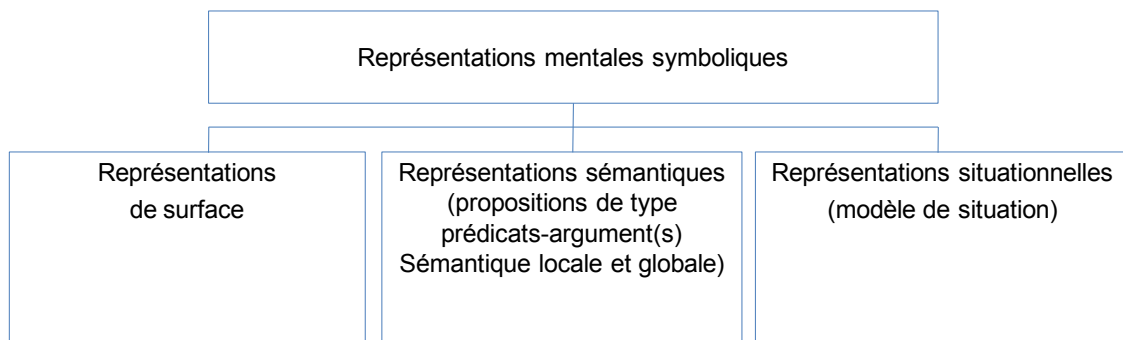


Tableau n°1.3.3 : Les trois niveaux de représentation mentale d'après Van Dijk et Kintsch 1983

Dans ce modèle, 3 niveaux de représentation du discours sont impliqués dans le processus de compréhension incluant le code de surface, le texte de base et le modèle de situation.

Le premier niveau (représentations de surface) est le plus « élémentaire » pour les psycholinguistes (Singer 2000), faisant référence à l'identification des mots à la syntaxe ou à la morphologie du message original (code). Il est donc, transposé à la situation de la compréhension orale, totalement dépendant de la perception de la parole.

Le niveau intermédiaire de représentation (texte de base ou proposition conditionnelle), renvoie aux propositions qui sont extraits du code de surface : essentiellement la sémantique du message. Il est organisé en deux niveaux: le niveau local et le niveau global, désignés par la microstructure et la macrostructure.

La microstructure d'un texte est constituée d'un réseau de propositions hiérarchisées, correspondant à ce qui est directement exprimé en surface. Elle résulte du traitement de l'information du texte, phrase par phrase. Elle se compose de prédicats et d'arguments. Les prédicats sont des termes relationnels comme les verbes, les adjectifs ou les adverbes, les arguments sont représentés par les noms. C'est ainsi qu'une phrase simple peut être représentée par une proposition complexe qui consiste en un prédicat associé à de nombreux arguments, des circonstances de temps, de lieu et des éléments modificateurs.

La macrostructure correspond au traitement global du texte réalisé en fonction des buts du lecteur. Elle représente le noyau de la signification, tel qu'il apparaît dans un résumé. Elle est composée d'une série de propositions hiérarchiquement organisées qui représentent la structure globale du texte. La présence de titre, de sous-titre et le thème de la première phrase de chaque paragraphe, sont des indices textuels utilisés dans la construction de la macrostructure. Un grand nombre de facteurs intervient dans cette construction, comme les connaissances générales du lecteur et le contexte.

Le plus haut niveau de représentation, le modèle de situation, correspond à ce que l'individu a appris du texte, détaché de la structure textuelle et intégré à ses connaissances. Il intègre les propositions dans la connaissance du monde et les buts du lecteur. Il s'agit d'une représentation cognitive des événements, des actions, des individus et de la situation en général qui peut incorporer des expériences antérieures et des particularisations de connaissances portant sur ces expériences. Ces travaux permettent de prendre en compte au delà des connaissances partagées, les connaissances spécifiques et le système des « connaissances/croyances » des individus construits dans leur culture.

La construction d'un « modèle de situation » ne se limite pas au domaine de la compréhension de textes. Il peut être élaboré à partir de support issus d'autres sources d'information.

Dans le prolongement de cette théorie, Kintsch 1988, 1998, propose le modèle de « construction-intégration » qui est un modèle à visée explicative. Cette conception décrit comment les textes sont représentés en mémoire durant le processus de compréhension et comment ils sont intégrés aux connaissances du lecteur. L'auteur simplifie son approche en s'intéressant plus précisément aux processus responsables de la sélection des informations congruentes et pertinentes pour construire la cohérence de la signification (Tapiero 2007).

Le modèle de Kintsch est devenu le modèle princeps de la compréhension de texte. Il n'exclut pas d'autres théories, notamment celle de Johnson-Laird qui a proposé également en 1983 une théorie des modèles mentaux comme base de connaissance et univers référentiels des expressions langagières. Comprendre un discours n'est pas progressivement construire un réseau de propositions issue du traitement de chaque phrase, mais d'élaborer d'un « modèle mental » qui est progressivement remanié et enrichi. La mise à jour du « modèle mental » passe par l'acquisition de connaissances nouvelles et l'actualisation de connaissances anciennes, c'est-à-dire leur spécification, ou la substitution de connaissances supposées exactes à des connaissances supposées erronées. La



compréhension de texte consiste en une séquence d'activation de modèles mentaux préexistants en mémoire. (Marin & Legros 1992).

La compréhension automatique des langues sous la forme écrite ou orale est devenue un enjeu majeur du fait de l'essor de la communication amorcé à la fin du siècle dernier par l'ouverture du « net » et l'avènement de la téléphonie mobile. Elle explique la richesse des modèles issus des théories de Chomsky (1965,1991), les applications potentielles de la compréhension étant un puissant « moteur » pour les recherches dans ce domaine.

#### 2.3.2.2.2. DE LA COMPREHENSION D'UN DISCOURS ORAL AU DISCOURS EN INTERACTION

La pragmatique contemporaine (Garric N & Calas F 2009) a insisté sur le caractère sémantiquement sous-déterminé de la communication linguistique. En d'autres termes, ce que le locuteur entend communiquer est rarement en rapport avec la signification de la phrase qu'il utilise lorsque celle-ci est interprétée de façon purement linguistique.

Dans la théorie des " actes de langage " de John Roger Searle (1972), les énoncés que nous préférons ne sont pas seulement porteurs d'informations. Ils expriment des demandes, des ordres et appellent des réactions comportementales. A ce titre, ce sont des " actes ", autant que des mots. En même temps qu'il a un sens, le langage a une "force illocutionnaire", à laquelle, parfois, se réduit toute sa signification. À travers les conventions extra-linguistiques qui gouvernent l'usage des expressions dans des contextes donnés, on peut distinguer les relations entre ce qui est dit, ce que cela veut dire, ce que veut dire celui qui parle, etc. Utiliser le langage et parler, c'est donc s'engager.

Proposer que la communication linguistique se fasse sur la base d'un processus d'encodage-décodage, accompagné par un processus inférentiel prenant pour prémisses un ensemble de connaissances partagées par les interlocuteurs a été la première manière d'aborder ce problème. Cependant cette solution n'est pas apparue réaliste car les connaissances ne sont pas effectivement partagées de manière universelle, ce qui a donné lieu au développement de théories au sein de la pragmatique.

Sperber et Wilson (1989), sur la base de la sous détermination sémantique de l'interprétation des énoncés supposent un processus double, linguistique et codique d'une part, pragmatique et inférentiel

de l'autre. Cette division linguistique des tâches est explicitée dans les termes de l'analyse modulaire hiérarchique que propose Fodor (1983) du fonctionnement de l'esprit humain et donne lieu à la théorie de la pertinence mettant l'accent sur l'intention informative (rendre manifeste ou plus manifeste à l'interlocuteur un ensemble d'hypothèse) et l'intention communicative (rendre manifeste ou plus manifeste à l'interlocuteur et au locuteur que le locuteur a cette intention informative). Selon Sperber et Wilson, la pertinence est affaire de rendement entre coût de traitement et effets cognitifs produits par l'énoncé, ces deux facteurs étant articulés de la façon suivante :

Définition de la pertinence : plus un stimulus a d'effets cognitifs, plus il est pertinent, moins un stimulus demande d'efforts d'interprétation, plus il est pertinent.

Garantie de pertinence optimale : L'énoncé est suffisamment pertinent pour mériter l'effort que consacre l'interlocuteur à le traiter.

La théorie de la pertinence s'adresse aux énoncés pris individuellement. Mais Moeschler (2006) qui défend cette théorie, émet le point de vue que le jugement de cohérence que l'on porte sur un discours dépend directement de la pertinence globale de l'ensemble du discours, à savoir la facilité avec laquelle l'interlocuteur récupère une intention informative globale et la richesse de cette intention informative, dépendant bien évidemment des effets cognitifs de ce discours.

Dans le cadre de la théorie de la pertinence, les processus pragmatiques se réduisent à deux grands types : l'explicitation d'une part, et l'implication d'autre part.

L'explicitation est un processus d'enrichissement de la forme logique fournie par le module linguistique. L'implication est un processus d'inférence forte si l'interlocuteur peut légitimement l'attribuer à l'intention communicative du locuteur, faible si elle est du seul ressort de la responsabilité de l'interlocuteur.

« Ainsi cette théorie est basée sur l'idée que l'interlocuteur ne fera pas d'efforts cognitifs inutiles, qui ne soient contrebalancés par des effets contextuels. La stratégie n'est par ailleurs pas une reconstruction ex post facto d'un raisonnement, elle est au contraire supposée être justifiée à la fois sur le plan cognitif et sur le plan communicatif. Le défi pour le système quelque soit la finalité d'application reste bien évidemment de déceler l'intention informative du locuteur et d'y apporter la réponse la plus pertinente possible » Moeschler (2006).

Vulgarisée et abordée du point de vue de l'enseignement, voici une petite « vignette » extraite de l'article de Gremmo et al 1990 :

*Un comportement de compréhension, c'est QUELQU'UN qui écoute.*

L'auditeur n'est pas uniquement défini par son rôle dans l'interaction. C'est un individu psychologiquement, socialement et culturellement déterminé. Ses caractéristiques au moment où il se comporte en auditeur (certaines sont permanentes, d'autres momentanées) vont influencer sur sa participation à la communication.

*Un comportement de compréhension, c'est quelqu'un qui écoute QUELQUE CHOSE dans une SITUATION DONNÉE.*

Le message se présente toujours sous la forme d'un type de discours particulier : émission de radio, dialogue avec un interlocuteur en face à face, conférence, pièce de théâtre, etc.

*Un comportement de compréhension, c'est quelqu'un qui écoute quelque chose dans une situation donnée POUR UNE BONNE RAISON.*

On n'écoute pas pour écouter mais pour atteindre un objectif, pour satisfaire un besoin extra-communicatif (augmenter ses connaissances, guider son action, se faire plaisir...).

Mais l'interaction constitue l'environnement naturel de la langue. Kerbrat-Orecchioni C (2005) précise dans son ouvrage sur le discours en interaction, que « Comprendre comment les énoncés sont compris », lorsqu'on travaille sur du discours en interaction, se reformule ainsi : « comprendre comment les locuteurs se comprennent mutuellement ». C'est ce qui change fondamentalement l'approche du processus.

Le processus interprétatif reste toujours un véritable « travail » qui dans le cas des contenus implicites s'apparente à un « calcul ». Le sens doit être extrait de son enveloppe signifiante selon un processus complexe et tâtonnant. Il s'agit toujours pour les participants à la conversation de construire à propos d'un segment donné une hypothèse interprétative, qui va en principe leur servir de base à l'enchaînement.

Il est nécessaire de distinguer trois grands niveaux constitués d'unités que l'on peut dire respectivement « pratiques », « fonctionnelles » et « relationnelles » :

1/ Niveau de gestion locale de l'alternance des tours de parole (la « machinerie de la conversation ») ;

2/ Niveau de la cohérence syntactico-sémantico-pragmatique du dialogue et de son organisation aussi bien « micro » que « macro ». De ce niveau relèvent les unités pragmatiques que sont les actes de langage, les interventions et les échanges, mais aussi les unités syntaxiques que sont les phrases ou « clauses » pour l'oral ;

3/ Niveau de la relation interpersonnelle (gestion des divers types de « relationèmes », contraintes rituelles, face-work).

Voici deux exemples de l'analyse de clauses

*il fait bon chez vous !*

1/ illocutoirement : une assertion évaluative positive

2/ conversationnellement : un type particulier « d'ouvreur »

3/ relationnellement un « amadoueur », c'est-à-dire un procédé permettant de s'attirer les faveurs de son interlocuteur

*je peux vous demander quelque chose ?*

1/ une question,

2/ un énoncé préliminaire,

3/ un adoucisseur d'incursion

Nous n'avons pas abordé ici les données théoriques du rôle dans la compréhension en conversation et en interaction des facteurs extralinguistiques mais nous pouvons conseiller la lecture de Kröger et al (2009) qui propose un modèle de production, perception et acquisition des actions en communication face à face à partir de travaux incluant les mimiques faciales et les gestes des membres supérieurs.

#### 2.3.2.2.3. LA COMPREHENSION DE LA PAROLE DEGRADEE

Les discussions sur les notions d'intelligibilité et de compréhension dans la littérature traitant des TPPP, justifient la recherche des théories et des modèles explicités ci-dessus. Les débats seront présentés dans la section suivante consacrée aux principes de l'évaluation de l'intelligibilité dans ce domaine d'application.

Mais dès à présent, nous pouvons recenser les connaissances sur la compréhension de la parole dégradée dans la mesure où les principales sources d'informations sont d'ordre « thérapeutique » et concernent le domaine de la remédiation de l'impact du TPPP sur la communication des personnes dans leur quotidien. Cette approche croise celle des besoins de la communication augmentée. Une autre source d'informations comprend les facteurs interprétés comme des « biais » de la mesure de

l'intelligibilité quand elle est réduite à son niveau acoustico-phonétique. Il s'agit en fait des effets top-down provenant des niveaux de traitement des représentations mentales.

Nous pouvons lister ainsi :

- la familiarité (Beukelman & Yorkston 1980) avec la parole dégradée
  - o à travers la voix (qui est interprétée comme un effet sur la perception quand il s'agit des effets de la fréquence de la voix mais qui joue aussi un rôle dans la segmentation et sur le traitement de l'information linguistique avant la constitution du « stimulus » linguistique
  - o mais également à travers les aspects « habitudes » élocutoires de la personne
  - o ceci peut être complété par le fait que dans le contexte de l'apprentissage d'une langue, Bradlow et al. (1996) montrent que les femmes sont plus intelligibles, en tant que groupe, lorsqu'elles parlent une langue étrangère.
  - o l'expérience de l'interlocuteur (Tjaden & Liss 1995) est à rapprocher de la familiarité. Elle a conduit Liss (2007) à établir un programme de prise en charge pour les interlocuteurs des personnes souffrant d'un TPPP.
  - o La familiarité sous entendant un lien de parenté a également été étudiée (Flipsen 1995).
- la prédictibilité du message (Garcia & Cannito 1996) à travers les différents indices contextuels permet d'améliorer l'intelligibilité avec des indices visuels (Hunter L, Pring T, Martin S. 1991), visuels gestuels (Garcia & Cannito 1996) visuels faciaux (Hustad & Cahill 2003)
- La longueur du message (Yorkston & Beukelman 1981) peut être déterminante si elle dépasse les capacités de maintien en mémoire des différentes représentations nécessaires à la construction du sens.

Hustad (2007) et Hustad et al (2007) à travers ses travaux pour l'amélioration de la compréhensibilité des TPPP (particulièrement des personnes souffrant d'une infirmité motrice cérébrale, IMC) a colligé de nombreuses informations. La poursuite de ces travaux sur l'indigage alphabétique montre que l'utilisation d'une source audio visuelle améliore la compréhension de 20% (10 à 15% par canal auditif seul, 20 à 40% par un canal audio-visuel). La figure n°1.3.4 illustre l'importance de la multi modalité dans la perception de la parole.

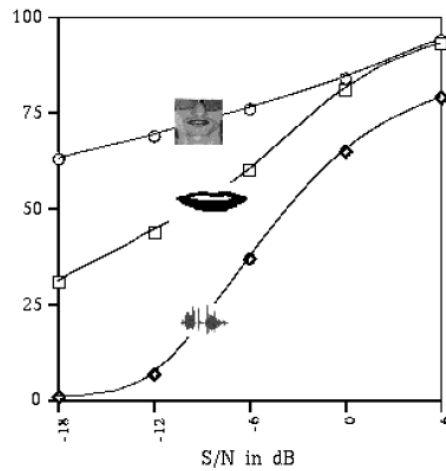


Figure n°1.3.4 : la multi modalité dans la perception de la parole d'après Benoit (1994)

En 2007, Hustad elle montre que l'intelligibilité est meilleure pour les mots « fermés » que pour les mots « ouverts » sur le plan sémantique.

Sur les constats que 1/ le jugement de l'intelligibilité et de la compréhensibilité peut intégrer des indices visuels, alphabétiques et sémantiques (Blanchet & Snyder 2010), 2/ la vision du handicap liée aux TPPP doit dépasser la perception isolée et intégrer le contexte de communication et de participation, Bloch & Wilkinson (2004, 2009) ont travaillé sur l'analyse de conversation dans les TPPP.

Les principes clés de l'analyse de conversation sont :

- L'importance des échanges naturels et quotidiens (naturel en opposition à une tâche expérimentale) ;
- L'ordre dans la conversation (sur l'hypothèse que l'ordre en conversation est créé à travers et maintenu par les participants eux-mêmes) ;
- Les tours de parole en interaction sont orientés contextuellement ;
- La construction au fur et à mesure de la compréhension.

L'observation de la parole en conversation en l'absence de TPPP montre que la parole et le langage produits sont très rarement parfaits et qu'il y a « naturellement » des troubles de la production de la parole, de la perception (de l'audition) et de la compréhension. Ils se traduisent par des faux départs, des répétitions de mots, des chevauchements, des demandes de clarification...Le terme de « réparation » décrit les pratiques utilisées par les participants pour éclairer et gérer les problèmes au

cours de l'interaction. Cette réparation peut être faite par celui qui l'a produite ou par « l'autre ». Schegloff et al 1977 ont montré une prépondérance des autoréparations dans la conversation. Ils font le constat que dans les troubles de la parole, l'attention est focalisée sur la théorie de l'erreur par rapport au trouble de la production et non sur la manière avec laquelle ce trouble est géré Schegloff (2007). De nombreux facteurs peuvent affecter la compréhension d'un tour de parole, entendu comme la manière dont le tour est compris par un participant en relation immédiatement avec le tour précédent (Bloch & Wilkinson 2004, 2009). Dans les observations faites par Bloch & Wilkinson 2009, un manque d'essai d'autocorrection par le locuteur non dysarthrique même si une opportunité se présente par une « suspension - initiation de réparation » apparaît comme banal. Comprendre et valider la compréhension devient une activité explicite. Les traits identifiés sont le management des troubles et les adaptations aux changements avec une importance particulière de la défaillance des essais ou des efforts d'auto réparation dans les TPPP. Ainsi, l'analyse de conversation dans le domaine des TPPP a démontré qu'il y a plus dans la parole distordue en interaction que la simple intelligibilité.

A ce stade de la rédaction, les données recueillies démontrent un important décalage entre la réflexion sur l'intelligibilité telle qu'elle est utilisée dans la pratique médicale et les notions de perception, compréhension, interprétation qui peuvent toutes renvoyer au terme d'intelligibilité. Nous allons dans la section suivante essayer de déterminer les définitions que nous pourrons ensuite appliquer au concept d'intelligibilité, objet de ce travail.

## **2.4. LES PRINCIPES DE L'EVALUATION DE L'INTELLIGIBILITE DE LA PAROLE**

---

### *2.4.1. LES GRANDS COURANTS*

---

Différents points de vue sont à l'origine de plusieurs conceptions de l'intelligibilité. Ils sont inhérents aux préoccupations et à la culture des différents acteurs impliqués dans la prise en charge des TPPP. Yorkston et al (1999), dans un chapitre intitulé « perspectives on dysarthria and apraxia of speech » listent un certain nombre de points de vue : celui du neurologue, du scientifique fondamental, du rééducateur, du locuteur. Cette équipe a défini l'intelligibilité (p 236) : "Intelligibility can be defined as the extent a listener understands the speech produced by individuals with motor speech disorders. It is

an overall measure reflecting not only the cumulative impact of the impairment X but also the compensatory strategies used by the speaker”. Associé à la mesure du débit de parole, cela constitue un index de sévérité de l'ensemble de la limitation fonctionnelle. Ceci explique la création par cette école en 1981 du « Assessment of Intelligibility in Dysarthric Speakers (AIDS, Yorkston & Beukelman, 1981) puis en 1996 du « Sentence intelligibility test » (Yorkston et al 1996b), qui associe à partir du même corpus de phrases, une mesure de l'intelligibilité, une mesure du débit de parole et une analyse des erreurs non seulement sur la production phonémique mais également sur les frontières des mots.

Ces choix n'étaient pas partagés par tous les spécialistes de la dysarthrie à l'époque de la publication du test, ajoutant une autre dimension aux différences de point de vue, celle de la finalité du test avec son parallèle : l'objet du test.

Ainsi toujours dans le même ouvrage, écrit en association avec KC Hustad mais cette fois ci dans le chapitre intitulé « Optimizing communicative effectiveness : bringing is together », l'amélioration de l'intelligibilité devient le but primordial de l'intervention thérapeutique et justifie l'introduction du terme de « comprehensibility » (figure n°1.4.1).

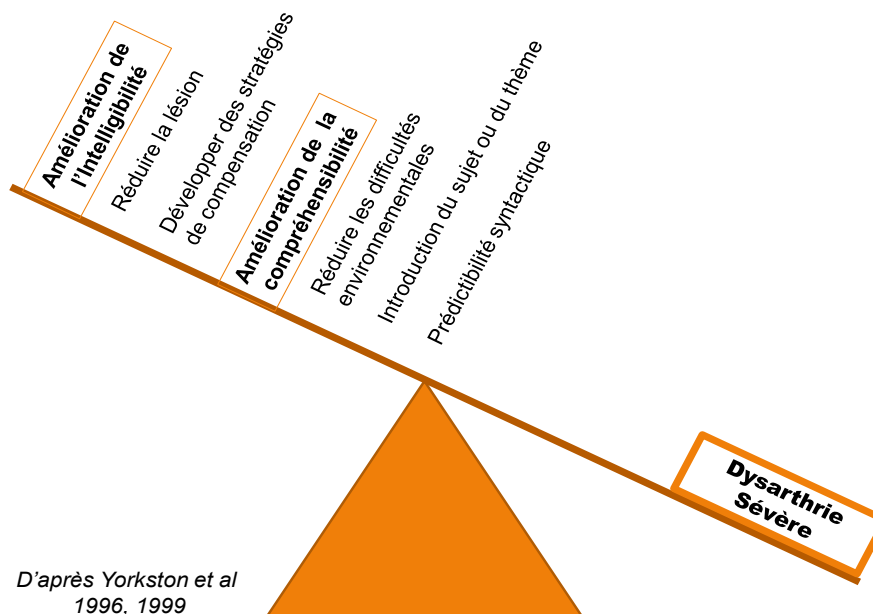


Figure n°1.4.1 : Représentation schématique des approches interventionnelles pour les individus présentant une dysarthrie sévère d'après Yorkston et al 1996.



Ce terme est alors employé en référence au processus dynamique par lequel les personnes partagent leurs pensées en utilisant toutes les informations disponibles (quelles qu'elles soient). Il avait déjà fait l'objet d'une publication en 1996 (Yorkston et al 1996 a) sa conceptualisation est schématisée sur la figure n°1.4.2.

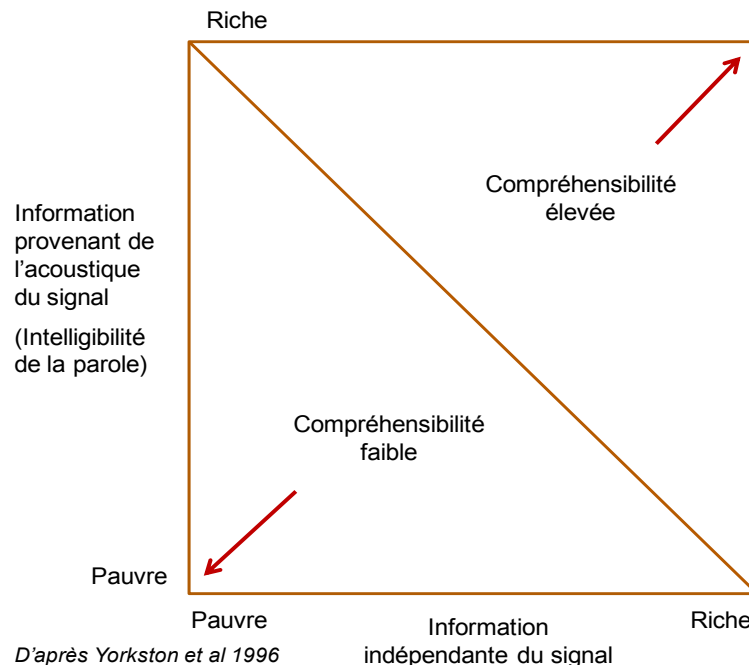


Figure n°1.4.2 : Contribution de l'information du signal acoustique et de l'information indépendante du signal à la compréhension du locuteur présentant une dysarthrie d'après Yorkston et al 1996.

Est décrite ainsi une stratégie d'intervention organisée en fonction du degré d'altération de l'intelligibilité compatible avec le profil évolutif des différentes étiologies. Dans cet ouvrage, dont le titre « management » des troubles moteur évoque la préoccupation thérapeutique sous toutes ces formes, la notion d'intelligibilité évolue du statut de mesure du degré d'altération du signal vers un indice de sévérité par sa mise en rapport avec la compréhension utilisée pour déterminer une conduite à tenir. Cependant aucune étude n'a permis d'établir une classification qui puisse servir de référence.

Un groupe de chercheurs faisant référence dans le domaine des TPPP a mené une réflexion conduisant à la publication du livre consacré à l'« intelligibility in speech disorders » en 1992 (Kent RD 1992). Dans son introduction (p1), Kent précise que la notion d'intelligibilité est beaucoup plus complexe que sa définition ne le suggère et que « although two persons may share some common idea of what intelligibility is, they may use very different methods to measure it and understand its

correlates in the act of speaking ». Un des grands axes de recherche de cette équipe a été le développement d'un test explicatif de l'intelligibilité. Cela signifiait (Kent et al 1989) un test qui ne se limite pas à un simple index de sévérité mais qui puisse donner lieu à une interprétation des erreurs phonétiques à l'origine de l'altération de l'intelligibilité. Il s'agit ici d'obtenir les connaissances sur les cibles à corriger pour réduire le trouble de l'intelligibilité cela situe l'intervention potentielle à gauche et haut du graphique de Yorkston). La création du « World Intelligibility Test » basé sur l'examen de 19 contrastes acoustico-phonétiques associe les propriétés suivantes : sensibilité aux troubles de la production dans les dysarthries et contribution de manière significative à l'intelligibilité de la parole. Les travaux se poursuivent avec la recherche de mesures acoustiques qui seraient hautement corrélées à l'intelligibilité et pourraient servir à prédire le degré d'altération de l'intelligibilité. Les dernières publications sont ainsi en faveur de l'utilisation de la mesure du F2 (Bunton & Weismer 2001, Kim et al 2009).

Pisoni et Luce (1986) semblent être les premiers à évoquer l'importance de la perception de la parole dans un contexte pathologique avec toutes les stratégies d'adaptation que l'auditeur peut utiliser pour pallier la détérioration de la qualité du signal produit. Il renvoie la problématique de la notion d'intelligibilité dans le cadre des TPPP à une diade auditeur-locuteur. Liss (2007) présente une argumentation fournie, analyse les principes de remédiations en rapport avec les difficultés de perception des TPPP et justifie l'éducation ou l'entraînement des interlocuteurs pour améliorer leur capacités à comprendre la personne dysarthrique.

Hustad & Beukelman (2002), constatant qu'il était rarement fait état de test mesurant la capacité des auditeurs à comprendre le message du locuteur présentant un TPPP (Beukelman & Yorkston 1979) propose d'évaluer celle-ci par des questions sur le contenu d'un message ou une narration en adaptant la méthode employée par Higginbotham et al (1994) dans une étude sur la compréhension d'un discours synthétique en communication alternative (résumer le contenu d'un passage narratif). Cette approche aboutit à la publication (Hustad 2008) d'un travail démontrant l'absence de corrélation entre un test d'intelligibilité par un taux de reconnaissance de mots et un test de compréhension de réponse à des questions dès lors que le degré de sévérité est un facteur contrôlé.

Ainsi, une autre dimension intervient dans l'évaluation : la sévérité du trouble, concept dont il faudra étudier le statut par rapport à celui de l'intelligibilité.

Ainsi à la fin de ce travail, nous essayerons de préciser dans le cadre des TPPP, ces différentes notions que sont la sévérité, l'intelligibilité et la compréhensibilité, l'interprétabilité. Nous allons tout d'abord nous recentrer sur la mesure de l'intelligibilité telle qu'elle nous vient des laboratoires Bell Telephon dans le cadre du modèle de Shannon (1948).

Le désavantage de l'ensemble des tests que nous venons de décrire dans la pratique clinique courante, est le temps et le personnel qu'ils mobilisent. Certains imposent une personne qui administre le test et une personne qui affecte le score, d'autres nécessiteraient de nombreuses personnes susceptibles de réaliser la passation. A noter, à ce sujet, l'exception du remarquable travail de Ziegler &, Zierdt (2008) qui en créant un test informatisé contrôlé et paramétré utilisable en télédiagnostique développe une conception pouvant résoudre ce problème. En attendant, de nombreux praticiens choisissent pour estimer la sévérité ou l'intelligibilité ou la compréhensibilité des échelles moins objectives telles que des échelles analogiques visuelles ou des échelles plus ou moins ordinales (Enderby 1983, Duffy 2005, Duffy 2007) dans des conditions variables en terme de support, de tâche... rendant difficile toute comparaison.

---

#### 2.4.2. LES DIFFERENTES MANIERES DE MESURER L'INTELLIGIBILITE.

---

Si dans le domaine de l'acoustique, on distingue généralement trois types d'évaluation de l'intelligibilité de la parole : les mesures subjectives (faisant appel à des locuteurs et des auditeurs), des mesures prédictives (se basant sur des paramètres physiques correspondant au canal de transmission, aux salles...) et des mesures objectives (effectuées à l'aide de signaux artificiels), le contexte médical des TPPP réduit le type de mesure du fait de « l'objet/sujet » lui-même : le signal spécifique au locuteur. Les méthodes subjectives sont donc la pratique courante et des recherches évaluent des options prédictives à partir d'indices acoustiques du signal du locuteur, des options objectives en utilisant des systèmes automatiques de reconnaissance pour « remplacer » les auditeurs par une référence « canonique » artificielle.

A noter que les mesures objectives ou prédictives n'ont pour objet que de retrouver des scores proches de ceux qui seraient obtenus par des mesures subjectives, tout en se passant de jury d'écoute. Elles restent donc le « gold standard » et les autres types de mesures n'ont donc pas de réalité différente.

#### 2.4.2.1. LES MESURES SUBJECTIVES

---

Les mesures subjectives se basent sur un protocole mettant en jeu un ou plusieurs locuteurs, et un ou plusieurs auditeurs. Les auditeurs sont soumis à différents stimuli, directement prononcés par les locuteurs préenregistrés. L'intelligibilité correspond alors au score obtenu par le jury (c'est à dire les auditeurs) dans les tâches qui suivent chaque stimulus, à savoir généralement la transcription de sons perçus ou la sélection d'unités – phonétiques ou lexicales - parmi plusieurs choix proposés.

Il existe de nombreux tests d'intelligibilité subjectifs et nous nous limiterons ici à retracer les tests en langue française dans la mesure où nous avons explicités les grands courants.

##### 2.4.2.1.1. LES TESTS PAR ESTIMATION DIRECTE D'AMPLITUDE

Ces tests reposent sur la capacité d'un ou de plusieurs locuteurs à extraire des informations d'un signal acoustique « déformé » par rapport à la normale. Cela reste une fonction proche de la détermination du degré d'intégrité sur signal acoustique, donne une information sur l'intégrité du signal acoustique par rapport à la « normale ». Ils sont souvent utilisés également pour décrire la sévérité de la dysarthrie et comme argument de suivi.

Dans *l'Evaluation Clinique de la Dysarthrie ou ECD* (Auzou et al., 1998). La mesure de l'intelligibilité utilise trois épreuves évaluant la parole lors de la lecture de mots, de phrases et au cours de la conversation. Le nombre de mots ou de phrases correctement interprétés sert à établir une note sur 8. L'évaluation de la parole au cours de la conversation est également notée sur 8. Ces trois épreuves aboutissent à un score d'intelligibilité noté sur 24 (Gaillard, 2001). Cette échelle donne un score d'intelligibilité sans prendre en compte le type d'erreurs phonétiques réalisées (Batterie d'évaluation clinique de la dysarthrie, Auzou & Rolland Monnoury 2000)

Le *Test Lillois de Communication ou TLD* (Rousseaux et al., 2001) comprend une évaluation de l'intelligibilité du patient pendant une conversation avec l'examineur (réponses à des questions, narration, description d'une image, lors de la lecture de mots, de phrases, et de textes). Le test comprend également une épreuve d'appréciation des éléments supra-segmentaux (prosodie, intonation) de la parole.

Weismer & Laures (2002) confirme l'influence d'un stimuli de référence permettant de calibrer le jugement des auditeurs avant la réalisation de ce type de tâche en obtenant des résultats différents à

partir de la présentation de 4 standards différents avant et au cours de la tâche, la consigne étant de mettre un score, l'intelligibilité étant définie comme la facilité avec laquelle les phrases peuvent être comprises.

Ce résultat conduit à proposer soit l'utilisation de cette technique d'évaluation sans calibrage par un stimulus de référence, soit la normalisation des tests par un ou des stimulus/i « étalon(s) » pour la recherche. Il observe également l'absence de concordance entre les résultats de l'estimation d'amplitude par cette consigne et le taux de reconnaissance de mots. L'explication serait la prise en compte holistique des paramètres déterminants l'intelligibilité.

Aucun test français ne propose en pratique de calibration préalable.

#### 2.4.2.1.2. LES TESTS « EXPLICATIFS »

Le but des travaux de Peckels et Rossi (1973) dans les années 70 a été d'adapter à la langue Française le « Diagnostic Rhyme Test » de WD Voiers, un outil apte à tester des vocodeurs ou de la parole synthétique mais qui pourrait aussi « apporter une aide précieuse dans d'autres domaines ». Ce test se présente sous la forme de liste mots où deux mots ne se distinguent que par la consonne initiale. L'auditeur est guidé dans son choix par le fait qu'il n'est pas demandé de transcrire mais de choisir dans un couple constituant une paire minimale. On entend par paire minimale deux mots qui ne se distinguent que par la consonne initiale et où en plus, ces deux consonnes représentent les deux membres d'une des oppositions. Le test se fait sur 6 traits caractéristiques (voisé, grave, compact, interrompu, vocalique, nasal) avec 2 oppositions dans chaque cas. Les auteurs ont fait remarquer que le choix forcé entre 2 possibilités ne semble pas une méthode adéquate pour tester l'intelligibilité de certaines oppositions notamment les oppositions de lieu d'articulation qui sont des oppositions à 3 termes. Ce test est peu utilisé dans le domaine des TPPP.

Gentil (1992) a travaillé sur l'adaptation française du Word Intelligibility Test selon la conception de Kent (1989) dans l'article « Toward phonetic intelligibility testing in dysarthria ». Elle a ainsi proposé des listes de mots pour un test de mots isolés à choix multiples présentés sous la forme de 4 colonnes avec dans une colonne la cible et dans les 3 autres un distracteur se différenciant de la cible par un contraste. (ex cible = bac, distracteurs = bec, bague, pack). Ce test explore 17 contrastes phonétiques. Est également proposé un test d'intelligibilité de paires de mots pour les troubles

sévères (une cible, un distracteur) explorant 13 traits chacun par 3 paires de mots. Dans tous les cas les contrastes peuvent être explorés par des mesures acoustiques. Ces tests sont à la base des descriptions des troubles phonétiques de la parole produite par des patients. La version « choix multiples de ce test a ensuite été adaptée par Crochemore et Vannier (2001) sous la direction de P. Auzou pour permettre une évaluation plus adaptée à la passation de bilan. Dénommé *Test Phonétique d'Intelligibilité ou TPI*, il a été élaboré dans le but de repérer les perturbations phonétiques qui conduisent à une mauvaise identification du phonème, et donc à une réduction de l'intelligibilité. Ce test d'identification à choix multiples est à la fois quantitatif et qualitatif : il mesure la sévérité de la dysarthrie par un score d'intelligibilité et identifie les troubles articulatoires sous-jacents. Il est présenté dans la batterie de *l'Evaluation Clinique de la Dysarthrie* (Auzou & Rolland Monnoury 2000)

#### 2.4.2.1.3. LES TESTS DE RECONNAISSANCES D'UNITES LINGUISTIQUES

Ces tests dont l'objectif est de calculer un taux de bonne reconnaissance sont très variables dans leur modalité en l'absence de recommandation et surtout d'essai de standardisation dans le domaine des TPPP.

Une revue critique de la littérature concernant les mesures de l'intelligibilité dans les TPPP (Dos Santos Barreto et al 2008) confirme l'absence de concordance entre les différentes méthodes de mesure de l'intelligibilité colligées dans la littérature accessible sur le web jusqu'en octobre 2007. Ces auteurs insistent sur les limites que cela induit dans l'interprétation des travaux cliniques ou de recherche. De plus apparaissent des interférences entre ces mesures et le type de tâche et de stimuli, les modalités de présentation du signal, le type de réponse demandé.

En français, contrairement à ce qui s'est produit en audiologie, aucune liste de mots ou de phrases n'a fait l'objet d'un consensus pour la pratique clinique. Les seules phrases équilibrées, validées et publiées demeurent les phrases de Combescure (1981). D'autres listes de mots équilibrées existent mais restent d'utilisation confidentielle (Dittner et al 2010, Zuliani 1988). Les listes de mots utilisées dans les batteries de test y compris la BECD ne sont pas équilibrées. De grandes variations sont également retrouvées concernant :

- des unités utilisées pour cette évaluation : du phonème à la phrase en passant par la syllabe, le mot...

- du type de tâche qui est destinée à être évaluée par les utilisateurs : identification, reconnaissance, compréhension
- des consignes pour la détermination du score surtout dans les tâches utilisant des mots ou des phrases (juste : si compréhension du mot même ou seulement si mot est parfaitement prononcé).
- la calibration des tests pour contrôler l'effet de familiarisation.

A noter que peu d'études se sont spécifiquement intéressées à la reproductibilité des tâches d'identification d'unités segmentales comme les phonèmes. Van Nuffelen et al (2008) confirment les résultats antérieurs en faveur d'une bonne reproductibilité intra-juge mais mettent en évidence un insuffisance de reproductibilité inter-juge lorsque l'intelligibilité est très altérée.

C'est cette absence de fondement scientifique dans les bilans des TPPP qui a motivé mon engagement dans une recherche en sciences du langage. L'idée bien naïve était de construire un test à partir d'un texte assez court pour ne pas fatiguer les patients, contenant des cibles qui en fonction de modalités d'analyse adaptées permettraient de recueillir différentes informations à intégrer au bilan. Bien naïf en effet ... voilà où cela mène !

#### 2.4.2.2. LES MESURES PREDICTIVES

---

Depuis une quinzaine d'années, plusieurs techniques automatiques ont été proposées pour évaluer la parole sans avoir recours à des auditeurs mais en utilisant des machines standards. Ce dernier point est important d'une part pour rendre accessibles les outils développés et d'autre part pour que la difficulté à manipuler les appareils soit la source de la variabilité dans les résultats (Kayasith & Theeramunkong . 2009). En première intention, l'utilisation des systèmes automatiques pour réduire le handicap des personnes dysarthriques nécessitait une reconnaissance préalable de certains segments du signal (technique de contrôle d'assistance électronique – projet STARDUST (Parker M, et al. 2006). Ainsi à partir des recherches sur les interfaces « homme-machine » adaptables au handicap de parole et l'évolution de la reconnaissance automatique de la parole, plusieurs équipes de recherche ont testé et adapté des systèmes pour quantifier l'intelligibilité (Maier et al 2009, Van Nuffelen et al 2009). Ces travaux seront développés dans la section 4.2 de cette thèse.

## 2.5. QUELS SONT LES FACTEURS DETERMINANT L'INTELLIGIBILITE DANS LES TROUBLES DE LA PRODUCTION DE LA PAROLE

---

De bodt et al (2002) ont exprimé, à l'aide d'un modèle de régression multiple, l'intelligibilité en fonction d'une combinaison linéaire de 4 principaux facteurs : la qualité de la voix, l'articulation, la nasalité, et la prosodie. Chaque facteur était jugé par des auditeurs. A partir d'un groupe de 79 patients dysarthriques, plusieurs étiologies confondues, ils démontrent que la dimension « articulation » est nettement dominante. Cependant, l'inclusion des autres dimensions (nasalité, voix et prosodie) contribuent à une meilleure estimation de l'intelligibilité. Ces résultats confirment les données cliniques en utilisant le cadre perceptif habituel des bilans cliniques.

Pour recenser les informations sur les paramètres acoustiques et linguistiques contribuant à l'intelligibilité (intelligibilité définie comme l'efficacité du processus de décodage, dans les troubles de la production de la parole), nous allons recourir à toutes les ressources disponibles, balayant les travaux issus des TPPP aux études plus fondamentales sur la perception de la parole.

---

### 2.5.1. TRAVAUX ISSUS DU DOMAINE DES TPPP

---

#### 2.5.1.1. INFORMATIONS PROVENANT DE L'EXPLORATION DES TPPP

---

En utilisant le « word intelligibility test » à base de paires minimales pour un groupe de patients atteints d'une Sclérose latérale amyotrophique, Weismer & Martin (1992), par une analyse de régression en paliers, ont mis en évidence la responsabilité prépondérante du contraste « fricatif alvéolaire » versus « fricatif palatal » (82,3% de la variance) en anglais. Interviennent également le contraste « occlusif » versus « nasal » et enfin du /h/ initial versus aucun contraste mais ce dernier n'est significatif que pour les hommes. Comme explication pour cette différence entre les genres Weismer & Liss (1991) ont évoqués les variations de fonctionnement laryngé avec l'âge entre les 2 sexes. Au cours du même travail, la valeur prédictive de la mesure du F2 a été découverte ouvrant le champ de la recherche des facteurs acoustiques.

Par la même méthode :



- Billman (1986), rapporté par Kent (1992) retrouve une influence plus importante sur l'intelligibilité de l'omission des singletons pré vocaliques et de la postériorisation dans une population d'enfants avec des troubles phonologiques ;

- Whitehill (2010) dans un travail consacré aux paralysies cérébrales montre le rôle dominant du contraste d'aspiration en mandarin et un rôle moins important des contrastes entre voyelles.

En ce qui concerne les paramètres acoustiques, la poursuite des recherches initiées par Weismer & Martin (1992) ont débouché sur un modèle statistique de la pente de F2 en relation avec la sévérité de la dysarthrie (Kim Y et al. 2009). Evalués dans plusieurs contextes pathologiques (Parkinson : Weismer et Martin 1992, Tjaden & Wilding 2004, Sclérose en Plaque : Tjaden & Wilding 2004, Sclérose latérale amyotrophique : Weismer et Martin 1992, glossectomies partielles pour cancer : Whitehill et al 2006), les paramètres relevant du second formant que sont l'étendue et surtout la trajectoire du F2 et l'espace vocalique ont fait preuve d'une bonne corrélation avec les mesures d'intelligibilité à partir de la perception d'un groupe d'auditeurs.

Les résultats obtenus par cette méthode intéressent les contrastes phonétiques et paraissent dépendant des spécificités de la pathologie causale sur le profil d'erreurs de production. Les contextes linguistiques (anglais, mandarin...) sont également différents. Ainsi, comme le fait remarqué Whitehill (2010), malgré les arguments en faveur de l'universalité des lésions neurophysiologiques du système de production de la parole et le fait que dans certaines langues des patrons d'erreurs communes aient été observés, des différences spécifiques aux traits linguistiques ont été signalés. La fréquence relative des contrastes liée à la pathologie et à la langue limite la généralisation des conclusions rendant plus pertinent la recherche à un niveau plus acoustique que linguistique.

#### 2.5.1.2. TRAVAUX ISSUS DE LA MANIPULATION DE SIGNAL DANS LE CADRE DES TPPP

---

Quelques travaux pour améliorer la parole dysarthrique par des systèmes automatiques contribuent à la connaissance des facteurs déterminant l'intelligibilité. Menendez-Pidal et al (1996) en créant des stimuli « hybrides » pour « corriger » de la parole dysarthrique avec de la parole naturelle sur des enregistrements de phrases sans sens ont trouvé que le remplacement du spectre à court terme d'une parole de dysarthrique par celui d'une parole non dysarthrique (en gardant la prosodie du

patient dysarthrique) conduit à une amélioration de l'intelligibilité à 87% par rapport à une ligne de base de 68%. Le remplacement du spectre à court terme uniquement des voyelles, des liquides et des glides conduit à une intelligibilité de 75% alors que le remplacement des autres types de sons produit une intelligibilité à 73%. Le remplacement de la prosodie (fréquence, durée et énergie) de la phrase avec celle du locuteur non dysarthrique, en gardant le spectre à court terme du locuteur dysarthrique conduit à un score d'intelligibilité de 75%. Les facteurs qui influent sur l'intelligibilité sont donc nombreux et inter-dépendants. Ce travail a été réalisé à partir de la data base de Nemours qui comprend des locuteurs avec paralysie cérébrale et traumatisme crâniens.

Maassen et Povel (1984, 1985) dans une étude utilisant une approche similaire consacrée à la parole produite par des sourds montre que la correction artificielle de la trajectoire de la fréquence fondamentale produit peu de gain en terme d'intelligibilité (de 20 à 27%) alors que le remplacement de la durée des phonèmes avec la fréquence fondamentale par les valeurs d'une parole produite par un normo-entendant génère un meilleur gain (de 24 à 34%). Mais c'est surtout la correction segmentale qui produit le meilleur gain de 24 à 72%. Ceci a permis aux auteurs de conclure que dans la parole produite par un sourd, l'amélioration de l'articulation est plus importante que celle de la prosodie.

Kain 2007 par des manipulations de type analyse-correction et resynthèse de signal obtient un gain d'intelligibilité de 48 à 54% à partir d'un modèle utilisant la durée de la voyelle et les points stables de F1-F3 sur une tâche d'identification de voyelles. Il discute l'importance des consonnes par elles-mêmes et part le fait qu'elles sont déterminantes pour la reconnaissance des voyelles. Kain A et al (2008) poursuivent ces travaux en utilisant eux aussi une technique d'"hybridation" pour rechercher le degré de contribution de chaque trait à l'intelligibilité. Les résultats pour un locuteur sont en faveur d'une amélioration significative de l'intelligibilité quand on remplace certaines combinaisons des spectres à court terme, l'identité phonémiques, et la durée de phonèmes. En revanche aucune amélioration n'est survenue avec les combinaisons intéressant le spectre, la fréquence fondamentale, l'énergie et les événements de non parole (pauses par exemple).

---

### 2.5.2. TRAVAUX ISSUS DE LA PERCEPTION DANS LE BRUIT

---

Les premiers travaux réalisés par Nicely & Miller (1955) ont été confirmés par les nombreuses études qui ont suivi (Dubno & Levitt 1981, Gordon-Salant 1986, Hazan & Simpson 1998). Le trait de

voisement est très robuste dans le bruit (90%) alors que le mode mais surtout le lieu d'articulation sont peu robustes.

Améliorer l'intelligibilité dans le bruit est devenu un véritable enjeu pour rendre le système d'aides auditives plus efficace. Les techniques d'amplification de certaines parties du signal de parole ont été expérimentées pour optimiser les appareils auditifs et perfectionner les implants. L'intérêt de n'amplifier qu'une partie du signal a justifié des travaux contribuant à déterminer les cibles les plus pertinentes pour obtenir un gain utile d'intelligibilité alors que de manipuler la durée de la consonne s'est avéré inefficace (Gordon-Salant 1986).

Parmi les parties du signal à amplifier, plusieurs solutions ont été évaluées comme, l'amplification des zones consonantiques dans une syllabe CVC sur des mots et des non mots (Hazan & Simpson 1998), l'amplification des voyelles dans des syllabes de types VCV (Meyer & Morse 2003).

Les résultats sont en faveur :

-d'une information phonétique plus consistante pour le contexte /a/ et moins consistante pour les contextes /i/ et /u/ (Gordon-Salant 1986, Hazan & Simpson 1998, Meyer & Morse 2003) traduisant le fait que le contexte /a/ est plus robuste dans le bruit ;

-de l'importance de l'amplitude relative des consonnes c'est de dire de l'importance de maintenir une différence d'intensité entre les consonnes permettant faire la différence entre les consonnes de faible et forte amplitude. Les consonnes de faible amplitude étaient plus facilement identifiées dans ces manipulations (Stevens 1985, Meyer & Morse 2003).

Ces travaux retrouvent l'importance des transitions avec F2 (Meyer & Morse 2003).

---

### 2.5.3. TRAVAUX ISSUS DE LA MANIPULATION DE SIGNAUX HORS CONTEXTE DES TPPP

---

La recherche motivée par la technologie des télécommunications est une source importante d'information sur la robustesse du signal définie comme la conservation de son efficacité en termes d'intelligibilité malgré un certain degré de dégradation ou modification.

### 2.5.3.1. LA PAROLE COMPRESSEE

---

L'intelligibilité de la parole en fonction d'une compression temporelle a été étudiée dès la mise au point d'un appareil électromécanique par Fairbanks et col. Les stimuli utilisés étaient 50 monosyllabes phonétiquement équilibrées avec une durée moyenne de 180 ms par phonème. Les données montrèrent une reconnaissance subnormale pour des compressions raccourcissant les mots stimuli à 20% de leur durée d'origine. Quand les mots étaient réduits à 10% de leur taille d'origine, l'intelligibilité atteignait 50% de la normale. Daniloff et al (1968) examinèrent les effets de différents degrés de compression temporelle sur la reconnaissance des voyelles. Ils trouvèrent que la compression temporelle affectait l'intelligibilité à partir d'une réduction temporelle à 30% de la durée originale. Nagafuchi (1976) analysa l'intelligibilité de la distorsion de sons de la parole chez les enfants normaux entendants. Utilisant 20 mots monosyllabiques, il démontra que la discrimination n'était pas altérée jusqu'à un allongement de 300% de la durée originale des mots. La reconnaissance commençait à diminuer progressivement quand les mots «étaient réduits à 50% de leur durée d'origine, elle chutait rapidement à partir de 25%. Wilson et al 1994 ont montré que l'effet de la compression sur l'intelligibilité des scores de reconnaissance des mots chute de 90 à 25% quand la compression augmente de 45 à 75 %. La diminution est graduelle sur la performance de reconnaissance de mot de -1% pour 1% de compression quand le pourcentage de compression se situe entre 45 à 65%. La chute est plus importante quand la compression se situe entre 65 et 75% : - 2,8% pour 1% de compression. Gordon-Salant (1993) dans une étude sur l'effet de la compression temporelle sur la reconnaissance de la parole chez les sujets âgés et les sujets jeunes montre que l'âge et les troubles de l'audition contribuent indépendamment au déficit de reconnaissance de la parole compressée.

Ces différents résultats sont en faveur de l'absence d'altération sur l'intelligibilité dans une gamme de vitesse allant de 50% de réduction à 300% d'allongement.

### 2.5.3.2. DANS LE DOMAINE FREQUENTIEL

---

Dans le domaine fréquentiel, la perception du signal utilise essentiellement les fréquences comprises entre 100 et 5000 Hz mais comme nous le démontre quotidiennement le téléphone, malgré la suppression d'une large partie de ces fréquences le signal demeure compréhensible (Fletcher, 1929). Le signal peut être échantillonné, voire discrétiser complètement en remplaçant les valeurs positives par +1 et les valeurs négatives par -1 (Licklider. 1946, 1950) et demeurer compréhensible.

Les phonèmes avec des parties stables (les voyelles et les fricatives) seront plus résistants que les phonèmes rapides ou sans partie stable. En effet, l'identification de la parole est sensible aux changements rapides ou sans partie stable : si la durée des transitions diminue, les changements de fréquence sont plus difficiles à détecter (Wieringen et al 1995).

Les travaux récents de Li et al (2007) précisent la fréquence et la durée totale du masquage déterminant pour la reconnaissance de la parole. Les meilleures performances sont obtenues quand les auditeurs sont capables de percevoir les informations dans les régions fréquentielles de F1/F2 soit 0 à 3kHz sur au moins 60% de la phrase.

### 2.5.3.3. L'ENVELOPPE TEMPORELLE

---

Comme nous l'avons dans la section concernant la perception de la parole, plusieurs études plaident en faveur de l'importance fondamentale pour l'intelligibilité d'une modulation de fréquence de l'enveloppe temporelle pour permettre la segmentation du signal.

Les travaux de Houtgast & Steeneken (1985) sur la diffusion de la parole dans les salles acoustiques ont mis en évidence une meilleure intelligibilité de la parole lorsque l'enveloppe temporelle présente des modulations de fréquence d'environ 3-4 Hz, fréquence correspondant au rythme syllabique. Shannon et al (1995) ont également apporté des arguments pour l'importance de l'intégrité de l'enveloppe en présentant à des auditeurs une bande de bruit lissée avec la forme de l'enveloppe d'intensité d'un signal de parole. Dans ces conditions, malgré la perte de toute information fréquentielle fine, il reste possible d'accéder à une représentation linguistique du signal.

Une information précise quant au moment d'occurrence des différents événements fréquentiels du signal dans le temps ne semble pas nécessaire. Aarai & Greenberg (1998) ont appliqué une désynchronisation artificielle provoquant une distorsion du spectre de modulation. Ils montrent que ces ruptures de synchronies ne gênent pas considérablement l'identification de phrase pour un décalage inférieur à 150 ms.

Une meilleure intelligibilité malgré une altération ou une distorsion du signal lorsqu'une modulation temporelle favorable est conservée est également un argument pour l'importance cruciale de la forme de l'enveloppe temporelle. Miller et al (1950) ont les premiers démontré que malgré un masquage ou la suppression de signal de parole à intervalle régulier, le signal peut rester intelligible. Meunier et al

2002 utilisant une parole inversée temporellement démontre l'importance de la dynamique temporelle dans la reconnaissance de mot et de pseudo mots en français chez l'adulte. Récemment, Grataloup et al (2007) précisent les résultats obtenus en prenant la syllabe comme unité de base. Ils montrent que l'intelligibilité reste bonne jusqu'à une taille d'inversion temporelle égale à une syllabe pour des mots et des pseudo mots bi syllabiques.

La redondance du signal de parole permet au cerveau d'interpréter un signal incomplet à condition qu'une modulation de fréquence soit respectée. Cette fréquence est proche du rythme syllabique.

#### 2.5.3.4. LA RESTAURATION PHONEMIQUE

---

Cole (1973), étudie la détection par des auditeurs « sains » de mauvaises prononciations produites par le changement d'un son consonantique d'un mot trisyllabique par un, deux ou quatre traits distinctifs. Les résultats révèlent que le changement d'un seul trait distinctif sont rarement détectés alors que la modification de deux ou quatre traits sont facilement détectés. La syllabe (1<sup>ere</sup>, 2<sup>ème</sup> ou 3<sup>ème</sup>) dans laquelle survient la mauvaise prononciation n'influence pas la probabilité de détection. Cependant les temps de réaction pour les mots avec mauvaise prononciation étaient au moins de un tiers de seconde plus lent quand elle survient dans la première syllabe du mot.

Le fait que les mots mal prononcés par un seul trait distinctif étaient rarement entendus comme mal prononcés suggère que les sujets n'ont pas besoin de toutes les informations acoustiques présentes dans la forme sonore.

L'un des arguments les plus souvent cités en ce sens est le phénomène de restauration phonémique mis en évidence par Warren (1970).

Les travaux sur la restauration phonémiques et l'induction auditive principalement mené par Warren et al sont synthétisés dans le chapitre intitulé « auditory illusions and perceptual processing of speech » (Warren 1996). Ils apportent des informations sur le rôle de ce processus sur l'intelligibilité de la parole. Ainsi, lors des premières expériences, la suppression d'un phonème d'un mot, situé à la place du point d'unicité (c'est-à-dire qu'à partir du phonème précédent le son supprimé, il n'existe pas d'autre mot dans le lexique de la langue du test qui pourrait conduire à une confusion), a été remplacé en respectant sa durée initiale par un bruit (toux, son ou buzz). Les auditeurs n'ont pas repérés l'absence du son même s'il leur était spécifié qu'il en manquait un. De plus, il était impossible aux

auditeurs de définir précisément la place du son rajouté (Warren 1970). Cette incapacité à localiser le son introduit pour remplacer le son manquant, a elle-même été étudiée (Warren and Obusek 1971). Ainsi, plus le son est long, plus il est perçu avant sa localisation exacte dans le signal. Et plus l'auditeur réécoute le signal dans les mêmes conditions expérimentales, plus le délai s'allonge, s'éloignant du « lieu » de la restauration. En revanche, si le phonème supprimé n'est pas remplacé et qu'un silence demeure pendant la durée de son absence, l'auditeur identifie immédiatement la suppression. Des mesures de perception de la discontinuité et de l'intelligibilité ont démontré que le remplissage par du bruit des « trous » dans le signal augmente l'impression de continuité et d'intelligibilité. Cette impression d'amélioration de l'intelligibilité a été confirmée (Bashford 1992). L'importance de cette amélioration dépend de l'information syntaxique et sémantique contribuant à l'identification des fragments manquants. L'effet est maximum autour de 304 ms et si le bruit a un noyau fréquentiel proche du signal supprimé. La restauration phonémique apparaît ainsi comme une adaptation linguistique d'une capacité plus générale à restaurer des portions de signal qui sont masquées ou détériorées par un autre son. Cette illusion perceptive d'un son induit par la présence d'un autre a été appelée l'induction auditive.

Cet effet est très dépendant du contexte. Samuel 1981 a mis en évidence une meilleure restauration avec une augmentation des informations lexicales et rend compte d'effets top down mais également Bottom up car cet effet de restauration phonémique existe également pour les pseudo mots, des unités infralexicales pouvant influencer l'identification de la parole.

Ainsi l'intelligibilité de la parole est remarquablement robuste aux distorsions du signal acoustique mais cette conclusion est à relativiser par rapport au nombre de traits altérés par unité.

---

#### 2.5.4. LE CARACTERE NATUREL DE LA PAROLE

---

Le caractère naturel de la parole a particulièrement intéressé les spécialistes du bégaiement. Lors de premières études menées par Metz et al (1990) le « Voice onset timing » et la durée des phrases étaient particulièrement associées au caractère naturel de la parole avec une plus faible corrélation pour la durée des voyelles. L'influence de la longueur des voyelles a cependant été confirmée (Schaeffer & Eichorn 2001) avec la conservation du caractère naturel jusqu'à 200%. Cet effet de

l'allongement est réparti inégalement en fonction des voyelles. Par exemple, dans cette étude en langue américaine, la voyelle de « cake » perd de son caractère naturel dès 50% d'allongement alors que la voyelle de « cup » le perd seulement à 250%.

---

#### 2.5.5. EN SYNTHÈSE

---

Cette revue de la littérature confirme l'existence d'unités plus contributives pour l'intelligibilité avec la possibilité d'un phénomène d'induction malgré la détérioration d'un segment de signal. Par contre, il existe peu d'informations sur la capacité des auditeurs à restaurer la parole dans les TPPP. Ceci justifiera la recherche d'une contribution distribuée de manière inégale sur les différents sons de la langue française traité dans la cinquième partie de cette thèse.





Quelle intelligibilité donner à ce qui  
fait sens dans notre rapport sensible  
au monde ?

Danielle Dubois, Le sentir et le  
dire p13

### ***3. CATEGORISATION LIBRE DANS LES TROUBLES DE LA PRODUCTION PATHOLOGIQUE DE LA PAROLE***

---

3.	Catégorisation libre dans les troubles de la production pathologique de la parole	85
3.1.	Introduction.....	89
3.2.	Première étude.....	93
3.2.1.	Matériel et méthode	93
3.2.1.1.	Les locuteurs	93
3.2.1.2.	Le corpus	94
3.2.1.3.	Méthode	94
3.2.1.3.1.	Principe	94
3.2.1.3.2.	Méthode d'analyse des regroupements	95
3.2.1.3.3.	Méthode d'analyse des commentaires	95
3.2.1.4.	Les auditeurs	96
3.2.1.5.	Plan d'analyse des résultats	97
3.2.2.	Description de la procédure des tests	98
3.2.2.1.	Procédure du test de catégorisation	98
3.2.2.2.	Procédure de l'analyse des stimuli	99
3.2.2.2.1.	Analyse par un jury d'auditeurs experts	99
3.2.2.2.2.	Analyse acoustique	100
3.2.3.	Résultats	100
3.2.3.1.	Les locuteurs	100
3.2.3.2.	Les stimuli	101
3.2.3.2.1.	L'analyse perceptive par le jury	101
3.2.3.2.2.	L'analyse acoustique des stimuli	102
3.2.3.3.	Détermination des classes et construction des arbres	103
3.2.3.4.	Analyse des arbres	106
3.2.3.4.1.	Analyse des classes par rapport aux caractéristiques des locuteurs et des résultats de l'expertise du jury	106
3.2.3.4.1.1.	La « lecture » des arbres	106
3.2.3.4.1.2.	Corrélation entre les classes, le degré de sévérité et le degré d'altération de l'intelligibilité.	107
3.2.3.4.1.3.	Relation entre les classes, et les caractéristiqueS des stimuli en terme d'étiologie et de type de trouble et d'identité des locuteurs	108
3.2.3.4.1.4.	Analyse des classes par rapport aux caractéristiques de l'analyse perceptive issues de l'expertise du jury	111
3.2.3.4.2.	Analyse des classes par rapport aux caractéristiques des résultats de l'analyse acoustique de la prosodie par le logiciel EVA	112
3.2.3.5.	Analyse des commentaires	115
3.2.3.5.1.	grille d'analyse	115
3.2.3.5.2.	Références à l'intelligibilité et à la normalité	117
3.2.3.5.3.	Le support linguistique de référence	118
3.2.3.5.4.	Supposition par rapport au locuteur et référence au TPPP	120
3.2.3.5.5.	Sentiments « ressentis » par le locuteur à l'écoute du stimulus	121
3.2.3.5.6.	Critère principal et analyse de la stratégie des auditeurs	122
3.2.3.5.6.	Analyse des classes en fonction des commentaires	123
3.2.3.5.6.1.	Pour les experts	123
3.2.3.5.6.2.	Pour les non experts	124
3.2.3.5.6.1.	Pour les proches	124
3.2.4.	En résumé	125
3.3.	Seconde étude.....	126
3.3.1.	Matériel et méthode	127
3.3.1.1.	Les locuteurs	127
3.3.1.2.	Le corpus	127
3.3.1.3.	Méthode	130
3.3.1.4.	Les auditeurs	130
3.3.1.5.	Plan d'analyse des résultats	131
3.3.2.	Description de la procédure des tests	131
3.3.2.1.	Procédure du test de catégorisation	131
3.3.2.2.	Procédure de l'analyse des stimuli	131

3.2.2.2.1. Analyse par un jury d'auditeurs experts	131
3.2.2.2.2. Analyse acoustique	131
3.3.3. Résultats	132
3.3.3.1. Détermination des classes et construction des arbres	132
3.3.3.2. Analyse des arbres	135
3.3.3.2.1. Analyse des classes par rapports aux résultats de l'expertise du jury	135
3.3.3.2.1.1. Pour les experts (figures n°3.3.11 à n°3.3.13)	135
3.3.3.2.1.2. Pour les non experts	137
3.3.3.2.2. Analyse des classes par rapports aux caractéristiques des locuteurs (genre et pathologie)	138
3.3.3.2.2.1. Pour les experts	138
3.3.3.2.2.2. Pour les non experts	139
3.3.3.2.3. Analyse acoustique	141
3.3.3.2.2.1. Pour les experts	141
3.3.3.2.2.2. Pour les non experts	143
3.3.3.5. Analyse des commentaires	144
3.3.3.5.1. La grille d'analyse	144
3.3.3.5.2. Références à des critères identiques à ceux du GRBASI	144
3.3.3.5.3. Degré d'altération des paramètres et référence à la norme	145
3.3.3.5.4. Supposition par rapport au locuteur	145
3.2.3.5.5. Sentiments « ressentis » par le locuteur à l'écoute du stimulus	146
3.3.3.5.6. Référence à certains mécanismes et aux paramètres acoustiques des stimuli	147
3.3.3.5.7. Référence au support linguistique	147
3.3.3.5.8. Critère principal et analyse de la stratégie des auditeurs	149
3.3.3.6. Analyse des classes en fonction des commentaires	150
3.2.3.6.1. Pour les experts	150
3.2.3.5.6.2. Pour les non experts	151
3.3.4. Résumé-discussion	152
3.4. Comparaison résultat étude 1 et étude 2.....	153
3.4.2. Comparaison des résultats des classes	153
3.4.2. Comparaison entre l'analyse des commentaires du corpus comportant des pathologies vocales et du corpus comportant des pathologies de la production de la parole.	154
3.4.2.1. Les notions de normalité et les références au degré d'altération	154
3.4.2.2. Les supports de référence	155
3.4.2.3. La clé de tri principal	156
3.4.2.4. Les principaux critères de tri relevés	157
3.5. Discussion générale de la catégorisation .....	158
3.5.1. Proposition d'une hypothèse	158
3.5.2. La normalité et la déviance pathologique par rapport à la « normale »	161
3.5.3. Contribution à la définition de l'intelligibilité	163
3.5.3. La notion d'expert	164
3.6. .... Conclusion	166
.....	166



### 3.1. INTRODUCTION

---

La perception de la parole produite par des personnes souffrant de maladies sources de trouble de la production pathologique de la parole (TPPP) est à la base de la prise en charge du handicap qui en résulte. Elle est d'une part, un des piliers du bilan étiologique en ce qui concerne l'analyse perceptive des caractéristiques de la parole. Elle est, d'autre part, notamment à travers l'intelligibilité de la parole, l'indice de sévérité du trouble participant au suivi de l'évolution et à la mesure de l'efficacité des prises en charge.

Dans ces deux situations, la perception de la parole est abordée par un « interlocuteur – juge » idéalisé comme pouvant fournir une mesure de la déviance par rapport à une norme. Or, la notion de norme est largement débattue en linguistique. De nombreux auteurs tendent à considérer que la norme n'existe pas (Borrell & Billières 1989).

Ces considérations adaptées à notre approche des TPPP, soulève la pertinence des critères de jugement utilisés dans la pratique clinique. Plus précisément, est ce que ces critères de jugement répondent aux stratégies perceptives des différents protagonistes impliqués dans la prise en charge de ces personnes alors que leur sélection est principalement issue de travaux menés par des chercheurs préoccupés au premier plan par l'enquête étiologique neurologique? Si cette question est régulièrement soulevée pour les thérapeutes en charge de la réhabilitation de ces personnes, elle a rarement été abordée du point de vue d'un interlocuteur qui cherche à comprendre ce type de parole dans une situation de communication telle qu'elle se présente dans la vie courante des patients. (Bloch & Wilkinson 2009).

Ainsi, évaluer l'impact des facteurs qui peuvent améliorer la communication est un problème qui se pose différemment (Yorkston et al, 1999) incitant à revisiter la place de la perception de ce type de parole.

D'un point de vue historique, l'analyse perceptive a été fortement structurée par les travaux de l'équipe de Darley, Aronson et Brown (Darley et al 1969 a et b) avec un objectif diagnostique basé sur le constat qu'il était possible auditivement de distinguer plusieurs types de dysarthries en référence aux syndromes neurologiques en cause. Dans ce but, ont été utilisés des critères utilisés par les

professionnels impliqués avec une échelle de cotation ordinale à 7 niveaux. Cette approche a permis une description détaillée de la parole dysarthrique mettant en évidence que les mêmes critères déviants sont présents dans plusieurs types de dysarthrie avec des fréquences différentes en fonction du critère et du type de dysarthrie. Par exemple une imprécision des consonnes est présente dans tous les types de dysarthrie alors qu'une accélération paroxystique anormale du débit est retrouvée presque exclusivement dans les maladies de Parkinson. Les critères perceptifs issus de ces travaux sont donc le fruit d'une écoute par un groupe d'interlocuteur ayant un profil particulier, orientée par un objectif diagnostique. Elle reste utilisée (Bunton et al) avec différentes variantes ou adaptations dans les bilans d'évaluation cliniques y compris français (Auzou 1998). Les limites de leur utilisation dans ce but reflète les limites de la corrélation entre le déficit lié aux mécanismes étio-pathogéniques de la maladie et à l'incapacité fonctionnelle sur la parole qui en découle. Par exemple, il n'y a pas de parfaite concordance entre une paralysie du voile du palais et un défaut de fermeture vélopharyngée en phonation comme il n'y a pas de parfaite concordance entre un défaut de fermeture du voile du palais et la perception d'un défaut de nasalité dans la parole. Elles sont également en rapport avec les limites de reproductibilité de l'analyse perceptive. En effet, si comme pour la perception de la voix, les auteurs estiment que la fiabilité est suffisante pour une pratique clinique courante (Dejonckere et al 1998, De Bodt et al, 1997, Ozsancak et al. 2002), la concordance interjuges reste limitée pour les critères pertinents par la fréquence de leur altération chez les patients dysarthriques. Le taux de concordance est, par exemple, à 68% dans une étude sur les ataxies de Friedreich (Kearns & Simmons 1988). Pour la voix, les travaux de Kreiman et Garret (Kreiman et al 1998) précisent que les auditeurs sont relativement d'accord lorsque la qualité de la voix est presque normale ou très sévère mais qu'ils ne présentent jamais un accord suffisamment robuste dans le milieu des échelles et que seulement 32% de la variance de ces mesures est en fait en rapport avec la différence de qualité de la voix. La recherche d'explications d'une telle variabilité (Kreiman et al 2000) a révélé que les auditeurs présentent de grandes différences quant aux traits auxquels ils font attention quand ils jugent la qualité de la voix. De plus, le degré d'attention apporté par les auditeurs à une dimension dépend en partie de l'importance de ce changement au sein du corpus utilisé pour le test et donc du contexte. Enfin, si le nombre de stimuli est grand, la perception ne se fait pas de manière continue mais forme des « clusters » sans que ce regroupement recouvre les déterminants perceptifs classiques comme le caractère voilé ou éraillé par exemple. Cependant, l'utilisation d'une tâche

d'ajustement (Gerratt et al 2001) dans laquelle les auditeurs font varier les paramètres d'un appareil de synthèse pour créer des sons s'accordant de manière acceptable à un stimulus naturel conduit à des taux de concordance de 96% ce qui démontre que la variabilité individuelle est liée au design des tâches demandées et non à l'absence de fiabilité et reproductibilité des auditeurs (Kreiman et al 2007).

L'intelligibilité de la parole a été identifiée comme le principal critère permettant de suivre l'évolution de la sévérité des TPPP (Kent 1992). Malgré le consensus implicite autour de cette indexation de la sévérité par la mesure de l'intelligibilité, peu de travaux se sont attachés à définir objectivement ce qui détermine le caractère sévère du TPPP. Plusieurs auteurs relativisent la prédominance de ce critère (Bunton et al 2000). Ainsi, Auzou (2007) introduit la notion de « sévérité perceptive » qui prend en compte l'ensemble des « déviations » de la parole (qualité vocale, prosodie, intelligibilité...). De la même manière, Bunton et al (2000) suggère que la sévérité du TPPP peut être décrite par plusieurs paramètres dont le degré de modification de la prosodie. De plus, comme toutes les autres mesures perceptives, sa validité est controversée. Les procédures utilisés pour quantifier l'intelligibilité en dehors des pourcentages d'identification de mots sont discutées quand à leur fiabilité (Schiavetti 1992, Weismer & Laures 2002).

L'intelligibilité en tant que mesure de l'impact des facteurs participant à la communication est un concept plus récent (Yorkston et al 1999). Hustad (2008) a démontré que les scores d'intelligibilité obtenus par la transcription du signal de parole perçue ne reflètent pas les scores de compréhension obtenus par l'adéquation de la réponse à une question posée par une personne présentant un TPPP. En d'autres termes, l'intelligibilité et la compréhension sont des notions qui ne se recouvrent que partiellement lorsqu'elles sont définies de la manière suivante :

L'intelligibilité : précision avec laquelle l'auditeur récupère le signal acoustique produit par le locuteur

La compréhensibilité : capacité de l'auditeur à interpréter le sens du message produit par un locuteur sans tenir compte de la précision ou de la justesse phonétique ou lexical.

L'intelligibilité ne peut pas, par conséquent, être le seul critère de définition de la sévérité des TPPP en référence au contexte de communication du patient. Ceci s'intègre parfaitement à la conception que la parole est produite dans le but de se faire comprendre par un auditeur telle qu'elle a été définie



par le modèle (Hypo/hyper) de Lindblom (1990) où la variabilité en production du signal de parole s'explique par le fait que la production de la parole est une adaptation permanente aux diverses situations de communication, l'auditeur demeure la cible et ses capacités à réagir au signal de parole produit reste la finalité de l'action entreprise par le locuteur à l'origine d'une coopération entre locuteur et auditeur . Ceci justifie le fait que la notion de sévérité ne se soit pas cantonnée à des mesures d'altération du signal acoustique. Le jugement par le l'auditeur reste une référence faisant basculer la problématique sur le versant de perception de la parole malgré sa complexité (contraste perceptif minimal).

Ainsi, bien que la perception auditive de la parole soit largement utilisée en clinique, la pertinence des paramètres utilisés dans les tests reste discutable et très peu de travaux sont consacrés à la recherche des critères qui sous-tendent le jugement des auditeurs lors de l'écoute de TPPP. Or l'identification des sources de la dégradation de la compréhension en situation de communication serait fondamentale pour améliorer la prise en charge des patients atteints de TPPP (Bloch & Wilkinson 2009). Mais, la recherche des paramètres les plus pertinents pose des problèmes méthodologiques majeurs. En effet, cette recherche sous-entend la connaissance des critères de compréhension des interlocuteurs lors de l'écoute de ce type de production de la parole. Or, les solutions pour mettre en lumière ces paramètres posent des problèmes fondamentaux. Nguyen (2005) a fait le point sur les problèmes méthodologiques liés à l'étude de la perception. De récentes contributions (Magnen 2009) montrent que les méthodes issues de la psychologie cognitive et des théories autour de la catégorisation prototypique appliquée aux sensations (Dubois, 1993, 2010) sont adaptées à la perception de la parole. Les tests de catégorisations libres (Gaillard 2009, 2010) peuvent ainsi contribuer à compléter et améliorer les méthodes d'évaluation des TPPP.

Le cadre théorique dans lequel s'inscrit la catégorisation libre est en pleine évolution. Il est lié à la problématique récente de la catégorisation en psychologie cognitive, principalement développée dans le domaine verbal (Rosch & Lloyd 1978) et plus récemment, dans d'autres domaines sensoriels dont les sons environnementaux (Guyot, 1996 ; Maffiolo, 1999 ; Guastavino, 2007), la parole (Magnen, Billières, & Gaillard, 2005, Magnen, 2009), les sons musicaux (Gaillard 2000 ; Gaillard, Castellengo, & Dubois 2007 ; Morange, Fontaine & Dubois 2007, 2009) ou simplement de la voix (Morange, Fontaine, Vogel, & Dubois, 2005). Ce cadre théorique et les méthodologies attenantes (tris libres et arbres de distances) sont en cours d'application dans le domaine des pathologies (en aphasie par exemple).

Les concepts et méthodes élaborés dans ces études ont ainsi permis d'identifier les critères pertinents de la construction des catégories perceptives. L'application à la perception des voix pathologiques est une suite naturelle qui fait l'objet d'un programme de recherche spécifique actuellement en cours.

Face à ces différents constats, l'objectif principal de ce chapitre est d'étudier comment sont perçus les TPPP afin de comprendre les stratégies cognitives sous-jacentes à la perception de la parole pathologique. Les hypothèses de départ sont :

1/ que les auditeurs vont utiliser comme principal critère de tri le degré d'altération de l'intelligibilité de la parole et ce, quelle que soit l'expertise de l'auditeur et,

2/ que les critères et méthodes employés par les participants vont être différents selon leur expertise supposée. L'objectif complémentaire consiste à observer les commentaires des sujets à propos de la parole perçue. L'hypothèse à ce sujet est qu'un expert (dont on définira plus loin les caractéristiques) utilisera un vocabulaire spécialisé lui permettant de mieux définir les troubles.

Pour répondre à cet objectif, nous avons réalisé des expériences de catégorisation libre à partir d'un corpus de patients présentant un TPPP puis dans un second temps la même expérience a été effectuée à partir d'un corpus de patients présentant une dysphonie. Cette seconde expérience est en cours et nous ne présenterons ici que les résultats préliminaires qui permettront de répondre à une partie des questions soulevées par la première expérience.

---

## **3.2. PREMIERE ETUDE**

---

### **3.2.1. MATERIEL ET METHODE**

---

#### **3.2.1.1. LES LOCUTEURS**

---

Trente trois locuteurs adultes ont été enregistrés : 13 témoins et 20 patients. Les enregistrements ont été réalisés consécutivement lors des consultations pour bilan d'un trouble de la production de la parole à l'Unité de la voix et de la déglutition dans le service ORL de l'hôpital universitaire Larrey Rangueil à Toulouse. L'entourage des patients et le personnel soignant, ont pendant cette période, accepté d'être enregistrés pour constituer le groupe des témoins. Les enregistrements ont été effectués après information et obtention du consentement des patients.

### 3.2.1.2. LE CORPUS

---

Le corpus est constitué de la dernière phrase du texte « Pierrot » de Maupassant, utilisée en pratique dans la version de 1998 de l'ECD (Evaluation clinique de la dysarthrie, Auzou & al. 1998) : « Et elles s'épouvantaient pour l'avenir, comment dormir tranquillement maintenant ? ». Le choix de cette phrase a été déterminé par le fait que c'est la partie du texte qui fait l'objet de moins d'hésitation lié à la lecture. Elle a été enregistrée avec un microphone serre-tête Shure®, une carte d'acquisition Edirol® et un ordinateur Macintosh®. La phrase choisie a été extraite du texte lu avec le logiciel PRAAT (Boersma & Weenick 2009) sous la forme d'un fichier au format wav. Les 33 fichiers ainsi obtenus, ont été utilisés comme stimuli dans le logiciel d'aide informatisée à la réalisation de classifications libres de sons. Aucun traitement ultérieur n'a été effectué sur ces enregistrements. En particulier, aucune normalisation en intensité n'a été faite. En effet, la différence d'intensité entre les locuteurs fait entièrement partie des caractéristiques de la parole pathologique.

### 3.2.1.3. METHODE

---

#### 3.2.1.3.1. PRINCIPE

Les tests de catégorisation libre consistent à proposer une série de stimuli auditifs à des auditeurs et à leur demander de les classer par affinité, selon leurs propres critères (sans préciser les critères à utiliser et sans demander d'explication a priori), en une série de plusieurs catégories. Chaque auditeur est maître à la fois du contenu et du nombre de catégories. Aucune « bonne réponse » n'est attendue et l'évaluation se fait en comparant les mesures de « distance » qui séparent les stimuli entre eux. Ainsi ce qui est testé n'est pas l'audition mais bien les stratégies qui permettent de déterminer les critères pertinents qui vont aboutir à la classification des sons. Pour réaliser ces tests, nous avons utilisé le logiciel TCL-LabX spécifiquement développé à l'Unité de Recherche Octogone par Pascal Gaillard (Gaillard 2009, annexe 7.3). L'analyse des résultats de la méthode de tri libre repose sur la mesure de la distance entre les réponses des participants et la construction d'un arbre de proximité entre les stimuli (Sattah & Tversky, 1977; Barthelemy & Guénoche 1988). L'évaluation de la stratégie cognitive auditive des participants pour la construction des classes répond à l'analyse des résultats du classement et de l'étude des éventuels commentaires. Ces analyses sont explicitées dans Poitevineau (2009) et reprennent des algorithmes de A. Guénoche. A la fin de la catégorisation, nous

demandons aux sujets de s'exprimer sur chacune des classes ou catégories qu'ils ont formées. Ces commentaires sont ensuite analysés selon la méthode exposée plus loin.

#### 3.2.1.3.2. METHODE D'ANALYSE DES REGROUPEMENTS

Parmi les méthodes de reconstruction et de visualisation d'arbres, « Additive-tree » (ADDTREE, Sattath et Tversky1977) a été appliqué pour ce travail. Utilisé pour l'inférence d'arbres phylogénétiques à partir des distances évolutives, elle reconstruit la structure d'un arbre en commençant par un arbre en étoile qui contient  $n$  feuilles associées aux objets et  $n-1$  branches. L'étoile est développée au fur et à mesure, par ajout de nouveaux nœuds internes jusqu'à ce qu'un arbre binaire comprenant  $2n-2$  nœuds et  $2n-3$  branches soit obtenu. Sur l'arbre, les nœuds et les feuilles forment un ensemble de points reliés entre eux par les branches, « de telle façon que tout point soit connecté à un autre et qu'il n'y ait pas de circuit fermé » (Poitevineau, 2009). Les nœuds ou embranchements de l'arbre sont en fait définis par les distances existant entre les objets. Ces distances sont ensuite représentées en créant un nœud sur une arête. Les scores calculés pour chaque point sont définis comme la « force de voisinage » de chaque paire de stimuli. Selon le dessin particulier, l'arbre obtenu « n'induit rien sur une éventuelle disposition des objets dans le plan : la distance entre deux objets fournie par cette représentation est la somme des longueurs des arêtes du chemin qui les joint ; ce n'est pas leur distance euclidienne dans le plan » (Barthélémy & Guénoche 1988). Pour l'ensemble des calculs et des représentations des arbres, ont été utilisés les logiciels spécifiquement développés comme Partition et AddTree (Poitevineau 2009). L'affichage des arbres le logiciel Dendroscope (Huson et al 2007).

#### 3.2.1.3.3. METHODE D'ANALYSE DES COMMENTAIRES

L'analyse des commentaires est conduite selon la méthode de l'analyse thématique de contenu des discours utilisée dans les enquêtes par l'entretien (Blanchet & Gotman 2009). En effet, la possibilité offerte aux auditeurs de définir les critères utilisés pour réaliser le groupement des sons s'apparente à la réalisation d'une enquête par entretien à usage complémentaire. Le recueil de ces commentaires contribue à la construction et à l'interprétation des résultats obtenus préalablement par l'expérience du tri libre non verbalisé. L'analyse thématique défait la singularité du discours d'un auditeur et consiste à découper transversalement tout ce qui, d'un recueil de commentaire à l'autre, se réfère au même thème. Elle ignore ainsi la cohérence singulière du commentaire unique et cherche une

cohérence thématique inter-commentaires (se rapportant aux mêmes éléments de tris). L'unité de découpage est le thème qui représente un fragment de discours, ici un fragment de commentaire. Chaque thème est défini par une grille d'analyse élaborée empiriquement. Le mode de découpage est stable d'un entretien à l'autre, c'est-à-dire de l'ensemble de commentaires émis par un auditeur à l'autre. Ceci permet d'envisager des traitements statistiques descriptifs et la recherche d'inter corrélation. En raison du caractère exploratoire des expériences de tri libre, l'identification des thèmes et la construction de la grille d'analyse s'effectue autant à partir des hypothèses descriptives de la recherche qu'après la lecture de l'ensemble des commentaires. Ces hypothèses procèdent d'une itération entre hypothèses et corpus, ici le thème de l'intelligibilité ou de la sévérité et les autres thèmes qui seront éventuellement repérés après avoir pris connaissance de tous les commentaires émis par l'ensemble des auditeurs.

#### 3.2.1.4. LES AUDITEURS

---

Pour les auditeurs, aucune hypothèse sur les critères d'âge et de sexe n'a été formulée. Seul, le critère de l'expertise (Bunton et al 2007) nous a semblé pertinent dans une première phase puis le critère de familiarité dans le sens où peuvent le vivre les proches des personnes souffrant de TPPP (Bloch & Wilkinson 2004). Le premier critère participe à la compréhension des processus sous-tendant l'expertise dans le cadre des bilans réalisés par les professionnels pour la prise en charge des patients. Le second critère permet de chercher si il existe des processus particuliers chez les proches et qui seraient à considérer pour relativiser les objectifs de prise en charge.

Au nombre de 40, les auditeurs correspondent à 3 groupes en fonction de leur niveau d'expertise dans le domaine des TPPP.

Un groupe d'auditeurs non expert mais sensibilisé aux questions de pathologie est constitué de 15 personnes étudiant dans le domaine de la santé. Ce groupe est constitué de 9 étudiantes en orthophonie de première année (9 femmes), 4 étudiants en médecine de 3ème année (2 femmes et 2 hommes) et 2 étudiantes en pharmacie de 2ème année (2 femmes). L'âge moyen est de 23 ans avec des extrêmes de 20 ans et 27 ans. Aucun des étudiants n'a encore été en contact pour des raisons personnelles ou professionnelles avec des personnes présentant un trouble de la production pathologique de la parole.

Un groupe d'auditeurs expérimentés est constitué de 15 professionnels prenant en charge pour des bilans et ou de la rééducation des patients présentant des troubles de la production pathologique de la parole (familiarisation professionnelle) comprenant 12 orthophonistes, toutes des femmes et 3 ORL phoniatres (2 femmes et 1 homme). L'âge moyen est de 45 ans avec des extrêmes de 29 ans et 65 ans.

Un groupe de personnes proches de patients présentant un TPPP concerne des personnes venues accompagner un patient hospitalisé à l'Unité de la voix et de la déglutition du service ORL de l'hôpital Larrey au CHU de Toulouse pour un bilan de parole et/ou de la déglutition. Ce groupe se compose de 4 aidants (3 femmes, 1 homme) s'occupant plus de 6 heures par jour du patient, de 6 personnes de la famille dont une sœur, une belle fille et 4 époux (2 hommes et 2 femmes), soit au total 10 personnes (7 femmes, 3 hommes) dont l'âge moyen est de 56 ans avec des extrêmes de 25 ans et 76 ans.

#### 3.2.1.5. PLAN D'ANALYSE DES RESULTATS

---

Une première phase descriptive précise les caractéristiques de la population des locuteurs (âge, sexe, pathologie à l'origine du trouble de la production de la parole) et les caractéristiques de la production de la parole. Ces paramètres sont obtenus par l'écoute d'un jury de 5 experts (2 phoniatres et 3 orthophonistes) avec une méthode similaire à celle habituellement employée en clinique décrite en détail dans le chapitre suivant.

Elle comprend :

- la détermination d'un indice de sévérité du TPPP et d'un indice d'altération de l'intelligibilité sous la forme d'une variable continue étendue de 1 à 30,
- la mesure du pourcentage d'identification des mots du corpus comme indice d'intelligibilité (étendue de 0 à 100%),
- le jugement des qualités concernant la voix, la résonance, la prosodie et la qualité phonémique (ou qualité de l'articulation des phonèmes) avec une échelle ordinale à 4 niveaux (de 0 à 3).

Puis les arbres de distances et les classes reconstruites sont visualisés pour chaque groupe d'auditeurs. La lecture des arbres permet de déterminer les critères généralement utilisés par les 3 populations de sujets en fonction des classes observables. L'analyse des commentaires quantifie le nombre de référence à ces différents critères et à la notion d'intelligibilité qui fait partie de l'hypothèse de départ.

Les classes sont ensuite étudiées en fonction des commentaires des auditeurs, puis des caractéristiques des locuteurs et enfin des caractéristiques de la production de la parole déterminées par le jury. La recherche de différence de distribution entre les différentes classes et les variables quantitatives relève d'un test de t de Student alors que les variables qualitatives d'un test du Chi<sup>2</sup>.

Ceci sera réalisé pour chaque groupe d'auditeurs et les résultats seront comparés entre les groupes.

---

### 3.2.2. DESCRIPTION DE LA PROCEDURE DES TESTS

---

#### 3.2.2.1. PROCEDURE DU TEST DE CATEGORISATION

---

Les stimuli sont présentés sur un écran d'ordinateur, sous forme d'icônes numérotées qui correspondent à un ordre de présentation aléatoire. Le sujet écoute librement les sons proposés et doit ensuite les assembler en les faisant glisser sur l'écran à l'aide d'un pointeur constituant ainsi plusieurs groupes disparates. L'auditeur est libre d'écouter les sons autant de fois qu'il le souhaite. Aucune limite de temps n'est imposée. Un même stimulus ne peut pas se trouver dans plusieurs catégories à la fois et tous les stimuli doivent obligatoirement être classés (même seul). La procédure complète est expliquée dans Gaillard (2009) et dans l'annexe 7.1.2. Après cette phase de création « physique » des classes, il est demandé sans que cela soit obligatoire, de caractériser verbalement les catégories précédemment formées (libre, sans utilisation d'une liste de mots). Le recueil des commentaires se fait par écriture directement sur le clavier de l'ordinateur dans une fenêtre qui apparaît dès lors que la première étape de catégorisation est terminée. La consigne rédigée était la suivante : « *Vous allez écouter une série de stimuli auditifs. Il vous est demandé de les classer par affinité, selon vos propres critères, en une série de plusieurs catégories (une catégorie pouvant contenir un seul stimulus). Le nombre de catégories n'est pas limité par la consigne, et vous êtes maître à la fois du nombre et du contenu de chacune d'elles. Les stimuli sont présentés sous forme de boutons sur un écran d'ordinateur. Vous pouvez librement écouter les sons et/ou déplacer les boutons à l'écran à l'aide de la souris. Vous pouvez écouter chaque son autant de fois que vous le souhaitez et vous n'êtes pas limité dans le temps pour exécuter le test. Une fois que vous les aurez regroupé, vous devrez confirmer votre choix et il vous sera offert la possibilité de définir les critères que vous avez utilisé.* ». A la fin du test, les données sont archivées dans un répertoire spécifique.

Pour les proches, dans la mesure où une majorité d'entre eux n'était pas utilisateur des systèmes informatiques, l'expérimentateur est resté avec la personne pour l'aider dans la manipulation. Les dix premiers stimuli ont systématiquement été écoutés avant de proposer à la personne de les regrouper. Les commentaires ont été recueillis et tapés par l'expérimentateur.

### 3.2.2.2. PROCEDURE DE L'ANALYSE DES STIMULI

---

#### 3.2.2.2.1. ANALYSE PAR UN JURY D'AUDITEURS EXPERTS

Le degré de sévérité du trouble de la production de la parole et les caractéristiques acoustiques qui en découlent sont variables. Afin de les déterminer, les enregistrements ont fait l'objet d'une analyse par un jury d'expert. Ce jury d'experts est constitué de 5 personnes habituées à juger des productions de parole pathologiques (2 phoniatres et 3 orthophonistes travaillant dans le même hôpital et ayant déjà participé à un jury d'écoute). La tâche demandée consistait à écouter autant de fois que nécessaire les 33 enregistrements puis d'attribuer un score à l'aide d'une échelle catégorielle de 0 à 3 (0 : pas d'altération, 3 : altération importante) sur 1) la qualité de la voix, 2) la qualité de la résonance, 3) la prosodie et 4) la qualité de la production phonémique. Enfin, une fois cette tâche terminée, il était demandé de placer sur une échelle visuelle analogique l'icône de l'enregistrement en fonction de l'intelligibilité de la parole d'un côté et de la sévérité de l'atteinte de la parole de l'autre. La consigne était : « écoutez tant que vous voulez. Votre tâche consiste à les classer en terme d'intelligibilité et de sévérité après avoir défini pour chacun, entre 0 (normal) et 3 (atteinte sévère), le degré d'altération les quatre paramètres suivants : la qualité de la voix, la résonance, la prosodie, les altérations phonémiques (ou de l'articulation des phonèmes). L'intelligibilité est définie comme la compréhensibilité du message véhiculé par le signal, alors que la sévérité est défini comme le degré d'altération globale du signal sonore. En pratique, copier-coller le l'icône du son et glisser la copie le long de l'échelle en haut pour la sévérité, en bas pour l'intelligibilité. Après avoir rempli les cases en rapport avec la gradation des trois paramètres préalablement décrits. ». Cette tâche renvoie à une pratique professionnelle hebdomadaire voire quotidienne pour tous les membres de ce jury. L'échelle est de 30 unités, une unité correspondant à la taille de l'icône utilisée pour écouter les sons. 0 correspond à une bonne intelligibilité et une faible sévérité. 30 à la sévérité maximale ou à la plus mauvaise intelligibilité. Pour obtenir une mesure plus précise de l'intelligibilité ciblant l'identification des mots, une tâche de transcription orthographique du corpus a ensuite été demandée aux juges. Le



nombre de mots correctement retranscrits sur le nombre de mot total de phrases divisé par 100 à permis d'établir ce score.

### 3.2.2.2.2..ANALYSE ACOUSTIQUE

Chaque stimuli a été traité par le logiciel EVA de la société SQLAB permettant d'obtenir des mesures de Fréquence fondamentale et de l'intensité du signal ainsi qu'une analyse de sa durée totale et des pauses (pourcentage de temps passé en pause et nombre de pause). Les informations concernant l'analyse effectuée avec ce logiciel figurent en annexe (section 7.1.3).

---

## 3.2.3. RESULTATS

---

### 3.2.3.1. LES LOCUTEURS

---

Les 33 locuteurs se composent de 16 hommes, 17 femmes dont l'âge moyen est de 55 ans avec des extrêmes de 26 à 86 ans et une médiane de 56 ans. Les 13 témoins ont une moyenne d'âge significativement plus faible que celle des 20 patients (respectivement 45 et 61 ans,  $p=0,003$ ). Le groupe des témoins comprend également plus de femmes (10 femmes pour 3 hommes par rapport à 7 femmes pour 13 hommes,  $p=0,04$ ).

Les 20 patients présentaient des pathologies diverses dont la répartition est la suivante :

- Séquelles de traitement d'une tumeur ou d'une malformation des voies aérodigestives (9) dont
  - Séquelles de traitement de cancers (n=8)
  - Séquelles d'une fente faciale (n=1)
- Séquelles de pathologies neurologiques (3) dont
  - Infirmité d'origine cérébrale (n=1)
  - Maladie de Kearns dominé par un syndrome cérébelleux (n=1)
  - Paralysie des paires crâniennes (n=1)
- Pathologies neurologiques dégénératives (8) dont :
  - Sclérose en plaques (n=1)
  - Syndromes extrapyramidaux (n=7) (4 maladies de Parkinson idiopathique, 2 paralysies supranucléaires progressives et 1 démence corticobasale).

Il n'y a pas de différence d'âge et de sexe entre les 2 groupes de pathologie (neurologique ou structurelle).

### 3.2.3.2. LES STIMULI

#### 3.2.3.2.1. L'ANALYSE PERCEPTIVE PAR LE JURY

Les résultats de l'expertise des 5 juges sont présentés sur le tableau n°3.1 pour les scores de sévérité, d'altération de l'intelligibilité et le pourcentage d'identification des mots et sur la Figure 3.1 pour les scores décrivant les caractéristiques perceptives. La reproductibilité interjuge est correcte (entre 0,86 et 0,97). Les coefficients de corrélation intraclasse pour l'ensemble des locuteurs sont satisfaisants, meilleurs pour la sévérité que pour l'altération de l'intelligibilité, respectivement : 0.913, (95% CI: 0.860, 0.952) et 0.832, (95% CI: 0.707, 0.910).

Tous les témoins ont un score « normal » pour l'intelligibilité (à 1 pour l'altération de l'intelligibilité et 100% pour la tâche d'identification de mots) alors que 2 témoins ont un score de sévérité moyen 1,4 et 1,6 en rapport avec une cotation supérieur à 1 pour au moins 2 juges. La distribution de la sévérité couvre mieux l'étendue de l'échelle des 30 points. La corrélation entre les trois scores moyens est très forte ( $r = 0,97$  (IC 0,95 – 0,98)). Une différence significative ( $p < 0,001$ ) est retrouvée entre les témoins et les sujets pathologiques. Aucune différence n'est mise en évidence entre les 2 grands groupes de pathologies (structurale et neurologique ou entre les séquelles de cancer et les syndromes extra pyramidaux).

Population	Sévérité		Altération de l'Intelligibilité		% mots identifiés	
	Moyenne	DS	Moyenne	DS	Moyenne	DS
Témoins	1,09	0,22	1	0	100	0
Pathologies neurologiques	14,38	9,6	11,38	9,55	85,6	27,53
Pathologies Structurales	17,08	7,95	14,66	8,84	90,27	14,28

Tableau n°3.1 : Résultat de l'expertise pour la sévérité et l'altération de l'intelligibilité (échelle entre 0 et 30) et le pourcentage d'identification de mots.

Pour la suite de ce travail, le degré d'altération de l'intelligibilité a été distribué sur 6 niveaux comme suit :

Niveau 1 : score de sévérité et d'altération de l'intelligibilité égal à 1

Niveau 2 : score de sévérité supérieur à 1 alors que le score d'altération de l'intelligibilité égal à 1

Niveau 3 : score de sévérité et d'altération de l'intelligibilité supérieur à 1 et inférieur ou égal à 10

Niveau 4 : score de sévérité et d'altération de l'intelligibilité supérieur à 10 et inférieur à 20.

Niveau 5 : score de sévérité > à 20 alors que le score d'intelligibilité est < à 20

Niveau 6 : score de sévérité et d'altération de l'intelligibilité supérieur à 20

Le pourcentage de mots correctement identifiés n'a pas été utilisé en raison d'un effet plafond lié à l'effet de familiarisation maximal avec les stimuli dans la mesure où il s'agit de la même phrase. Une discussion sera consacrée à ce sujet dans le chapitre 4.

En ce qui concerne les caractéristiques perceptives des stimuli (figure n°3.1), les scores sont en accord avec le TPPP (De Bodt et al 1997). Ils sont plus sévère pour l'aspect phonétique puis par ordre décroissant, la résonance, la prosodie puis la qualité de la voix. Pour la suite de ce travail, ils seront utilisés sous la forme du score maximum attribué par le jury pour chaque caractéristique.

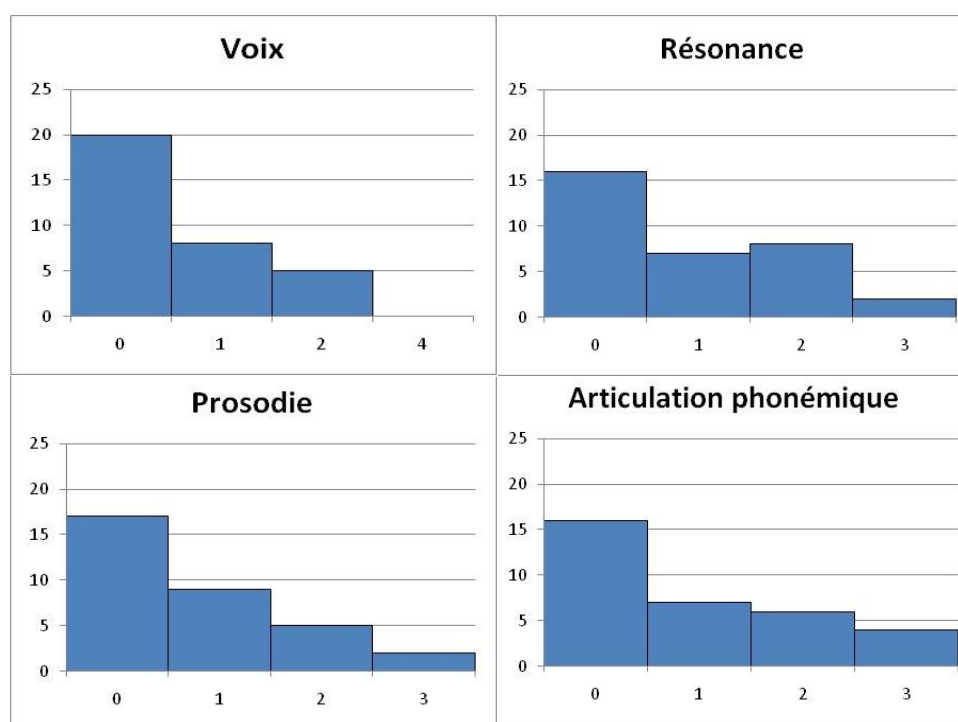


Figure n°3.1 : Distribution des scores maximum attribués par le jury aux différentes caractéristiques des stimuli

### 3.2.3.2.2. L'ANALYSE ACOUSTIQUE DES STIMULI

Les paramètres prosodiques de l'analyse sont décrits sur le tableau n°3.2.

	Moyenne	Ecart type	Minimum	Maximum
Fréquence moyenne Hz	173,82	31,56	113	222
Coefficient de variation de la fréquence %	14,65	5,68	5	31
Etendue en demi tons	10,54	3,7	3,1	17,8
Dynamique en dB	46,38	7,35	30	61
Durée totale	5,57	3,43	3	22,4
Pourcentage de temps en pause	22,6	10,12	6,3	50
Nombre de pause	1,879	1,43	1	6

Tableau n°3.2 : Résultat de l'analyse acoustique par station EVA sur les paramètres prosodiques

Pour permettre la lecture des résultats, les paramètres de Fréquence moyenne, dynamique, Durée totale et Pourcentage de temps en pause ont été distribués en variables ordonnées à 4 niveaux par quartiles.

L'ensemble des résultats est présenté dans l'annexe 7.1.4.

### 3.2.3.3. DETERMINATION DES CLASSES ET CONSTRUCTION DES ARBRES

---

Les arbres de distances représentant les classes «moyennes» construites par l'ensemble des sujets sont présentées sur les figures n° 3.3 à 3.5. Pour connaître et comparer la distance entre deux stimuli, il convient de suivre les arêtes les reliant, sans considérer leur éloignement dans le plan. Pour déterminer les classes, la valeur des arêtes et des nœuds aide à préciser l'impression visuelle globale. Elles ont été laissées sur la figure n°3.5, puis supprimées sur les autres pour une meilleure lisibilité.

L'interprétation se fait à travers les grandes classes (visibles à travers les nœuds principaux au centre de l'arbre) permettant de déduire des similarités entre les membres de chacune de ces classes, et des dissimilarités avec les membres des autres classes. Similarités et dissimilarités pourront être évoquées indifféremment pour comprendre les classes formées par les sujets. Ainsi, pour les experts, (figure n°3.3), 3 nœuds permettent de différencier 3 classes de respectivement 13, 7, 13 locuteurs. Pour les non experts, 3 nœuds permettent de différencier 3 classes de respectivement 5, 6, 13 locuteurs. Neuf locuteurs restent isolés et seront regroupés pour les traitements statistiques. Pour les proches, l'arbre dessine nettement 2 pôles avec à une extrémité 2 classes dont l'une est très consensuelle et comprend 5 locuteurs et l'autre moins consensuelle comprend 9 locuteurs. A l'autre extrémité, une classe de 13 locuteurs avec à l'intérieur des petites classes plus consensuelles. Entre ces 2 pôles, 5 locuteurs sont isolés et seront regroupés également pour le traitement statistique. La légende de ces figures repose sur les codes couleurs suivants :

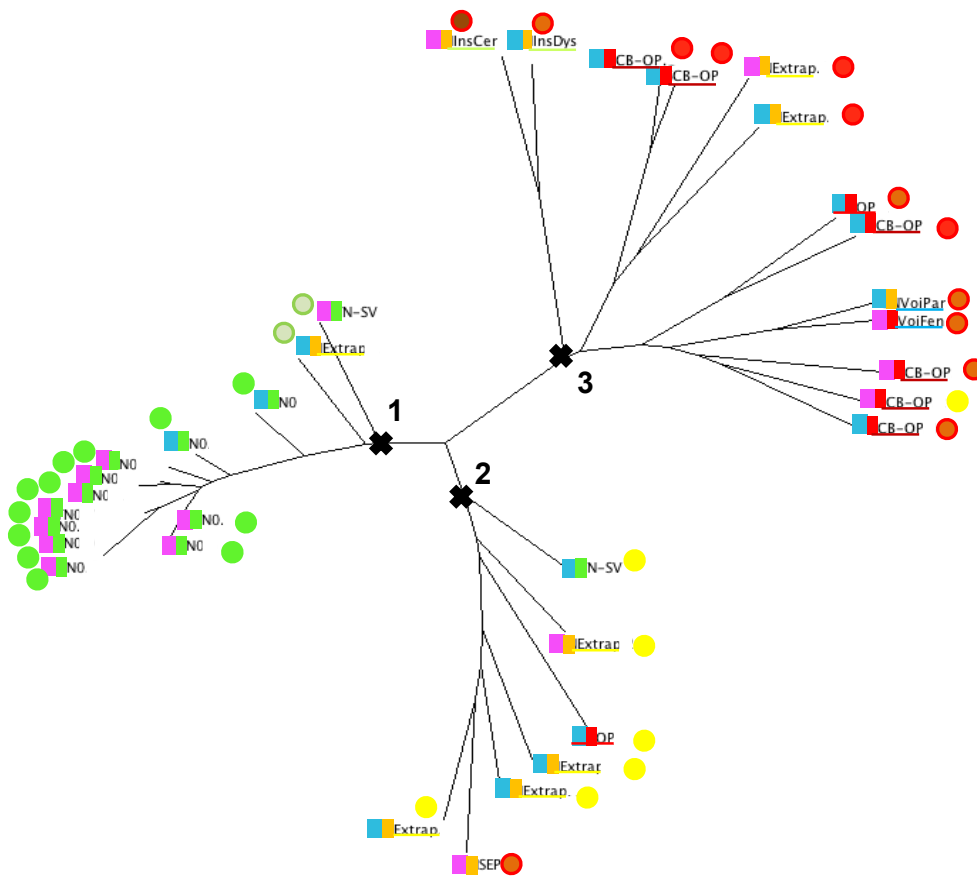
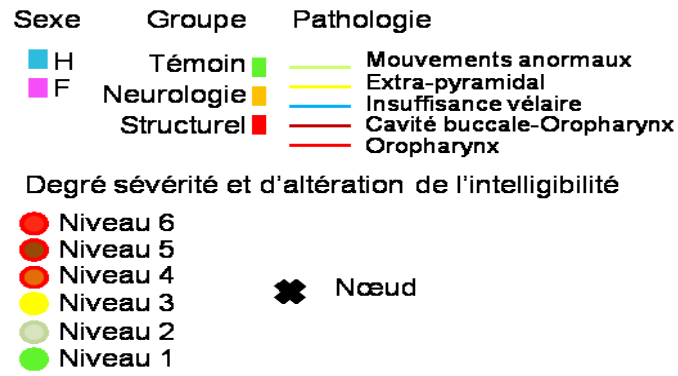


Figure n°3.3 : Arbre réalisé à partir des auditeurs du groupe des experts.

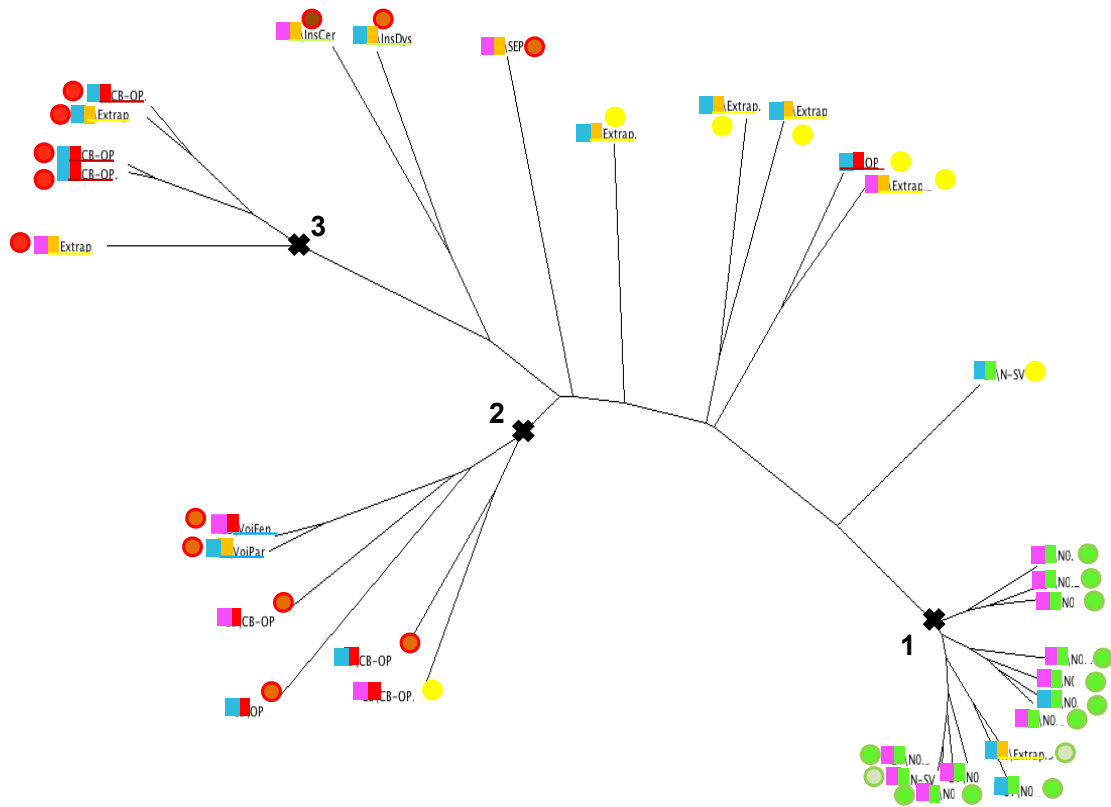


Figure n°3.4. Arbre réalisé à partir des auditeurs du groupe des non experts.

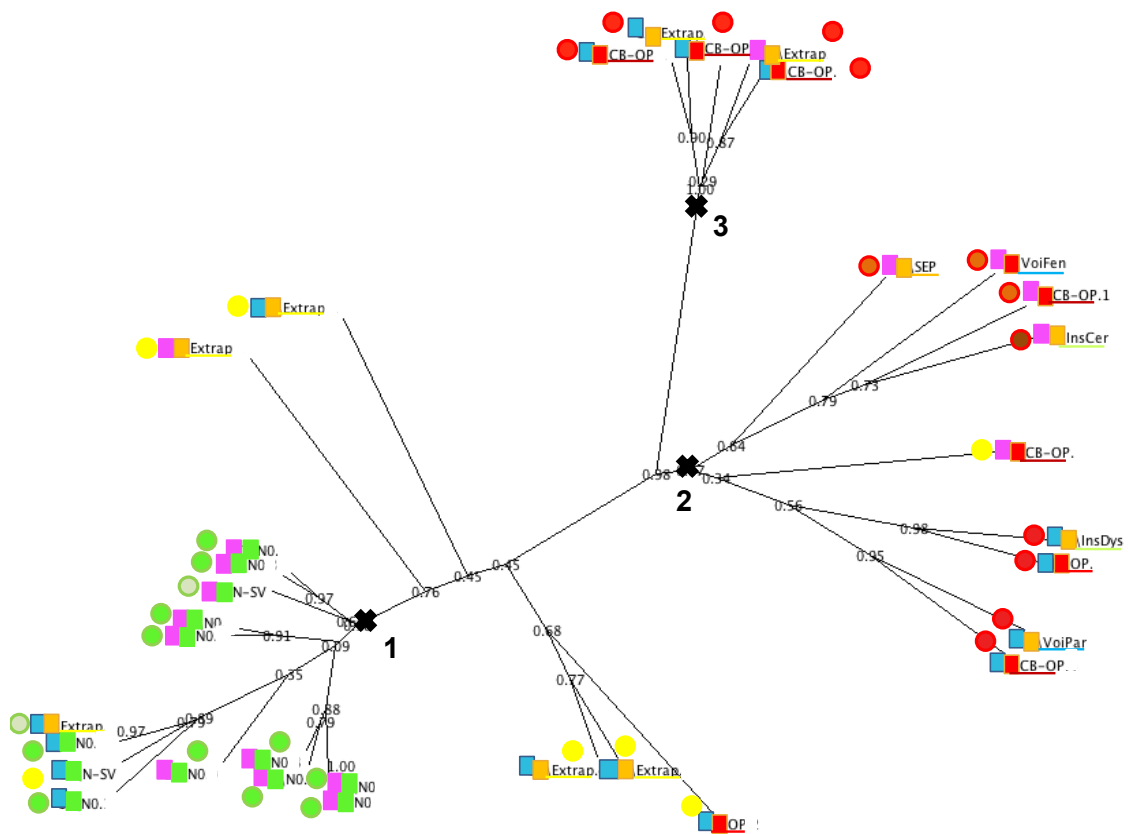


Figure n°3.5. Arbre réalisé à partir des auditeurs du groupe entourage proche du patient.

### 3.2.3.4. ANALYSE DES ARBRES

Les catégories créées par les auditeurs (figure n°3.6) sont en moyenne de 6,72 +/- 3,25 catégories avec un nombre moyen de stimuli dans la catégorie la plus petite (nombre minimum moyen) de 2,35 +/- 1,6 et dans la catégorie la plus grande (nombre maximum moyen) de 11,7 +/- 3,8 stimuli par catégorie. Aucune différence n'a été mise en évidence entre les groupes ( $p=0,5$ ).

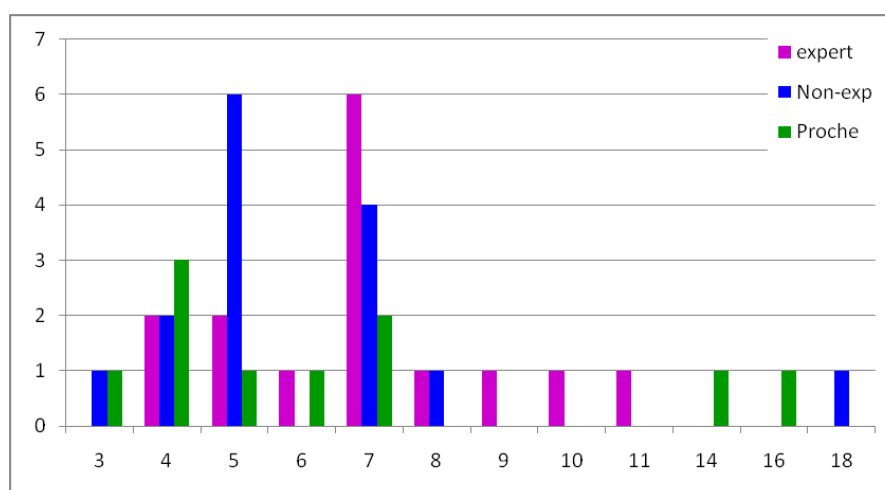


Figure n°3.6 : Distribution du nombre de catégories créées par les auditeurs.

#### 3.2.3.4.1. ANALYSE DES CLASSES PAR RAPPORT AUX CARACTERISTIQUES DES LOCUTEURS ET DES RESULTATS DE L'EXPERTISE DU JURY

##### 3.2.3.4.1.1. LA « LECTURE » DES ARBRES

La lecture des arbres complétée par les scores d'intelligibilité et les groupes pathologies permettent l'analyse globale suivante.

Pour les experts (Figure 3.3), la classe 1 comprend tous les sujets dont l'intelligibilité est « normale ». A noter que les enregistrements pour lesquels certains juges ont attribué un score de sévérité > à 1 sont les plus proches du nœud. Le groupe des voix de femmes est dans cette classe particulièrement bien identifiée. La classe 2 est composée de sujets présentant un trouble de l'intelligibilité plus léger, y compris un témoin qui a fait une pause et une répétition non linguistique se traduisant par un score supérieur à 1 au niveau de l'intelligibilité. Il comprend également l'enregistrement d'une personne présentant avec un trouble plus sévère dans le cadre d'une sclérose en plaque. Les étiologies et les mécanismes des troubles sont divers. La classe 3 comprend les patients dont l'intelligibilité est la plus altérée. On observe à l'intérieur de cette classe des regroupements par type de trouble de la

production de la parole (par exemple instabilité avec des productions instables hachées, scandées dans le cadre d'un syndrome cérébelleux et d'une paralysie cérébrale athétosique (Ins...) ou trouble dominé par une forte rhinolalie dans la cadre d'une fente faciale ou d'une paralysie des paires crâniennes (Voi...).

Pour les non experts (Figure n°3.4), les 3 classes sont encore plus concordantes avec le degré d'atteinte de l'intelligibilité. Dans le groupe des non classés, la disposition suit une progression suivant le degré d'intelligibilité. La classe 1 comprend tous les stimuli pour lesquels le degré d'altération de l'intelligibilité était à 1. La classe 2 comprend des troubles structurels et les 2 insuffisances vélares pour lesquels le trouble de l'intelligibilité est plus marqué mais avec des degrés différents. La classe 3 réunit les enregistrements des 5 patients les moins intelligibles dont les origines sont soit structurelles, soit neurologiques. Dans le groupe des stimuli non classés, on remarque une grande étendue dans le degré d'altération de l'intelligibilité et la présence d'étiologies principalement neurologiques.

Pour les proches (figure n°3.5), les classes sont également très concordantes avec le degré d'atteinte de l'intelligibilité. La classe 1 comprend tous les sujets témoins, quelque soit leur score de sévérité, et le patient chez qui on ne décèle pas sur l'enregistrement le TPPP. Une sous catégorie correspondant cette fois-ci aux voix masculines est nettement individualisée. La classe réunit des stimuli où l'altération de l'intelligibilité est nette, sans que l'on puisse déceler à l'intérieur un autre critère de tri. La classe 3 regroupe comme pour les non experts les enregistrements des 5 patients les moins intelligibles.

A noter que les non experts et les proches n'ont pas classé les mêmes stimuli (les « non classés » sont différents) : les experts ont plutôt regroupés dans une catégorie des voix de femmes alors que les proches ont eux regroupé plutôt des voix d'hommes.

#### *3.2.3.4.1.2. CORRELATION ENTRE LES CLASSES, LE DEGRE DE SEVERITE ET LE DEGRE D'ALTERATION DE L'INTELLIGIBILITE.*

Que ce soit pour les experts, les non experts ou les proches, la différence de moyenne des scores d'altération de l'intelligibilité et de sévérité entre chaque classe est statistiquement significative avec un  $p < 0,001$ . Elle paraît plus « évidente » pour les non experts et les proches que pour les experts. De ce fait, nous avons choisi d'illustrer ce résultat par les figures concernant le score d'altération de l'intelligibilité (figure n°3.7).



Il est remarquable que les groupes des non classés présentent également une différence significative avec les autres classes pour les 2 groupes d'auditeurs concernés. Mais pour les proches, ce score se situe entre les classes 2 et 3 alors que pour les non experts il se situe entre les classes 1 et 2. De plus en intégrant le groupe des non classés, l'altération de l'intelligibilité devient un critère particulièrement saillant pour les proches.

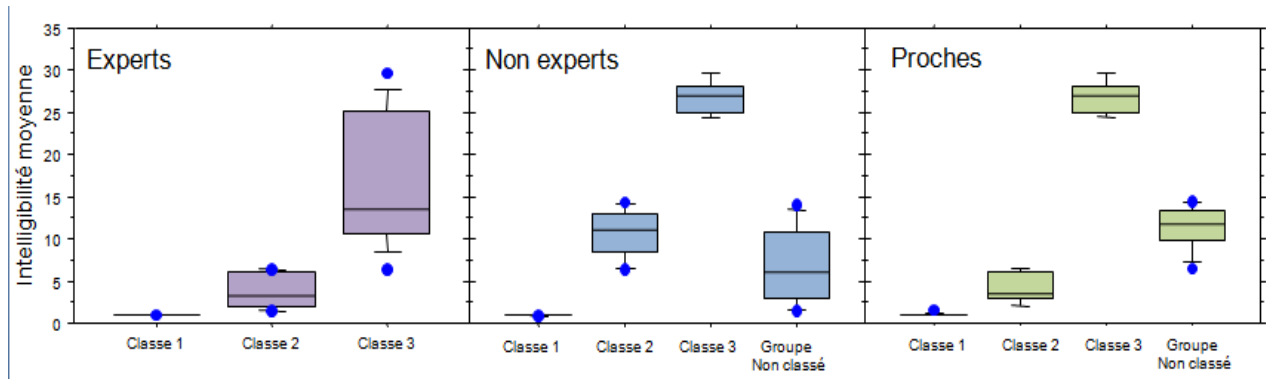


Figure n°3.7 : Graphique en boîte des scores d'intelligibilité en fonction des classes créés par les experts (à gauche) et les non experts (au milieu) et les proches à droite).

### 3.2.3.4.1.3. RELATION ENTRE LES CLASSES, ET LES CARACTERISTIQUES DES STIMULI EN TERME D'ETIOLOGIE ET DE TYPE DE TROUBLE ET D'IDENTITE DES LOCUTEURS

En ce qui concerne l'identité et la pathologie des locuteurs, une différence significative est retrouvée pour l'âge entre les classes 1 et 3 ( $p=0,02$ ) mais cette différence est associée à la présence d'une pathologie dans la mesure où les témoins sont plus jeunes. Il n'y pas de différence quant à la partition des genres. Les grandes catégories de pathologies sont disséminées dans les différentes classes (figure n°3.8). La différence de statut du sujet témoin pour lequel le score de sévérité est supérieur à 1 est remarquable. Dans la classe 2 des experts (altération modérée), le stimulus de ce sujet est classé avec les stimuli « normaux » par les proches et non classé par les non experts. Deux hypothèses non exclusives peuvent être émises : 1) les experts ont un niveau d'analyse qui se focalise sur le signal, et augmente la discrimination par rapport à un signal « parfait », 2) les proches se focalisent plus sur le locuteur en tant que personne et peuvent du fait de leur contact avec une personne présentant un TPPP être plus tolérant aux imperfections du signal.

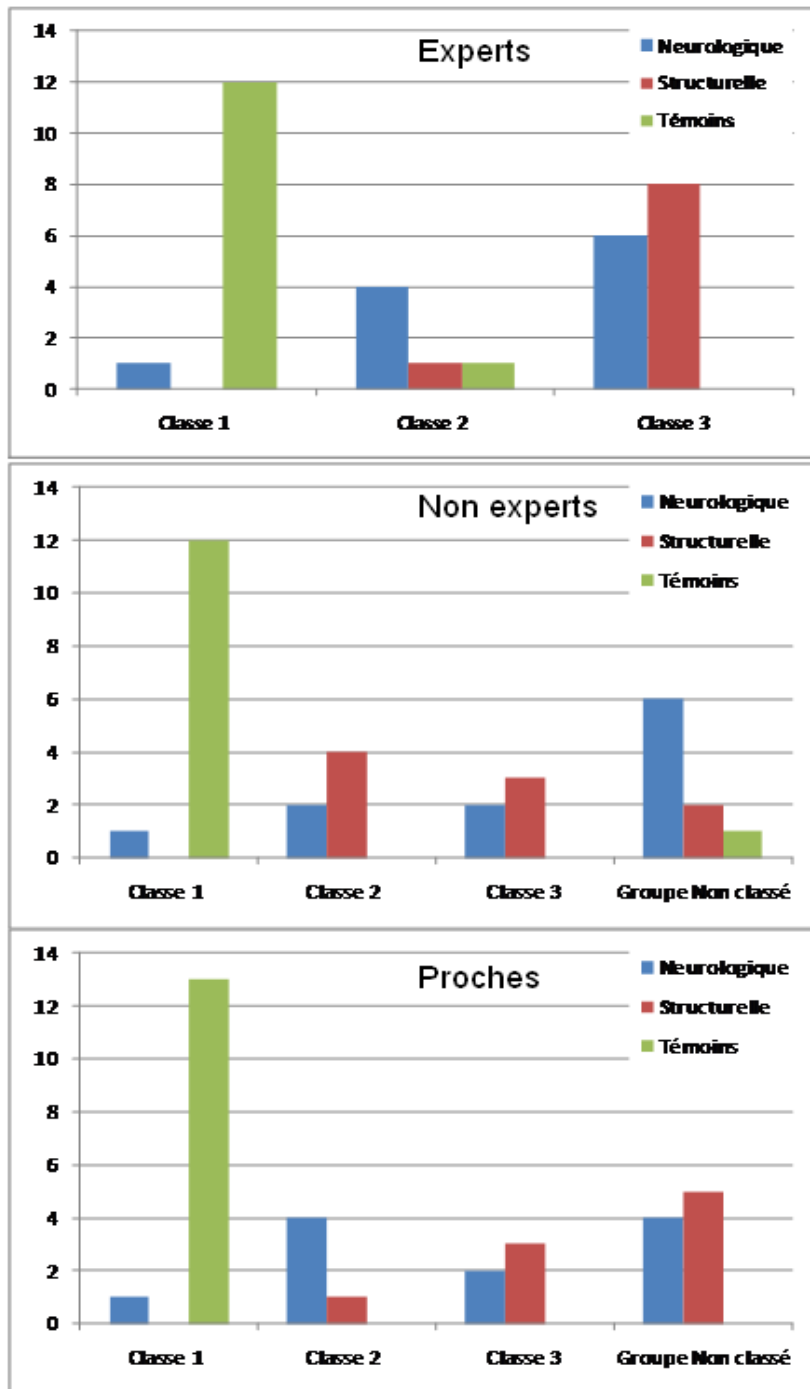


Figure n°3.8 : Histogramme des différents types (témoins, étiologie neurologique, étiologie structurelle) de stimuli en fonction des classes créés par les experts les non experts et les proches.

L'analyse plus détaillée des étiologies intégrant la notion de mécanismes confirme une stratégie différente par rapport aux stimuli des locuteurs témoins présentant des particularités (figure n°3.9). De plus, les différentes catégories d'auditeurs ont tous regroupé les TPPP avec des mécanismes très proches quelque soit leur étiologie comme les 2 stimuli marqués par une forte instabilité ou les insuffisances vélares. Le groupe des experts est celui qui regroupe le mieux les étiologies, ce qui

répond au développement de leur compétence professionnelle. A noter que les non experts ne font pas mieux que les proches.

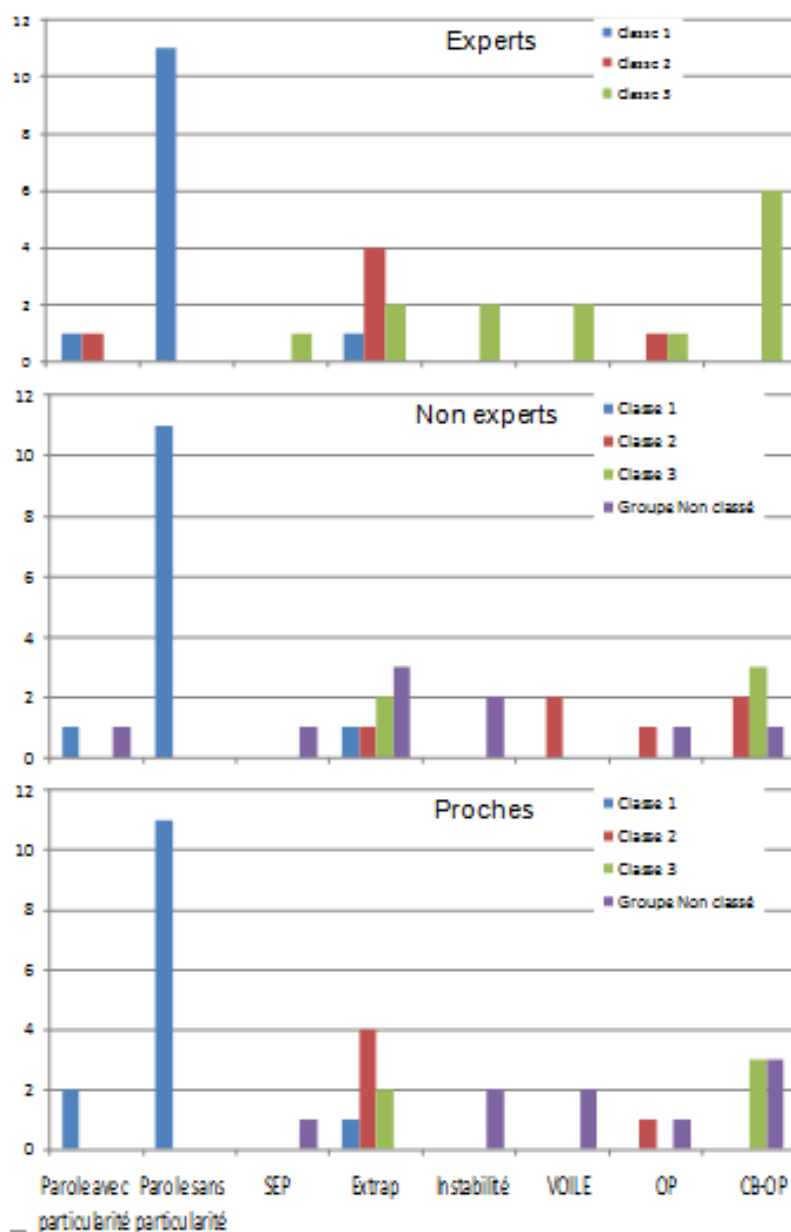


Figure n°3.9 : Histogramme des étiologies détaillées de stimuli en fonction des classes créés par les experts (en haut) et les non experts (au milieu) et les proches (en bas).

En intégrant les résultats précédents, il semble que pour les non experts et les proches, les personnes très peu intelligibles forme une catégorie bien à part, et que pour les experts d'autres critères rentrent en jeu. Il ne semble pas s'agir de l'étiologie mais plutôt du type de trouble à travers son mécanisme physiopathologique. Par contre, les classes créées par les « proches » ont naturellement tendance à intégrer ce paramètre.

#### 3.2.3.4.1.4. ANALYSE DES CLASSES PAR RAPPORT AUX CARACTERISTIQUES DE L'ANALYSE PERCEPTIVE ISSUES DE L'EXPERTISE DU JURY

Les scores des paramètres perceptifs (Figure 3.10) sont plus élevés pour les classes 3 par rapport aux classes 1. Ceci est particulièrement net pour les paramètres « articulation phonémique » et « résonance » quelque soit le type d'auditeurs. Les non experts et les proches semblent plus utiliser le critère de qualité de la voix.

La classe 1 pour toutes les catégories d'auditeurs ne comprend que des grades 0 pour tous les paramètres. De même, pour toutes les catégories d'auditeurs et tous les paramètres, les grades 3 sont présents exclusivement dans la classe 3.

La classe 2 des experts et des proches est constituée de grade 0 et 1 pour tous les paramètres alors que pour les non experts la composition de la classe 2 est hétéroclite comme le groupe non classé.

Il apparaît que les difficultés de consensus sont distribuées de manière différente sur les paramètres perceptifs entre les non experts et les proches. Les proches n'ont pas classé les stimuli qui ont des scores perceptifs de niveau 1 et 2 alors que les non experts n'ont pas classé ceux qui ont un score perceptif plus bas de niveaux 0 et 1. Cela est-il lié à un seuil de perception de la normalité différent ? Une hypothèse serait que les proches toléreraient plus « d'imperfections » du fait de leur familiarité avec des personnes présentant un TPPP.

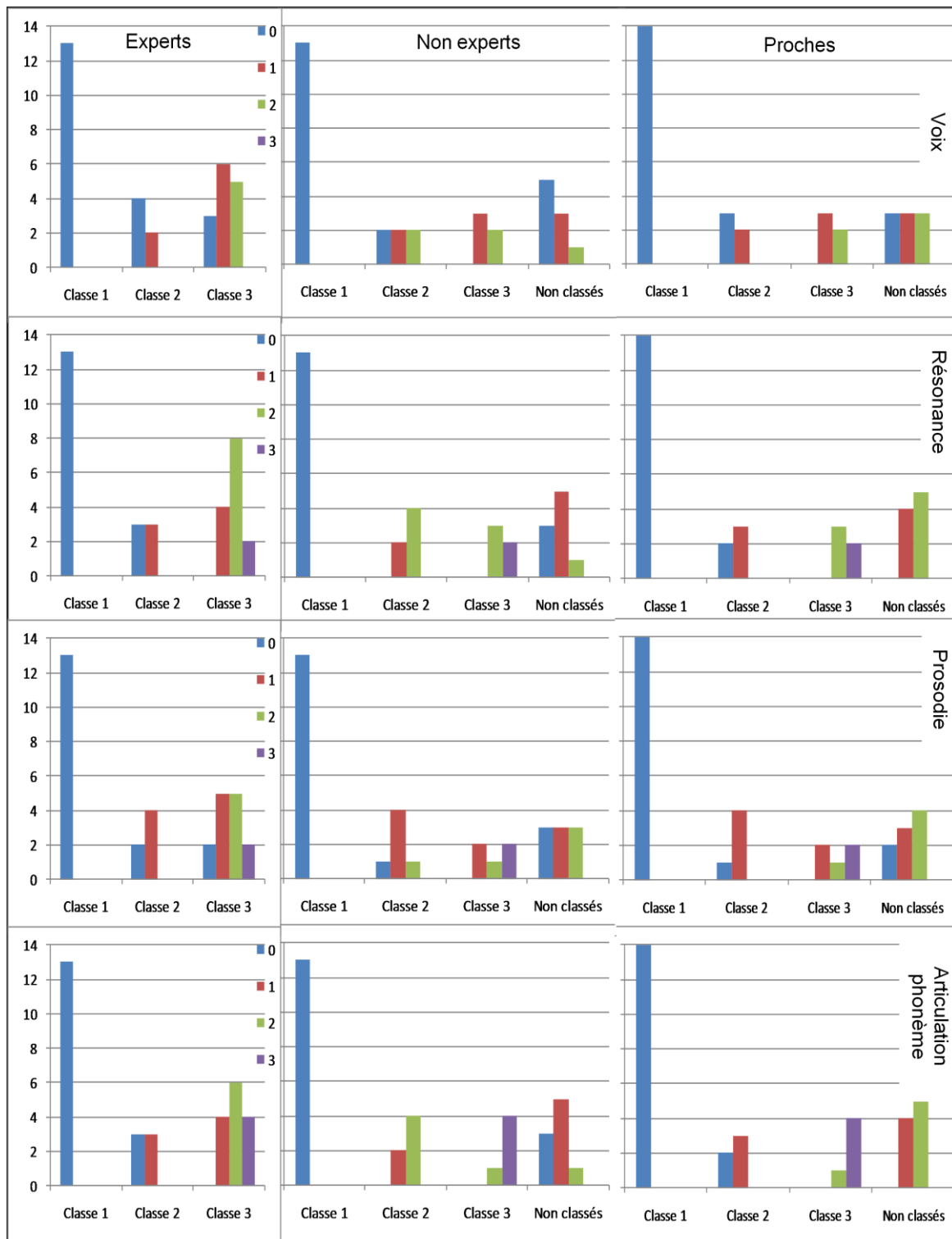


Figure n°3.10 :Scores perceptifs en fonction des classes créées par les experts

### 3.2.3.4.2. ANALYSE DES CLASSES PAR RAPPORT AUX CARACTERISTIQUES DES RESULTATS DE L'ANALYSE ACOUSTIQUE DE LA PROSODIE PAR LE LOGICIEL EVA

Les arbres annotés avec les résultats de l'analyse par la mesure en quartile des paramètres acoustiques sont présentés dans l'annexe 7.1.4.

Nous avons opté pour un codage chiffré renforcé par un codage couleur tel qu'il figure sur la légende des arbres. Malgré cela, leur lecture est difficile. Plusieurs raisons d'ordre cognitive peuvent être évoquées. La principale est que l'usage des quartiles procure une image gradant la croissance du paramètre alors que zone de « normalité » peut être dans le quartile bas ou dans la médiane. Nous allons donc analyser les résultats en privilégiant l'étude des paramètres en fonction des classes sans utiliser la lecture globale des arbres.

Par rapport à l'analyse de la prosodie (figure n°3.11, n°3.12 & n°3.13), les résultats sont superposables pour les non experts et les proches sauf pour les paramètres de dispersion de l'intensité où la déviation standard et la dynamique de l'intensité varient de manière significative pour les non experts et pas pour les proches (figure n°3.12).

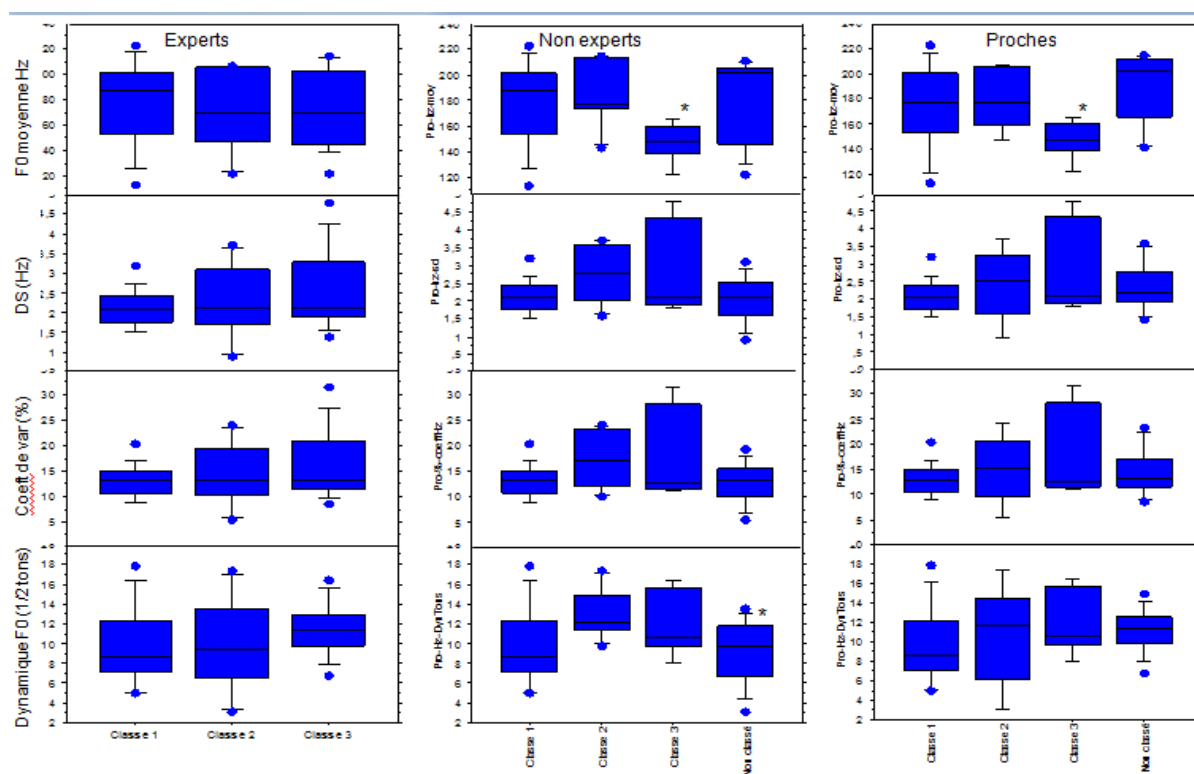


Figure n°3.11 : Graphiques à moustache des paramètres prosodiques de fréquence (\* = p <0,05 entre la boîte marquée par l'astérisque et la classe 1)

Alors que les paramètres de fréquence ne semblent pas être un critère intervenant dans la détermination des classes, la F0 moyenne paraît contribuer au tri pour les non experts et les proches (figure n°3.11). Une différence entre la classe 2 et la classe 3 démontre que s'il n'y a pas d'impact décelé du genre pour ces groupes d'auditeurs, la fréquence grave de la voix est un critère qui caractérise la classe 3. L'hypothèse pourrait être ici, l'impact de l'identité du locuteur dont ces groupes contrairement aux experts ne se sont pas affranchis

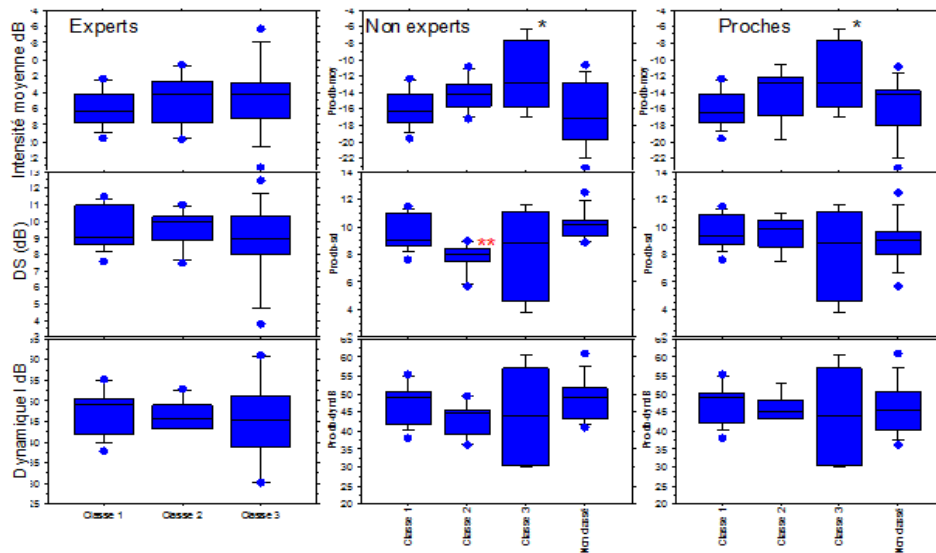


Figure n°3.12 : Graphiques à moustache des paramètres prosodiques d'intensité (\* =  $p < 0,05$ , \*\* $p < 0,01$  entre la boîte marquée par l'asterisque et la classe 1 ou le groupe non classé

. Il en est de même pour les paramètres d'intensité mais ici, seuls les non experts semblent utiliser ce critère.

Par contre, les paramètres de durée et les pauses sont des paramètres associés aux classes avec une différence significative de distribution pour la classe 3 pour les experts, et entre les différentes classes pour les non experts et les proches (figure n°3.13).

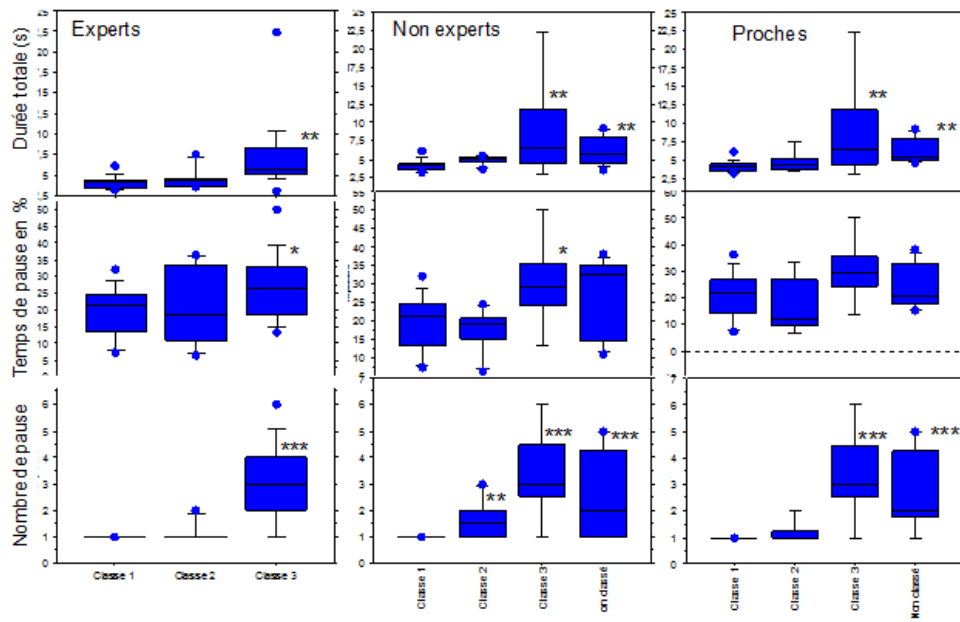


Figure n°3.13 : Graphiques à moustache des paramètres prosodiques de durée et pause (\* =  $p < 0,05$ , \*\* $< 0,01$ , \*\*\* $< 0,001$ ) entre la boîte marquée par l'astérisque et la classe 1)

Cet effet est particulièrement net chez les non experts sur la durée totale du signal traduisant la notion du débit de parole et le nombre de pause alors que la durée ou le pourcentage de temps relatifs aux pauses interviennent peu.

### 3.2.3.5. ANALYSE DES COMMENTAIRES

Rappelons que les sujets ont eu à commenter post-catégorisation les classes qu'ils avaient formés. L'ensemble de ces commentaires est intégralement rapporté dans les tableaux figurant dans l'annexe n°7.1.4.

#### 3.2.3.5.1. GRILLE D'ANALYSE

La lecture des commentaires a permis de construire la grille d'analyse suivante hiérarchisée en thèmes principaux et thèmes secondaires présentée sur le tableau n°3.3.

A noter que l'utilisation exclusive du terme « sévérité » par les experts et l'indexation connue de la sévérité par l'intelligibilité justifient que nous ayons associé ce mot au repérage de la notion d'intelligibilité tout en le traçant pour l'analyser.



<u>Intelligibilité</u>	
• Repérage des notions d'intelligibilité, de compréhension et de sévérité (ou gravité)	Intelligibilité, compréhension, sévérité
• Présence de mots ou de segments faisant référence à l'intelligibilité ou à la notion de compréhension	Absent /présent
<u>Normalité</u>	
• Présence de mots ou de segments faisant référence à des locuteurs ou des productions de parole sans anomalie	Absent /présent
<u>Référence au support linguistique (pour les auditeurs)</u>	
• Niveau principal : Voix, voix-parole, parole	Voix, voix-parole, parole
• Les différents composants : voix, prosodie, articulation (rhinolalie), langage, lecture	Toutes les combinaisons
<u>Supposition par rapport au locuteur</u>	
• Par rapport : à son identité (âge et/ou sexe), à son état de santé (maladie, trouble, déficit, noms de maladie, patients, asymptomatique, dysphonie, dysarthrique...), à son état émotionnel, à des traits de personnalité	Absent/ émotion/ identité-santé/ trait de personnalité
<u>Référence au trouble de la production pathologique de la parole</u>	
• En rapport avec l'étiologie du trouble de la production de la parole, avec le mécanisme du trouble de la parole	Absent/étiologie/mécanisme
<u>Sentiments ressentis par le locuteur à l'écoute du stimulus</u>	
	Aucun/effort/ agréable-désagréable...
<u>Estimation sur les critères de tri (pour les auditeurs)</u>	
• Estimation du critère principal	Intelligibilité/mécanisme/ étiologie/ qualité articulation/ qualité voix/ expressivité...
• Estimation de l'ordre d'importance du critère d'intelligibilité par rapport aux autres critères	Absent, 1 <sup>er</sup> , 2 <sup>nd</sup> ..., Accessoire
• Autres critères	liste
• Essai identification d'une stratégie de catégorisation	liste
<u>Précisions par rapport au support (pour les auditeurs)</u>	
• Si voix : intensité fréquence timbre	liste
• Si prosodie	liste
• Si articulation	liste

Tableau n°.3.3 : Grille d'analyse des commentaires

Un échantillon des résultats « bruts » de l'analyse sont présentés dans l'annexe 7.1.4.

### 3.2.3.5.2. REFERENCES A L'INTELLIGIBILITE ET A LA NORMALITE

En ce qui concerne la notion d'intelligibilité (tableau n°3.4 et figure n°3.5), le repérage de la notion d'intelligibilité, de compréhension et de sévérité révèle que la notion de « compréhensibilité » est largement prépondérante (28/40 auditeurs) et que cette dernière notion est une préoccupation des proches (8/10), puis des non experts (11/15) puis des experts (9/15). En se focalisant sur l'utilisation des mots, l'intelligibilité est le terme le plus utilisé par les experts alors que la compréhension appartient plus au vocabulaire des proches et des non experts. L'estimation de l'ordre d'importance de ce critère parmi les clés de tri repérées révèle une priorisation plus importante de celui-ci même si il est moins fréquent pour les experts par rapport aux non experts.

	Expert	Non Expert	Proches
Repérage des notions d'intelligibilité, de compréhension et de sévérité (ou gravité)			
absent	6	4	2
compréhension	2	7	7
compréhension et intelligibilité	0	2	1
intelligibilité	5	2	0
intelligibilité et sévérité	1	0	0
sévérité	2	0	0
Estimation de l'ordre d'importance du critère d'intelligibilité par rapport aux autres critères			
premier	4	2	5
deuxième	1	2	1
troisième	4	3	0
quatrième	0	3	0
accessoire	0	1	2

Tableau 3.4 : La notion d'intelligibilité ou de compréhensibilité dans les commentaires

La référence au caractère normal est également très fréquente (28/40). Elle est systématique chez les experts, alors qu'elle n'est présente que dans 6/15 commentaires pour les non experts. Elle est relativement plus fréquente pour les proches (7/10).

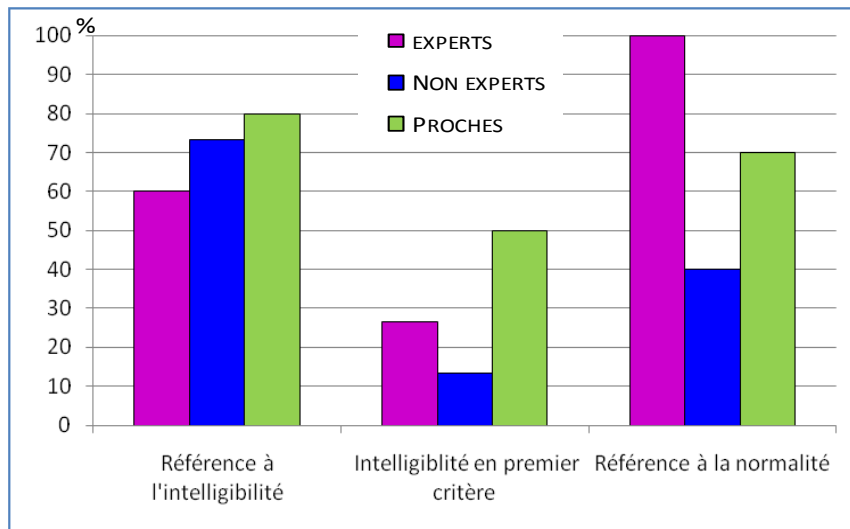


Figure n°3.5. Histogramme (en pourcentage) de la fréquence de la référence à la notion d'intelligibilité, de la priorisation de ce critère par les 3 groupes d'auditeurs et de la référence à la normalité.

A retenir : le critère d'intelligibilité est prégnant dans les trois groupes mais avec une fréquence croissante de la manière suivante : proches > non experts > experts.

Par contre, la référence à la normalité semble importante pour les experts. Peut être est ce que leur préoccupation professionnelle les incitent à détecter ce qui peut ou doit être « corrigé » par rapport à la pathologie, ce qui renforce leur conception de la normalité.

Les non experts y font moins référence que les proches. Ceci ne semble pas un problème de détection puisque le seuil semble augmenté (grade 0 et 1 versus 1 et 2). Peut être n'ont-ils pas conscience de cette augmentation de seuil mais qu'ils ont un besoin de référence par rapport à la normale du fait de leur proximité avec des personnes atteintes d'un TPPP ?

### 3.2.3.5.3. LE SUPPORT LINGUISTIQUE DE REFERENCE

Le niveau linguistique qui semble soutenir principalement la catégorisation est dominé par la notion de parole notée 30 fois sur 40. Cette tendance est nette pour les experts alors que les non experts font référence de manière équivalente à la voix et/ou à la parole et que les proches font principalement référence à la voix et la parole (Figure n°3.5). Le niveau prosodique n'est jamais apparu isolément.

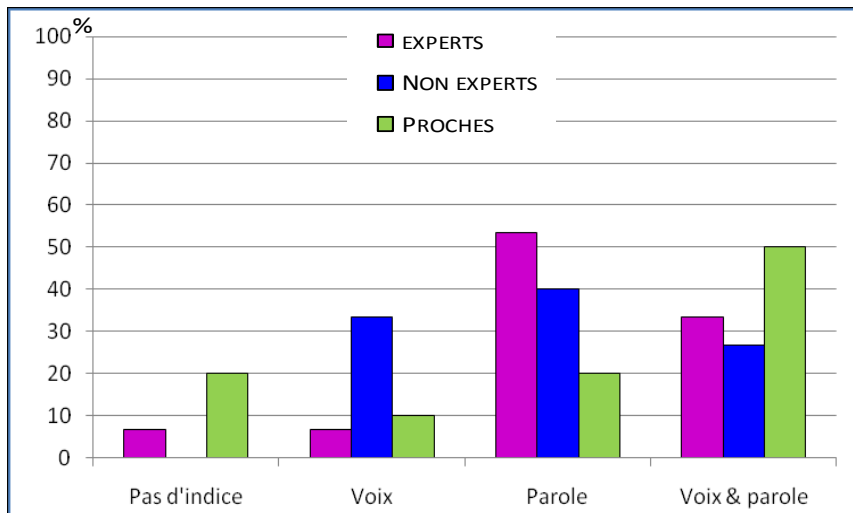


Figure n°3.6 : Histogramme en pourcentage de référence au support linguistique principal

Il est remarquable que les termes faisant référence à la prosodie ou la nasalité sont très rarement utilisés par les proches. Ils semblent répondre à une culture des professionnels de santé. Les non experts font plus souvent référence à la voix (figure n°3.7).

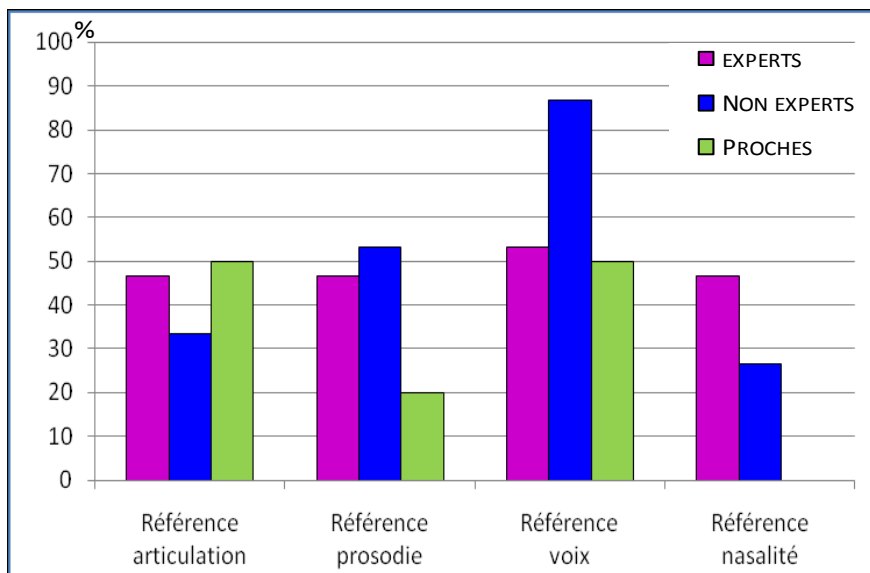


Figure n°3.7 : Histogramme (en pourcentage) de la répartition de toutes les références au support linguistique

A noter que des précisions sont explicitées dans chaque catégorie par les experts et les non experts. Elles sont très rares pour les proches et ne concernent que le débit de la parole. Il est remarquable que les précisions pour la voix sont dominées par le timbre (8/15 pour les experts et 12/15 pour les non experts) et rare sur l'intensité (respectivement 4/15 et 3/15) et quasi absent sur la fréquence (1/15 uniquement pour un expert). A noter des références décroissantes : timbre > intensité > fréquence

alors que la fréquence apparaît comme facteur de tri à la lecture des arbres. La fréquence apparaît comme un facteur implicite.

Cette différence de référence entre la voix et la parole est-elle liée à un concept indifférencié relevant plus de l'usage du terme « voix » ou est ce vraiment une sensibilité plus élevée à ce que véhicule la voix (identité) que la parole (message) ?

#### 3.2.3.5.4. SUPPOSITION PAR RAPPORT AU LOCUTEUR ET REFERENCE AU TPPP

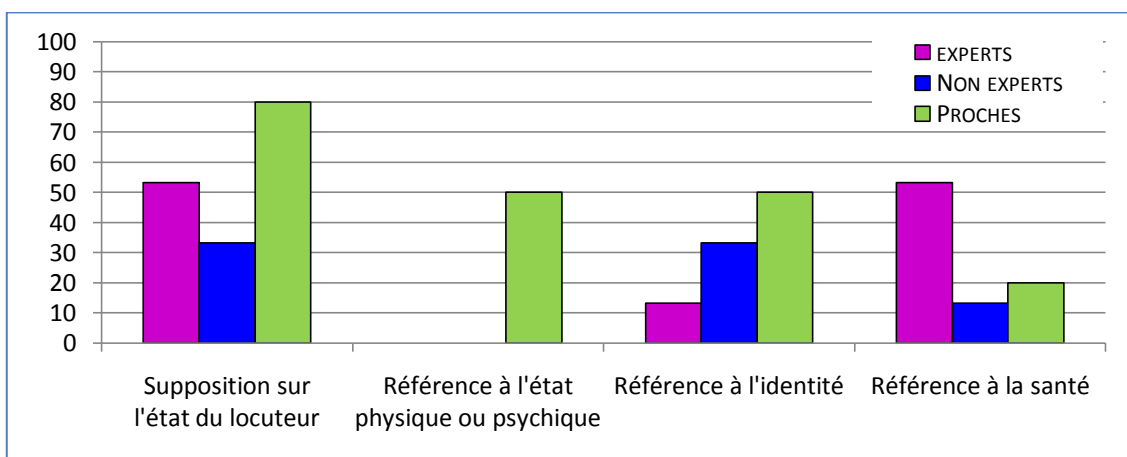


Figure n°3.8 : Histogramme en pourcentage des suppositions émises par rapport au locuteur

Les commentaires sont riches en suppositions émises sur l'état du locuteur au moment de l'enregistrement (figure n°3.8). Cela concerne particulièrement les proches qui établissent des inférences sur l'état physique (« fatigué » par exemple) ou psychique (« manque de pêche »...) en plus des références à l'âge et au sexe, des traits de caractère (« personne calme »...). Les experts font eux exceptionnellement ce type d'inférences et se limitent à des suppositions sur l'état de santé. Entre les deux, les non experts qui sont moins familiarisés avec les pathologies, font beaucoup plus de suppositions sur l'identité en termes de genre mais surtout d'âge. Il semble que le fait d'être des professionnels de santé impose un filtrage de ce type d'interprétation.

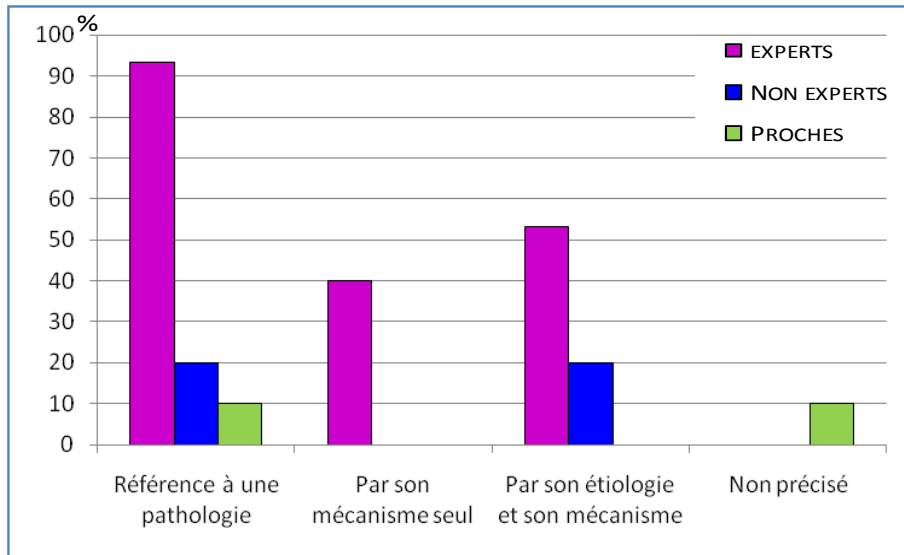


Figure n°3.9 : Histogramme en pourcentage des références à la pathologie

De plus, l'évocation des pathologies en rapport avec le TPPP est le fait des professionnels de la santé. Seuls les experts ont fait référence, de manière isolée au mécanisme du trouble en faisant « abstraction » de l'étiologie de celui-ci (figure n°3.9). Bien que « logique », ceci est surprenant quand on rappelle que tous les proches ayant participé à cette expérience accompagnaient une personne malade présentant un TPPP pour un bilan.

### 3.2.3.5.5. SENTIMENTS « RESSENTIS » PAR LE LOCUTEUR A L'ECOUTE DU STIMULUS

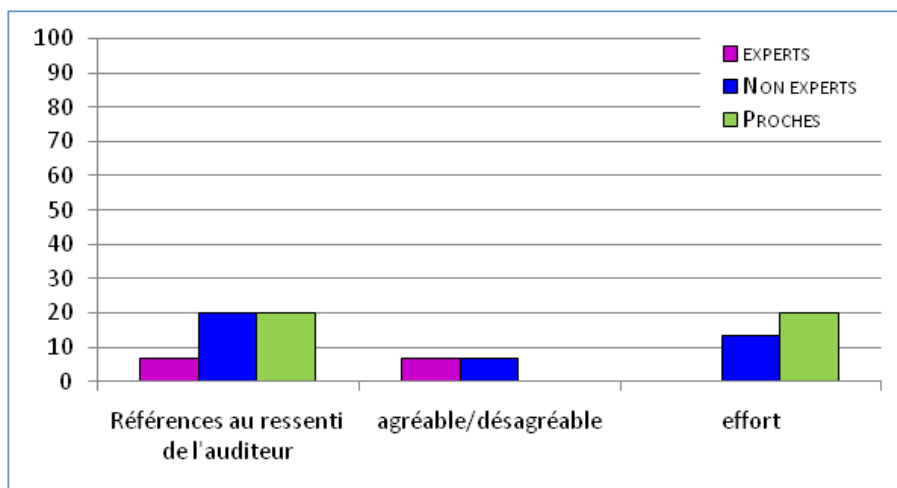


Figure n°3.10 : Histogramme (en pourcentage) des références aux sentiments ressentis par les auditeurs

Il est rare, mais cela a été observé dans les 3 groupes, que les auditeurs expriment ce que l'écoute des stimuli provoque sur eux principalement en terme « d'agréable ou désagréable » ou en terme d'effort que demande l'écoute. Là encore, un filtre s'impose probablement aux professionnels.

### 3.2.3.5.6. CRITERE PRINCIPAL ET ANALYSE DE LA STRATEGIE DES AUDITEURS

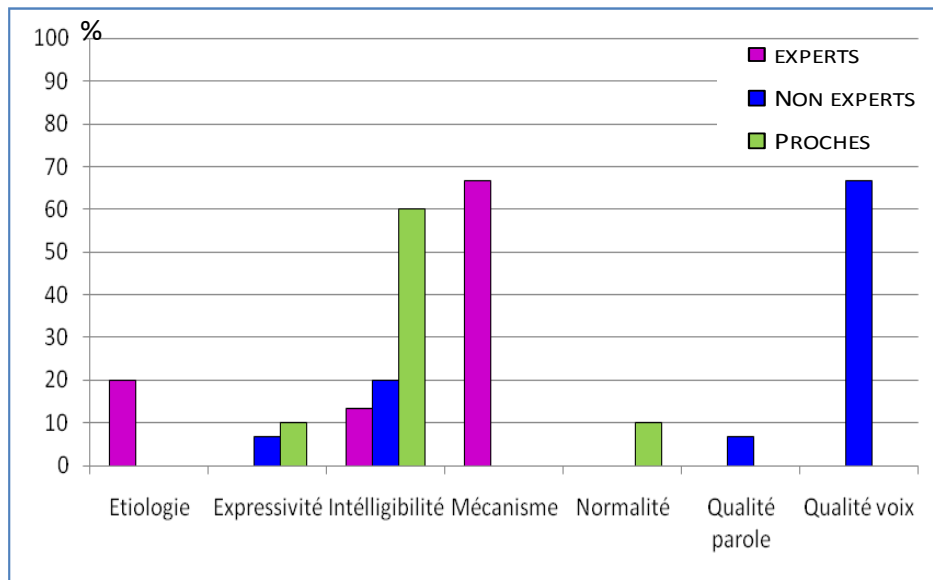


Figure 3.11 L'estimation du critère principal utilisé par les auditeurs

L'estimation du critère principal utilisé par les auditeurs (Figure n°3.11) révèle que les 3 groupes ont des références différentes : le groupe des experts par rapport au trouble de la production, le groupe des non experts par rapport à la qualité de la voix et les proches sur l'intelligibilité.

De nouveau, la question de la raison pour laquelle les non experts utilisent comme critère principal la qualité de la voix soulève la responsabilité de la terminologie. Est-ce un amalgame voix/parole mais le fait qu'ils utilisent des termes décrivant la voix n'est pas en faveur. Ou est-ce lié au début de leur formation professionnelle. Et pourquoi ne se réfèrent-ils pas comme les proches à l'intelligibilité ? Est-ce en rapport avec la confrontation quotidienne à des personnes avec un TPPP pour les proches ?

Il se dégage de cette analyse des commentaires des stratégies de classification (tableau n°3.5) :

- en fonction du degré d'altération de la production ou de l'intelligibilité pour 6 proches/10 et des suppositions faites sur le locuteur pour 3/10,

- en fonction de la description des troubles pour 8/15 auditeurs non experts avec une analyse du degré d'intelligibilité pour également 9 d'entre eux,

- en fonction du mécanisme et/ou de l'étiologie pour 10/15 pour les experts.

Étiquettes de lignes	Experts	Non-experts	Proches
Causale/étiologie	2		
Causale/étiologie et mécanisme	4		
Causale/mécanisme	4		
Degré d'altération	1	1	4
Degré d'intelligibilité		4	2
Degré d'intelligibilité/mécanisme	2		
Description des troubles	1	1	
Description des troubles de la voix		3	1
Description des troubles de la voix/ articulation/intelligibilité	1	5	
Expressivité/description des troubles		1	
Locuteur			3

Tableau 3.5 : Les critères et l'estimation de la stratégie des auditeurs

Ainsi, la stratégie exprimée dans les commentaires paraît renvoyer aux préoccupations probablement spécifiques à chaque groupe : une causalité permettant la mise en place d'une prise en charge pour les experts, une description des troubles souvent en référence à une normalité pour les non experts avec une part importante pour l'intelligibilité, une référence à la « compréhensibilité » en prenant en compte le locuteur en tant que personne pour les proches.

### 3.2.3.5.6. ANALYSE DES CLASSES EN FONCTION DES COMMENTAIRES

#### 3.2.3.5.6.1. POUR LES EXPERTS

Les trois classes peuvent être décrites de la manière suivante :

Une classe (1) de sujets plus ou moins normaux dont 2 présentent à la lecture des commentaires plus de 3 discrètes « anomalies ». Il n'est jamais question d'intelligibilité.

Une classe (3) qui comprend des personnes qui ont une altération sévère avec une atteinte de l'intelligibilité mais dont le degré est très variable ce qui n'en fait pas le critère dominant ni un critère exclusif. Dans les commentaires comme dans l'analyse des classes, le mécanisme du trouble le plus souvent cité est au niveau de l'articulation. Une logique de mécanisme ou de type de trouble semble ensuite conditionner les regroupements. Ceci est également étayé par les différences observées uniquement sur les paramètres temporels concernant la durée et surtout les pauses.

Une classe (2) où l'altération de l'intelligibilité n'est pas systématique et où les altérations sont légères sur le plan perceptif et sont principalement décrites au niveau vocal ou prosodique.



L'hypothèse émise peut être résumée ainsi : Normal ou presque / Pathologie marquée/ Pathologie légère.

#### 3.2.3.5.6.2. *POUR LES NON EXPERTS*

L'analyse des commentaires permet de décrire les classes de la façon suivante.

Une classe (3) déterminée par les difficultés importantes de compréhension traduites souvent par l'emploi de mots qualifiant une perte de «compréhensibilité» ou «d'intelligibilité» confirmée par les scores perceptifs d'intelligibilité et de sévérité.

Une classe (2) où les difficultés de compréhension du message sont moins marquées mais où la parole est altérée. A noter des différences dans les commentaires qui ne permettent pas d'expliquer certaines similitudes entre locuteurs que laisse supposer la construction de l'arbre.

Une classe (1) sans référence à l'intelligibilité, sans description de difficultés de prononciation ou d'articulation, avec des jugements sur la qualité de la voix, de l'intonation ou de l'expressivité correspondant à des scores d'intelligibilité normaux.

Les commentaires pour les non classés sont hétéroclites, il peut être question de trouble de l'intelligibilité mais les altérations de la parole sont signalées au niveau de la « prononciation » ou de « l'articulation », de « l'intonation » et de la voix.

L'hypothèse émise peut être résumée ainsi : Incompréhensible ou anormale mais sans critère par rapport à la normale relative à la voix, plutôt sous jacente la question normal ou pas normal ?

#### 3.2.3.5.6.1. *POUR LES PROCHES*

Les commentaires relatifs aux stimuli de la classe (3) sont totalement cohérents avec l'altération de l'intelligibilité traduite en termes de « incompréhension » ou « inaudible ».

La classe 1 est composé de stimuli « sans particularité » pour lesquels les références à l'identité des locuteurs en terme de genre et d'âge sont nombreuses.

Dans la classe 2, l'altération de l'intelligibilité est très souvent relevée dans les commentaires associée à des suppositions sur l'état du locuteur.

Le groupe des non classés fait souvent référence à la précision d'une intelligibilité correcte avec des particularités de « débit » de parole ou de « voix différentes ».

En résumé ici : Incompréhensibilité/Normal identité/ Normal mais atypique/ anormal pourquoi ?

---

#### 3.2.4. EN RESUME

---

L'analyse de l'arbre en fonction des caractéristiques des stimuli montre que les personnes très peu intelligibles forme une catégorie bien à part, mais pour les experts, d'autres critères semblent rentrer en jeu, notamment le type de trouble par son mécanisme. L'analyse des commentaires révèle un critère spécifique aux experts : la notion d'étiologie en tant que recherche de la maladie causale et globalement une préoccupation spécifique à leur objectif professionnel et à la formation.

Cette expérience met ainsi en évidence le caractère déterminant du critère d'intelligibilité dans la construction des catégories de signaux sonores produits par des sujets présentant un TPPP mais elle démontre également qu'il ne s'agit pas d'un prototype exclusif. La constitution des classes fait émerger un deuxième critère important : « la normalité ou l'absence de trouble ». Ces deux critères semblent jouer le rôle d'attracteur pour d'autres critères moins « stables » donnant lieu à des classes dont la constitution est moins homogène voir différente en fonction du niveau d'expertise des auditeurs. En fonction de la production de la parole, les stratégies engagées par les participants varient (prototype de normalité vs prototype d'intelligibilité/incompréhensibilité). On ne retrouve pas de classification selon le sexe et l'âge des locuteurs car dans cette expérience ce dernier paramètre est dépendant de la présence ou non d'une pathologie. Mais ces distinctions existent dans les commentaires post-catégorisation. Les différences de production liées à ces derniers critères sont tout à fait audibles (en témoigne les commentaires), mais n'ont pas été considérés comme pertinents pour la catégorisation. Ainsi, l'analyse des commentaires contribue à expliciter ces critères avec du côté de l'attracteur « normalité » des termes se réfèrent notamment à la qualité de la voix et à son impact sur l'auditeur (des caractéristiques acoustiques au caractère agréable), à l'identité des personnes (âge, sexe) et du côté de l'attracteur « intelligibilité » des termes faisant référence à une explication (accent, lecture, maladie) ou au trouble authentifié (mécanisme du trouble, étiologie, gravité).

Ainsi nous proposons l'hypothèse d'une classification déterminée à une extrémité par une comparaison à un « prototype de normalité » (qui n'inclut pas nécessairement la notion d'intelligibilité) et à l'autre extrémité par le caractère « d'intelligibilité ou de compréhensibilité ». En fonction de l'input ou entrée suggérant un prototype de voix/parole nous n'utiliserions pas les mêmes critères (Figure n°3.12).

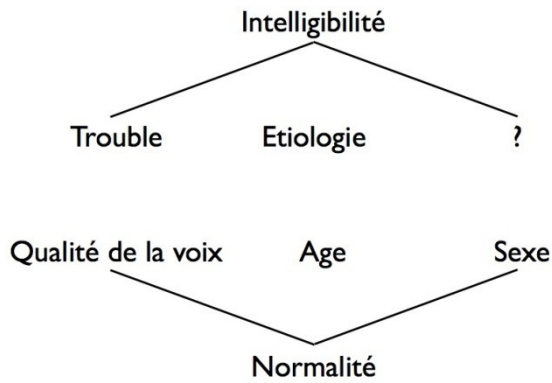


Figure n°3.12 : Hypothèse schématique des critères sous-tendant la perception des troubles de la parole pathologique.

Ceci a justifié une seconde étude sur la catégorisation libre de stimuli de patients présentant une pathologie vocale décrite dans la section suivante.

### 3.3. SECONDE ETUDE

---

Cette étude est en cours et fera l'objet d'un mémoire d'orthophonie. Elle a bénéficié du travail dirigé par Alain Ghio pour le mémoire d'orthophonie de Frédérique Weisz en 2010 permettant d'utiliser un corpus de stimuli vocaux analysés par un jury d'écoute pour une étude translinguistique de la perception des dysphonies (Weisz 2010).

Seules les données permettant d'appréhender la différence de classement par une méthode de catégorisation libre de stimuli comprenant des troubles de la production vocale pathologique isolés par rapport à un TPPP ont été traitées. Pour ne pas nous écarter de notre sujet, nous analyserons les résultats dans le but de vérifier si l'attracteur « normalité » des TPPP s'apparente à celui de la voix et de rechercher les critères relevés dans les commentaires induits par cette entrée par rapport à ceux relevés pour les TPPP.

Dans l'étude en cours, des experts et des non experts ont été recrutés de la même façon en tant qu'auditeurs mais ce sont des patients, eux même traités pour une dysphonie qui ont réalisé la manipulation. Nous excluons, pour les mêmes raisons de comparabilité la présentation des résultats

obtenus avec ce groupe d'auditeurs dans la mesure où un groupe de personnes proche d'un locuteur présentant un trouble est différent d'un groupe d'auditeurs présentant eux même le trouble.

---

### 3.3.1. MATERIEL ET METHODE

---

#### 3.3.1.1. LES LOCUTEURS

---

Le signal produit par quarante-cinq locuteurs adultes, 6 hommes et 39 femmes a été extrait de la base de données des « dysphonies » enregistrée à l'hôpital de La Timone à Marseille lors d'une consultation justifiant un bilan vocal.

La sélection des locuteurs a été faite sur les critères de pathologies permettant de couvrir un large spectre de caractéristiques vocales : 10 cas de dysphonie dysfonctionnelle (sans lésion), 9 cas de nodule, 9 cas de polype, 8 cas d'œdème de Reinke et 9 cas de paralysie laryngée.

#### 3.3.1.2. LE CORPUS

---

Le corpus est constitué du texte de la chèvre de monsieur Seguin. L'enregistrement a été effectué avec le module d'enregistrement de la station EVA décrite en annexe 7.1.3. L'ensemble des enregistrements a été conservé pour l'étude. Aucune manipulation du signal n'a été pratiquée.

L'analyse résultant du jury d'écoute permet de caractériser sur les 3 critères issus de l'échelle d'Hirano (Hirano 1981) les scores de G (grade de sévérité ou global), R (Roughness ou raucité), B (Breathness ou soufflé) pour chaque stimulus. Pour 2 stimuli, seul le grade a été attribué par un membre du jury.

Les caractéristiques perceptives ainsi obtenues sont présentées sur la figure n° 3.3.1.

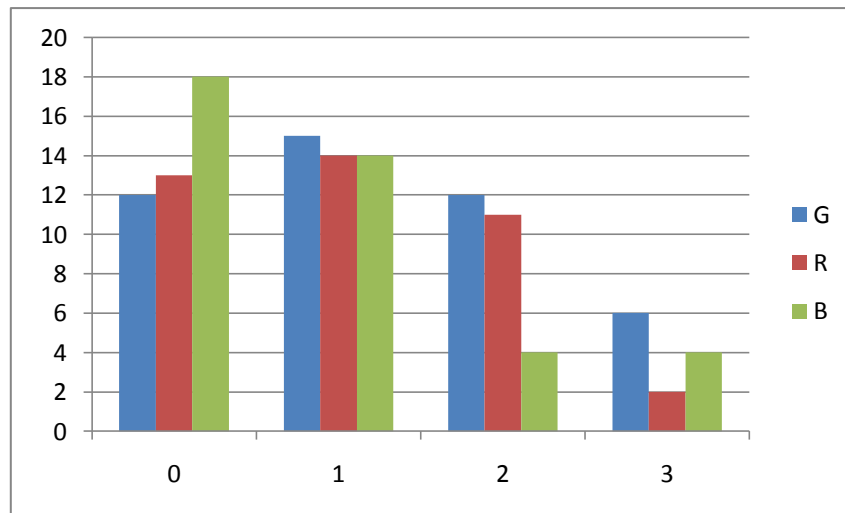


Figure n°3.3.1: Répartition en effectif des scores selon le grade de sévérité, la raucité et le souffle

Les grades 3 sont moins fréquents. Pour la raucité (R) et surtout le souffle (B) les scores sont moins élevés que pour le grade de sévérité (G). Pour rendre les comparaisons plus faciles lors de l'analyse des résultats, un regroupement en deux classes 0 et 1 / 2 et 3 sera proposé sur certains graphiques.

La distribution des différents paramètres en fonction des pathologies est présentée sur les figures n°3.3.2 à 3.3.4.

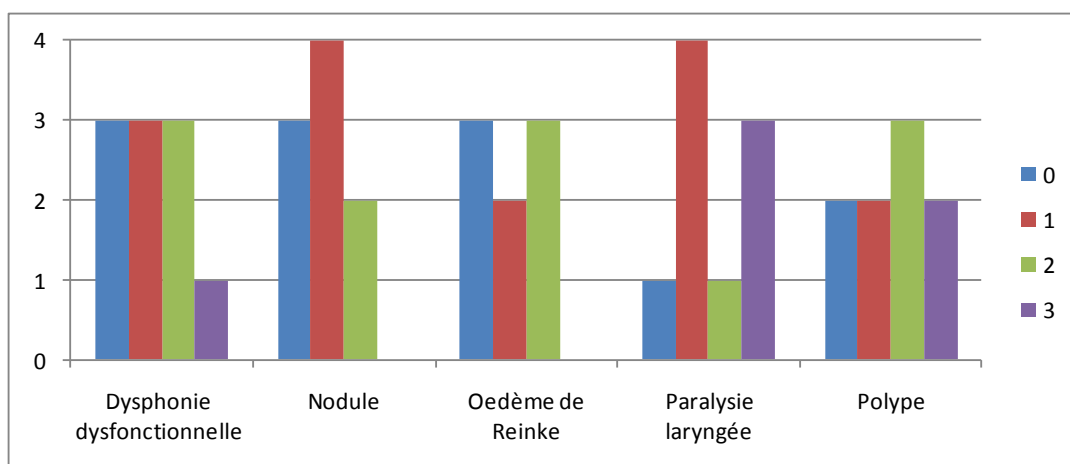


Figure n°3.3.2 : Répartition en effectif des pathologies selon la sévérité (G)

Les locuteurs ayant une paralysie laryngée ou un polype ont les grades les plus élevés, tandis que les locuteurs ayant un nodule ont les grades les moins élevés.

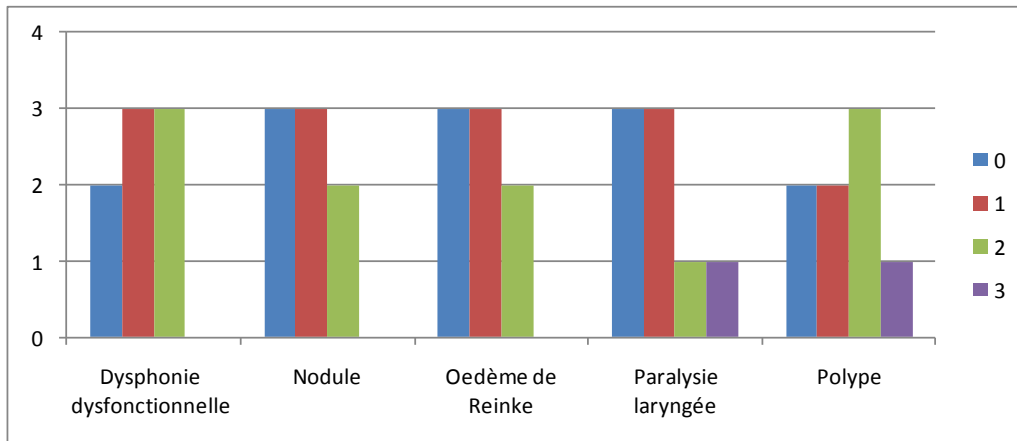


Figure n°3.3.3: Répartition en effectif des pathologies selon la raucité (R)

Concernant la raucité, les locuteurs ayant un polype ont les grades les plus élevés.

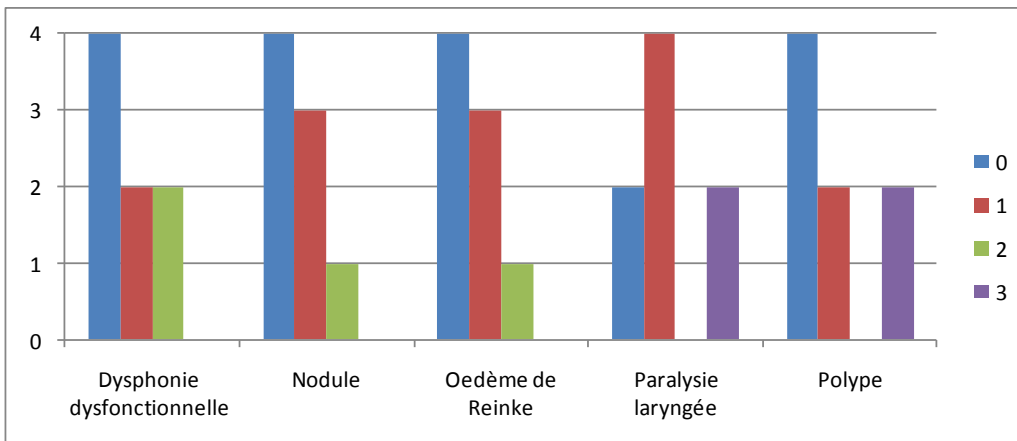


Figure n°3.3.4 : Répartition en effectif des pathologies selon le souffle (B)

Pour le souffle, ce sont les paralysies laryngées (avec plus de grades 1 et moins de grades 0 que les autres pathologies).

Comme cela était recherché, les différentes étiologies permettent de couvrir toute l'étendue de l'échelle choisie.

Les locuteurs ont tous une pathologie vocale. Néanmoins certains d'entre eux ont reçu le score 0 (aucune anomalie perceptive) sur les trois critères. Les pathologies sont toutes concernées : les nodules (2 locutrices), l'oedème de Reinke (3 locutrices), le polype (1 locutrices et 1 locuteur), la paralysie laryngée (1 locutrice) et la dysphonie dysfonctionnelle (2 locuteurs).

Les grades les plus sévères (3 sur les 3 critères) concernent deux locuteurs : une locutrice ayant un polype et un locuteur ayant une paralysie laryngée.

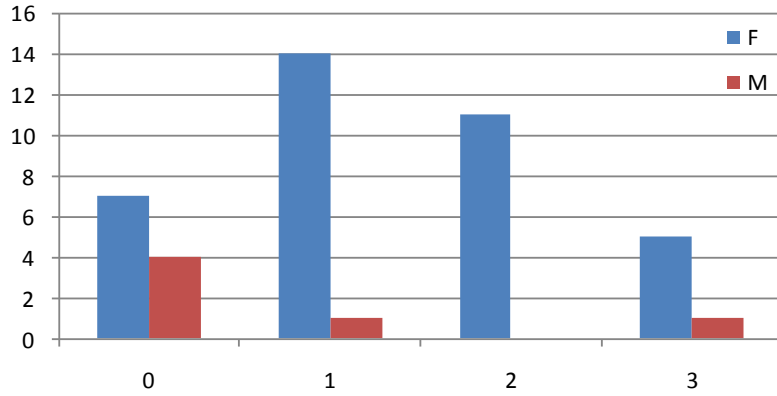


Figure n°3.3.5 : Répartition en effectif du grade de sévérité (G) selon le genre

La distribution des scores en fonction du genre montre que le faible nombre d'hommes représentés a des scores relativement moins sévères que les femmes. La répartition étant comparable pour le G, B, R, ne figure ici que le graphique pour le R (figure n°3.3.5).

Nous ne disposons pas de l'âge des locuteurs dans la base de données.

### 3.3.1.3. METHODE

La même méthodologie a été appliquée que dans la première étude.

### 3.3.1.4. LES AUDITEURS

Comme pour la première étude, aucune hypothèse sur les critères d'âge et de sexe n'a été formulée. Seul, le critère de l'expertise nous a semblé pertinent dans une première phase. Les 33 auditeurs correspondent à 2 groupes en fonction de leur niveau d'expertise dans le domaine des TPPP. Le test a été effectué par 2 groupes d'auditeurs : 15 avertis (13 femmes, 2 hommes), 18 experts (17 femmes, 1 homme).

Le groupe des non experts se compose de 4 externes en médecine générale et de 11 étudiantes en première année d'orthophonie. Aucun des étudiants n'a encore été en contact pour des raisons personnelles ou professionnelles avec des personnes présentant un trouble de la production pathologique de la parole.

Celui des experts comprend 3 phoniatres et 15 orthophonistes qui prennent tous en charge pour des bilans ou de la rééducation des patients dysphoniques.

### 3.3.1.5. PLAN D'ANALYSE DES RESULTATS

Il est comparable à celui de la première étude avec seulement une adaptation des critères perceptifs à la dysphonie puisque la tâche demandée au jury était d'associer un score à chaque stimulus avec l'échelle du GRBAS.

### 3.3.2. DESCRIPTION DE LA PROCEDURE DES TESTS

#### 3.2.2.1. PROCEDURE DU TEST DE CATEGORISATION

La procédure utilisée a été la même que dans la première étude.

#### 3.2.2.2. PROCEDURE DE L'ANALYSE DES STIMULI

##### 3.2.2.2.1. ANALYSE PAR UN JURY D'AUDITEURS EXPERTS

Le jury d'écoute été composé de 6 experts (phoniâtres et orthophonistes de Marseille et de Toulouse). Les stimuli ont été présentés, randomisés, à l'aide d'un programme d'écoute et d'évaluation vocale élaboré par Ghio (2003), le programme Perceval.

##### 3.2.2.2.2. ANALYSE ACOUSTIQUE

Chaque stimulus a été traité par le logiciel EVA de la société SQLAB permettant d'obtenir des mesures de fréquence fondamentale et de l'intensité du signal ainsi qu'une analyse de sa durée totale et des pauses (pourcentage de temps passé en pause et nombre de pause). Les informations concernant l'analyse effectuée avec ce logiciel figurent en annexe ainsi que l'ensemble des résultats (annexe 7.1.5).

Figure n°3.3.6 : Graphe en boîte de la fréquence fondamentale moyenne en fonction du genre.

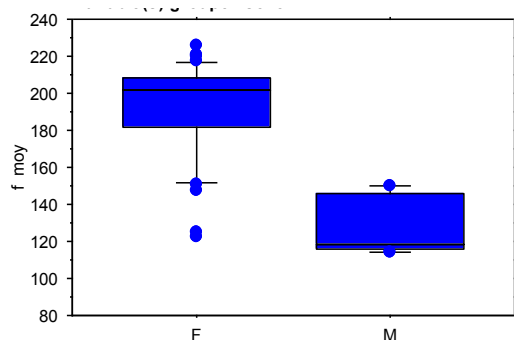
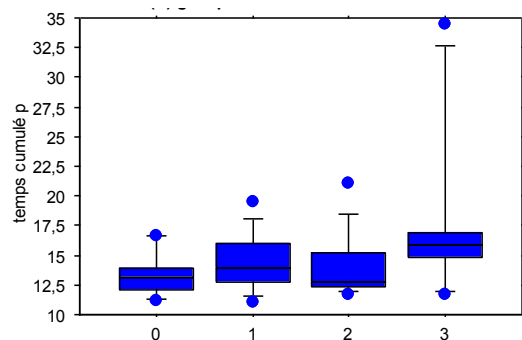


Figure n°3.3.7 : Graphe en boîte du temps cumulé passé en pause en fonction du grade (G) de la dysphonie





Seul deux résultats significatifs sont signalés ici :

- la différence des paramètres de fréquence avec le genre, observation attendue (fréquence plus grave des voix d'hommes) (figure n°3.3.6) ;
- la différence du temps cumulé passé en pause qui est nettement plus longue pour les grades 3 que les grades 0 (figure n°3.3.7).

A noter que parmi les paramètres d'intensité, seule la dynamique de l'intensité a été utilisée. En effet, il semble exister des différences de calibrage lors de l'enregistrement ne permettant la comparaison des valeurs moyennes, minimales et maximales. Mais ce calibrage étant le même pendant la durée d'enregistrement, le calcul de la dynamique est juste.

---

### 3.3.3. RESULTATS

---

#### 3.3.3.1. DETERMINATION DES CLASSES ET CONSTRUCTION DES ARBRES

---

Les arbres de distance représentant les classes «moyennes» construites par l'ensemble des sujets sont présentés sur les figures n°3.3.8 et 3.3.9. La détermination des classes en fonction du groupe d'auditeur est issue de l'analyse des scores des nœuds et des arêtes. Leur dénomination a été déterminée par les différentes catégories de stimuli partagées par les groupes d'auditeurs.

Pour les experts (figure n°3.3.8), 4 classes sont repérées.

La classe AB comprend 12 stimuli. Nous n'intégrons pas le stimulus s27 à l'intérieur de cette classe, mais nous nous interrogerons sur sa spécificité. La classe C est composée de 7 stimuli, la classe D de 6, la classe E de 7. Treize stimuli n'intègrent aucune classe, ils sont dits « non classés ».

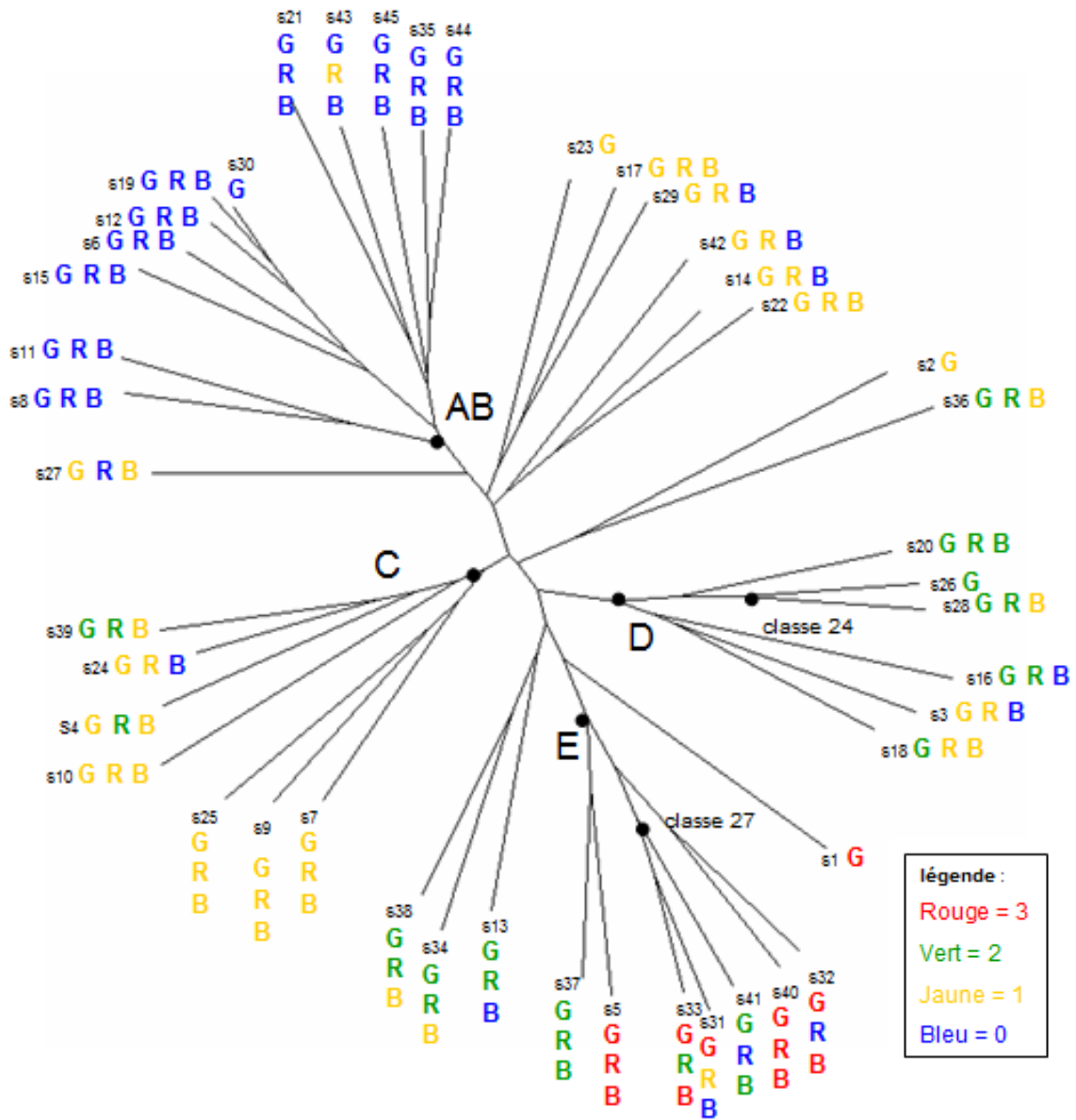


Figure n°3.3.8 : Arbre du groupe des auditeurs experts

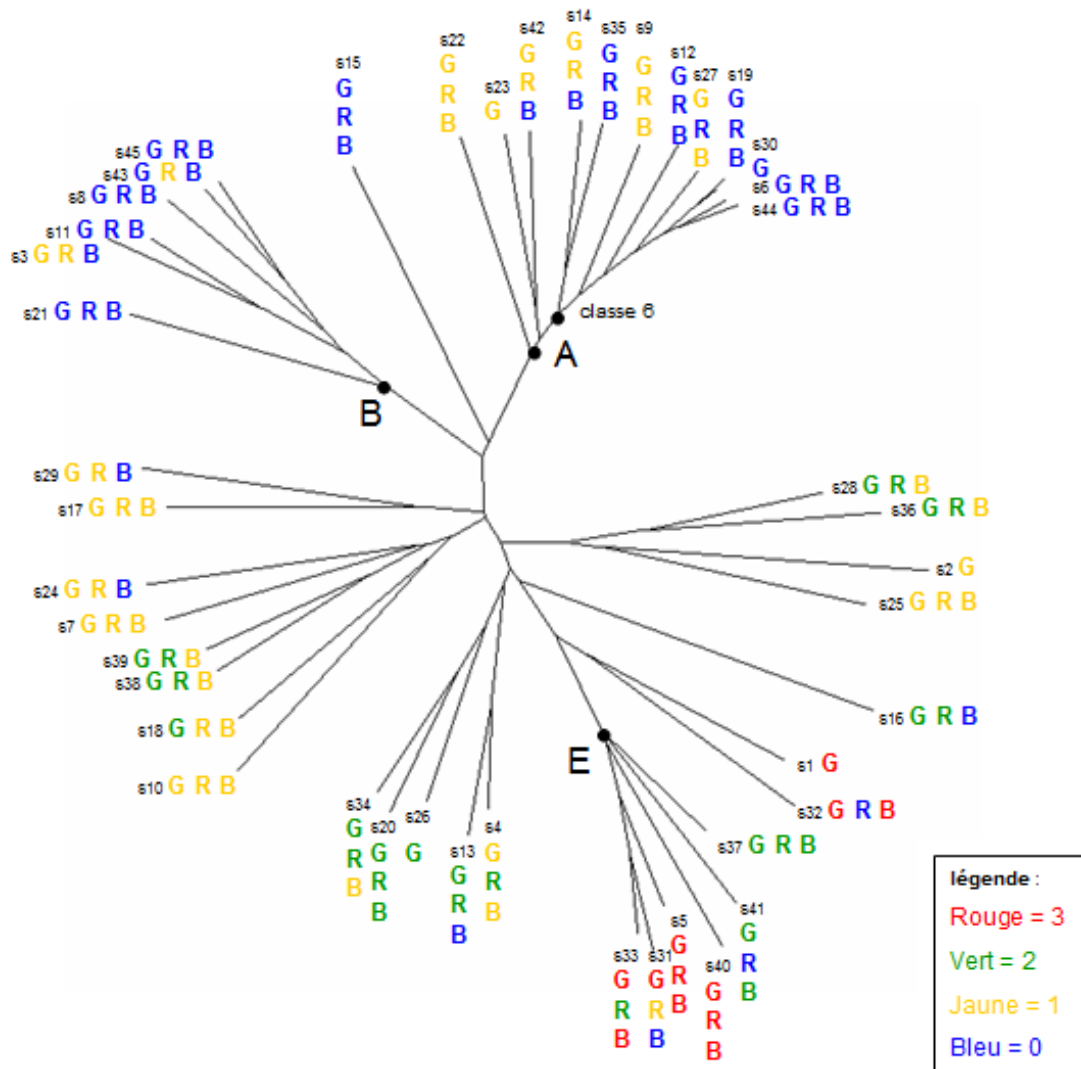


Figure n°3.3.9: Arbre du groupe des auditeurs non experts

Pour les non experts (figure n°3.3.9), 3 nœuds permettent de repérer 3 classes dénommés A (12 stimuli), B (6 stimuli), E (6 stimuli). Le reste des stimuli, soit 21, n'ont pas été classés.

On peut se questionner quant au stimulus s15, qui semble aussi proche de A que de B sans pour autant être intégré dans aucune de ces deux classes.

Afin d'étudier ultérieurement les différences de construction des arbres entre ces deux groupes d'auditeurs, voici listé les stimuli remarquables de part leur situation frontière entre les classes.

Pour les experts, il s'agit des stimuli s27 et s1. Le premier intègre la classe A des non experts alors que le seconde demeure non classé.

Pour les non experts, ce sont les stimuli s15, s1 et s32. Le premier est compris dans la classe AB des experts, le second reste non classé et le 3<sup>ème</sup> est dans la classe E.

### 3.3.3.2. ANALYSE DES ARBRES

Le nombre de catégories créées par les auditeurs est en moyenne de 9,6 +/-3,7 (9,1 pour les non experts, 9,9 pour les experts) avec un nombre moyen de stimuli dans la catégorie la plus petite de 1,5 pour les avertis, 1,7 pour les experts, et un nombre moyen de stimuli dans la catégorie la plus grande de 14,5 pour les avertis, 11 pour les experts.

La distribution du nombre de catégories créées par groupe d'auditeurs est illustrée par la figure n°3.3.10.

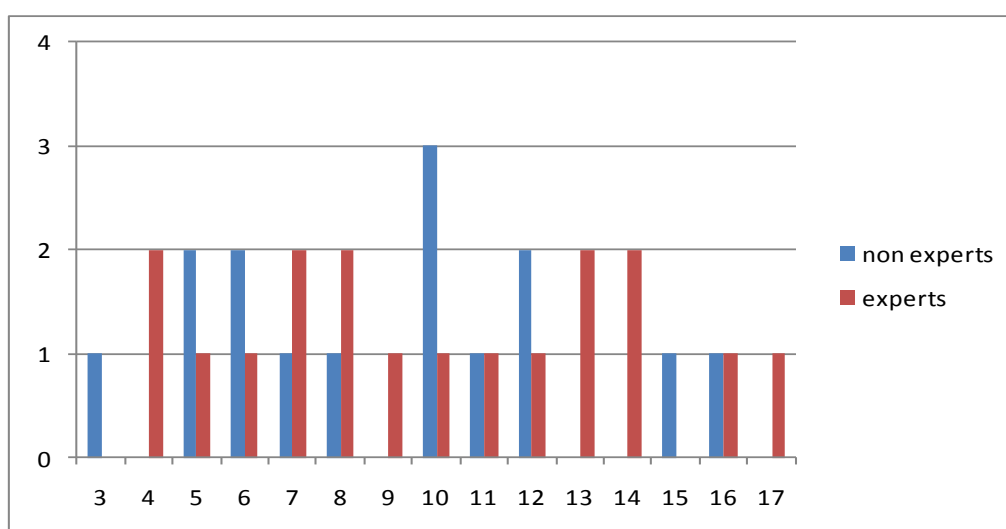


Figure n° n°3.3.10: Distribution du nombre (effectif) de catégories créées

Elle montre que les experts tendent à faire un peu plus de catégories.

#### 3.3.3.2.1. ANALYSE DES CLASSES PAR RAPPORTS AUX RESULTATS DE L'EXPERTISE DU JURY

##### 3.3.3.2.1.1. POUR LES EXPERTS (FIGURES N°3.3.11 A N°3.3.13)

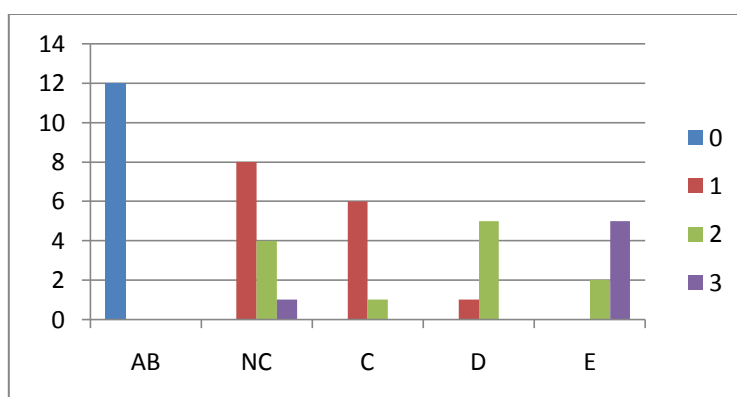


Figure n°3.3.11: Répartition des grades de sévérité (G) selon les classes des arbres

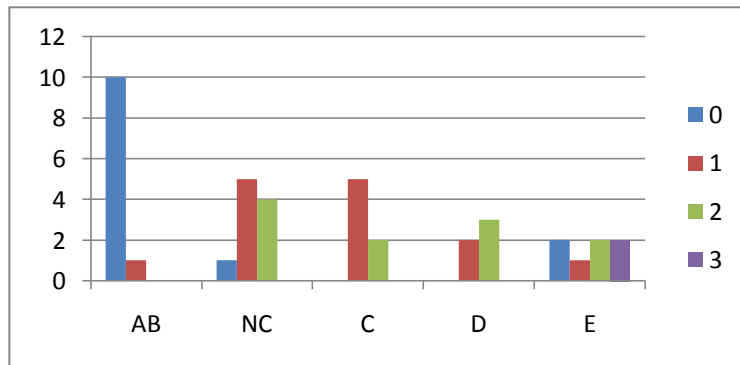


Figure n°3.3.12: Répartition des grades de raucité (R) selon les classes des arbres

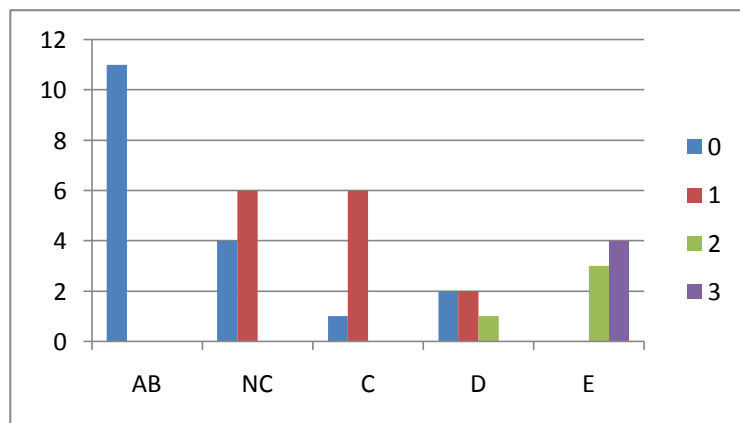


Figure n°3.3.13: Répartition des grades de souffle (B) selon les classes des arbres

La classe AB de l'arbre comprend toutes les voix sans anomalie perçue (cotées avec un grade 0) pour la sévérité, la raucité et le souffle, sauf s43 dont la raucité est de 1. A noter que s27, non classé mais proche de cette classe est différent du fait de scores G et B supérieurs à 0.

La classe E comprend toutes les voix de grade 3 sauf s1 (bien qu'il soit aussi G3) et deux voix un peu moins sévères (s37 et s41).

La classe C comprend essentiellement des voix de grade 1 pour tous les paramètres.

Le groupe des non classés (NC) de l'arbre comprend essentiellement des voix de grades 1. Ces stimuli se placent entre les classes C et D sur le plan des grades donnés par le jury

La classe D comprend des voix de grades 1 à 2 sauf sur le souffle où cela est moins marqué.

Globalement les classes traduisent un critère de jugement de sévérité sur tous les paramètres G, R dans l'ordre AB/C / NC /D/E et AB/NC/C/D/E pour le souffle (B).

### 3.3.3.2.1.2. POUR LES NON EXPERTS

Pour les non experts comme pour les experts, on observe une distribution entre 2 pôles distants de G0 à G3.

Les classes A et B incluent toutes les voix gradées à 0 pour la sévérité et quelques grades 1. Sont regroupés aussi la plupart des voix dont la raucité et le souffle ont été coté à 0 et en moindre proportion des voix cotées 1. S15 est la seule voix de grade 0 à avoir été mise à part. Nous chercherons par la suite ce qui distingue A de B.

La classe E regroupe les voix de sévérité 3, excepté s1 et s32, et deux voix (s41 et s37) de sévérité 2. On retrouve ici le même phénomène que chez les experts pour les stimuli s1, s41 et s37 avec en plus, l'exception du stimulus 32.

Le groupe NC se compose des voix de sévérité, raucité et souffle gradées de 0 à 3 pour G et B, 0 à 2 pour R.

On ne trouve pas de classe dont les voix ont été uniquement cotées avec des grades 1. Ceux-ci sont soit non classés, soit dans le groupe A.

Globalement, les classes traduisent un critère de jugement de sévérité sur tous les paramètres G, R B dans l'ordre B / A / NC /E. Dans la mesure où la situation est similaire pour les autres grades, ces résultats ne seront illustrés que pour le G par la figure n°3.3.14.

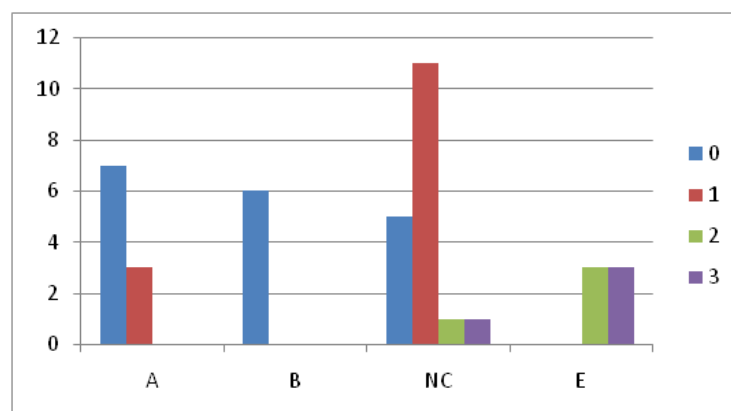


Figure n°3.3.14: Répartition des grades de sévérité (G) selon les classes des arbres

Cette homogénéité dans la catégorisation par rapport à ces 3 paramètres n'a pas été retrouvée chez les experts. Cela traduit-il l'influence de l'apprentissage inhérent à la formation des professionnels sur la capacité à différencier les stimuli à partir de ces paramètres ?

### 3.3.3.2.2. ANALYSE DES CLASSES PAR RAPPORTS AUX CARACTERISTIQUES DES LOCUTEURS (GENRE ET PATHOLOGIE)

#### 3.3.3.2.2.1. POUR LES EXPERTS

Pour les experts (figure n°3.3.15), les 2 pôles de l'arbre avec à une extrémité la classe AB et à l'autre les classes D et E sont mixtes dans le sens où les 2 genres féminin et masculin sont représentés.

Entre les classes NC et C ne comprennent que des femmes (figure n°3.3.16). Ainsi, le genre semble être un critère de tri prégnant probablement à travers les paramètres acoustiques dont la fréquence fondamentale.

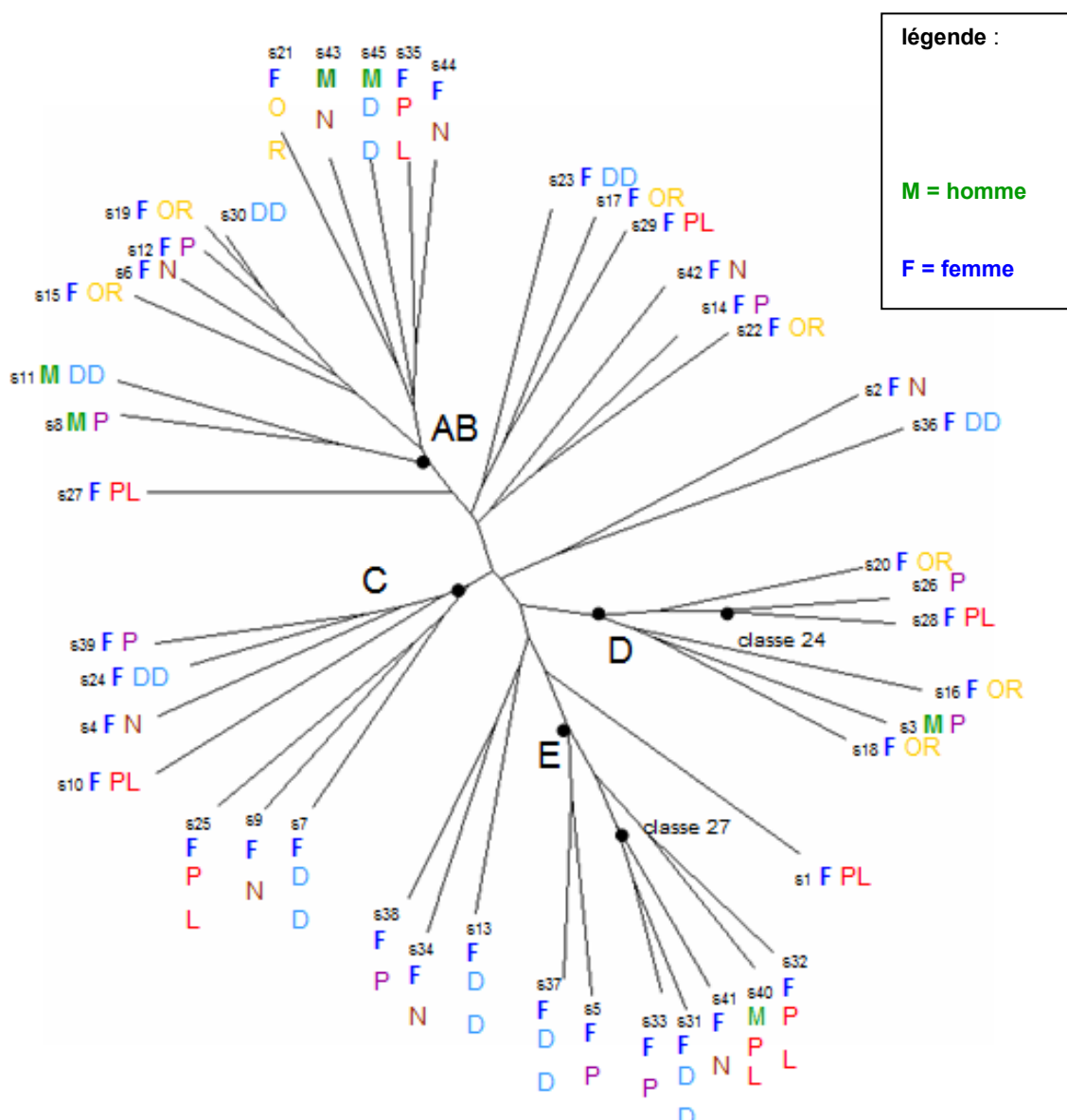


Figure n°3.3.15: Lecture de l'arbre des experts selon le sexe et la pathologie des locuteurs

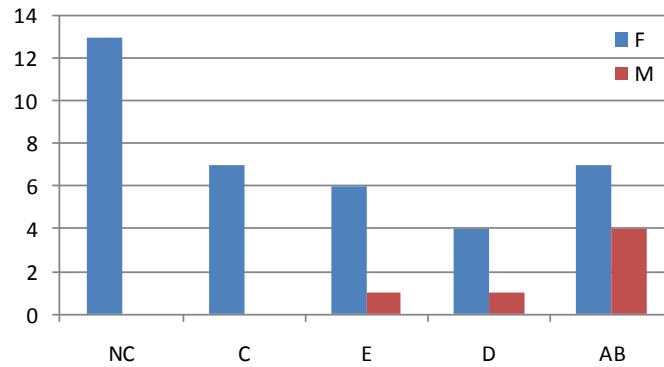


Figure n°3.3.16 : Répartition du genre dans les classes de l'arbre

Concernant la pathologie (figure n°3.3.17), on ne trouve pas de dysphonies dysfonctionnelles ni de nodules dans la classe D. Il n'y a pas d'œdème de Reinke ni dans C, ni dans E.

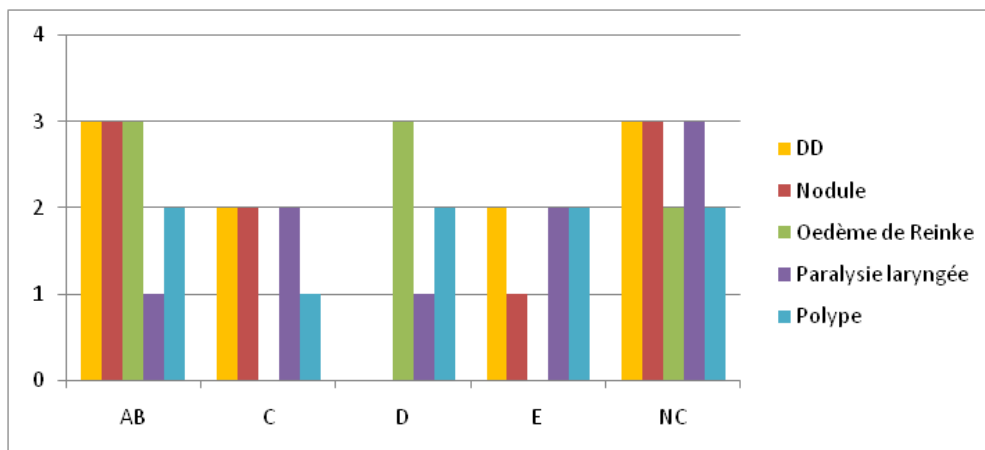


Figure n°3.3.17: Répartition des pathologies dans les classes de l'arbre (effectif)

Dans la mesure où la classe C ne comprend que des femmes, la fréquence de la voix, classiquement aggravée en fréquence dans les œdèmes, peut expliquer que ces voix ne soient pas intégrées. La classe E est constituée par des stimuli jugés sévères alors qu'aucun œdème de Reinke ne présente un score sévère > 2. Le reste se répartit de façon homogène.

#### 3.3.3.2.2.2. POUR LES NON EXPERTS

Pour les non experts (figures n°3.3.18), la classe B regroupe tous les hommes, sauf s40 qui figure dans la classe E, et une femme s21. Notons néanmoins que la locutrice s21 a un œdème de Reinke. Il semble que les auditeurs l'ait assimilé à un homme. Nous vérifierons cette hypothèse en analysant ultérieurement les fréquences.



La classe A ne comprend que des femmes. Ce qui distingue donc A de B est le sexe. A est la classe des voix féminines jugées normales ou peu sévères, et B est l'équivalent pour les voix masculines non sévères.

La classe NC ne comprend que des femmes, elle correspond aux voix féminines peu sévères à sévères.

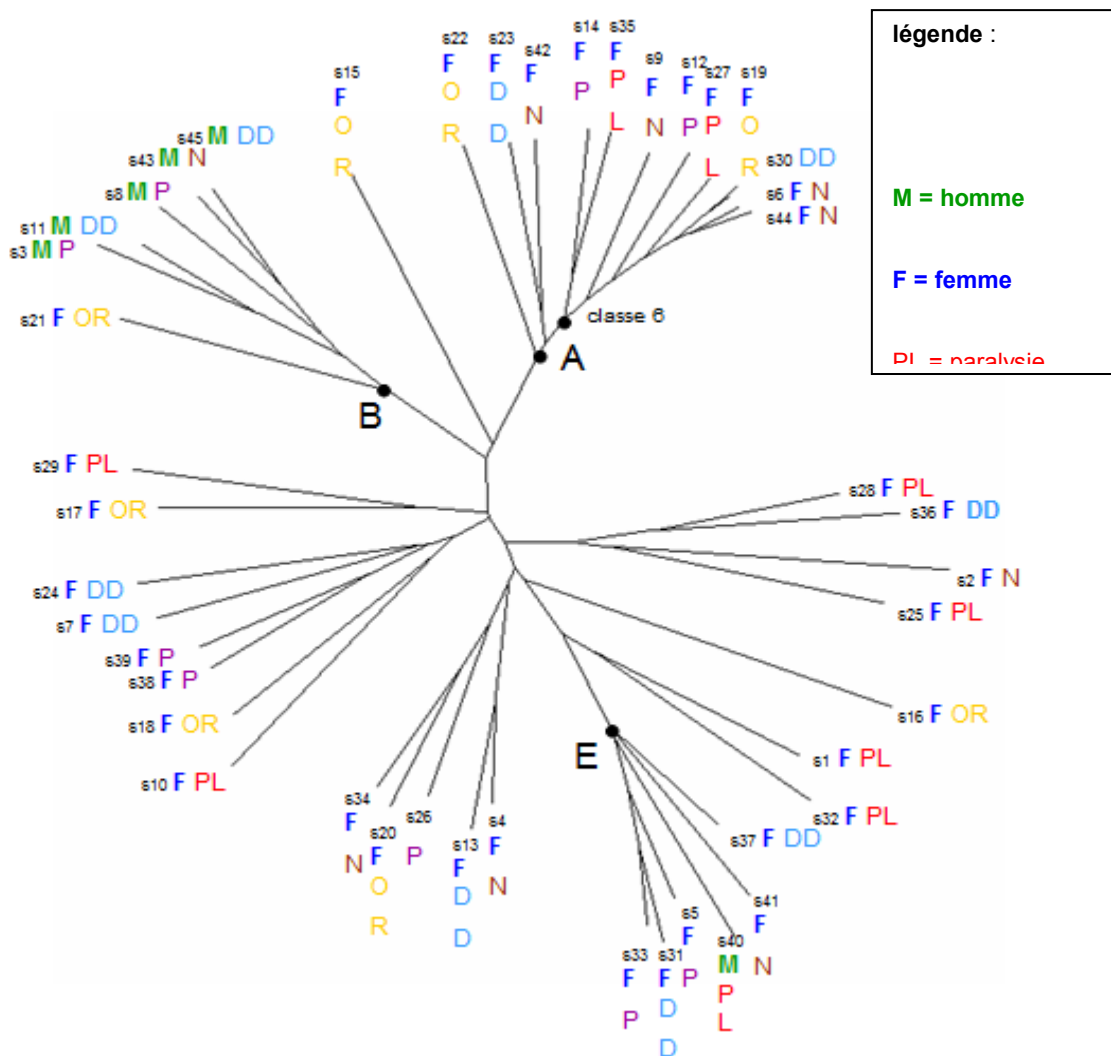


Figure n°3.3.18: Lecture de l'arbre des non experts selon le genre et la pathologie des locuteurs

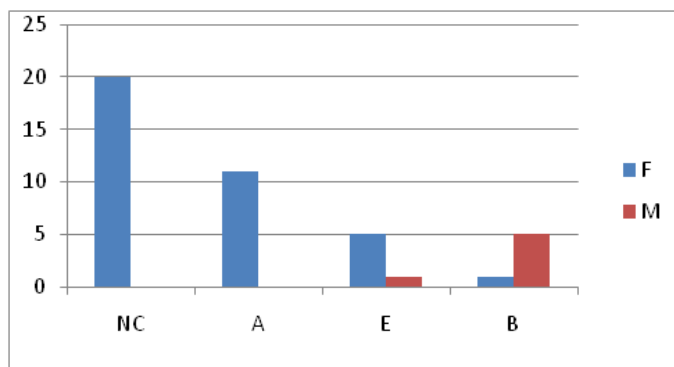


Figure n°3.3.19: Répartition du genre dans les classes de l'arbre

La classe E contient des femmes et un homme. Nous vérifierons avec l'analyse fréquentielle si le locuteur a été assimilé à une femme ou s'il n'a été classé avec les autres stimuli que sur le critère de la sévérité.

A noter que les hommes sont classés plus sévèrement que ce que ne laisserait prévoir le jury en terme GRB.

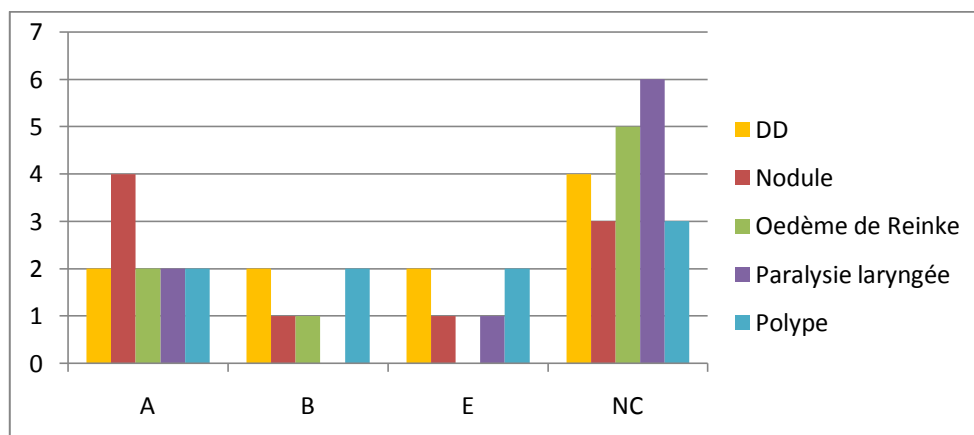


Figure n°3.3.20: Répartition des pathologies dans les classes de l'arbre (effectif)

Par rapport aux pathologies (figure n°3.3.20), on remarque qu'il y a dans A plus de nodules, pathologie plus féminine. Dans NC, on trouve plus de paralysies laryngées. Il n'y a pas d'oedème de Reinke dans E, et pas de paralysie laryngée dans B. Toutefois, nous savons que les paralysies laryngées et les polypes sont dans l'ensemble très sévères alors que les nodules et œdèmes de Reinke ne le sont pas.

Il n'y a donc vraisemblablement pas eu de classement selon la pathologie.

### 3.3.3.2.3. ANALYSE ACOUSTIQUE

#### 3.3.3.2.2.1. POUR LES EXPERTS

Il n'y a pas de différence significative entre les classes par la fréquence, mais la classe B est marquée par des fréquences plus aiguës (figure n°3.3.21) particulièrement différente de D. Cette opposition de fréquence peut être liée aux pathologies (œdème de Reinke, polype, paralysie laryngée).

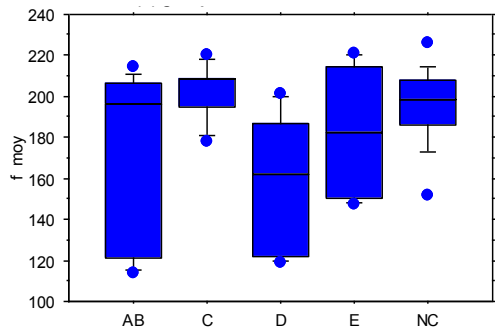


Figure n°3.3.21 : Graphe en boîte de la fréquence moyenne des stimuli dans les classes déterminées par les experts.

Une influence des facteurs temporels (figure n°3.3.22) notamment la durée du signal total, semble présente. L'autre facteur qui ne parait pas intervenir par sa durée concerne les pauses par leur nombre et leur répartition et ceux pour les classes C,D,E.

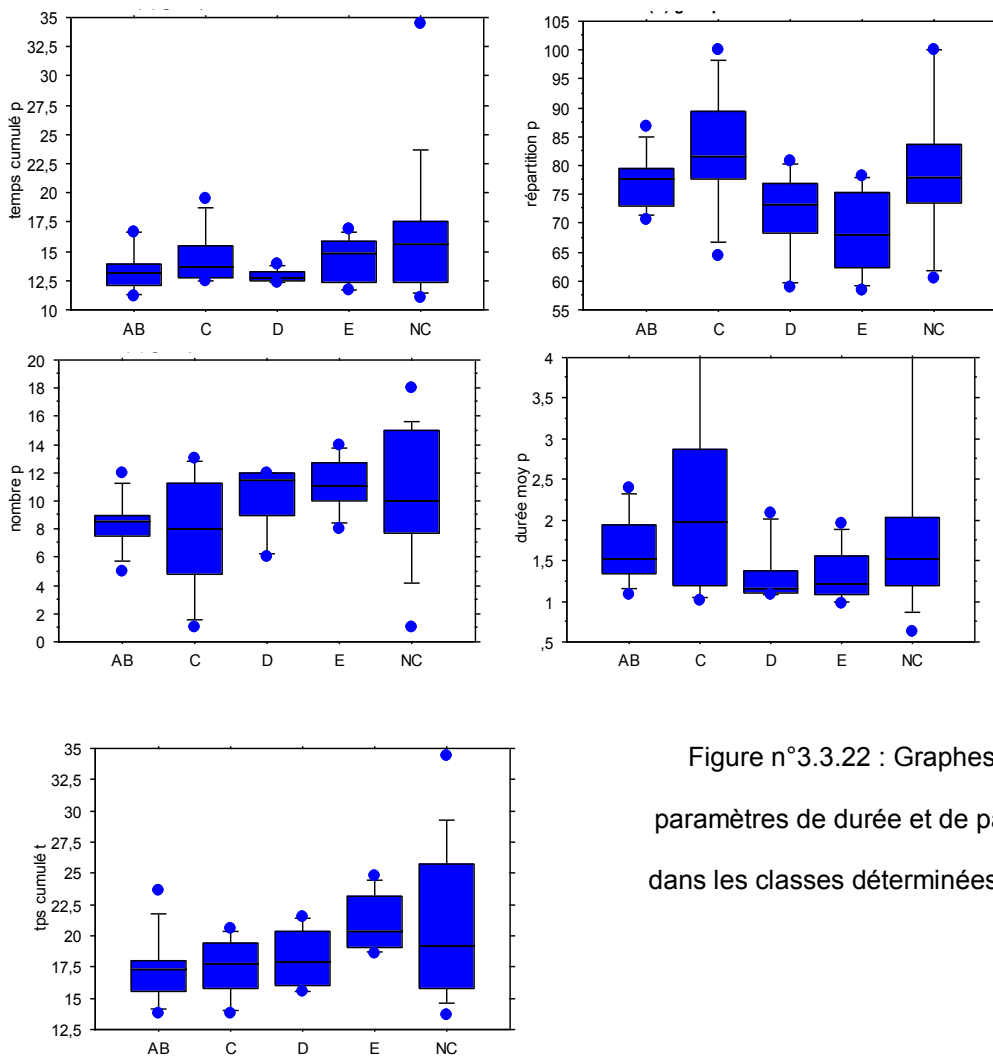


Figure n°3.3.22 : Graphes en boîte des paramètres de durée et de pause des stimuli dans les classes déterminées par les experts.

### 3.3.3.2.2. POUR LES NON EXPERTS

Le rôle prépondérant du genre dans la détermination de la classe B est concordant avec le phénomène observé sur la fréquence. A noter que le s21 est une femme avec un oedème de Reinke et une fréquence de voix particulièrement basse à 125Hz

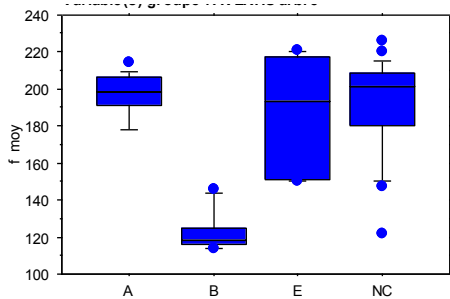


Figure n°3.3.23 : Graphe en boîte de la fréquence moyenne des stimuli dans les classes déterminées par les experts.

Le rôle des paramètres plutôt prosodiques concerne les pauses en terme de durée et de temps que du nombre.

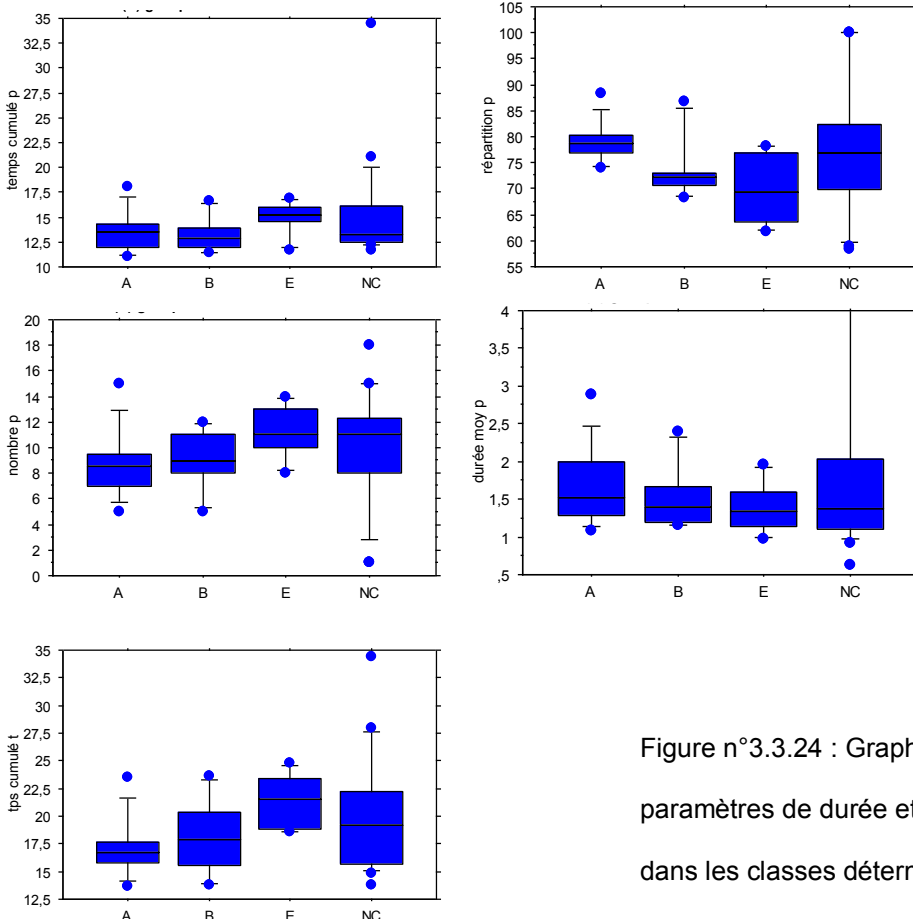


Figure n°3.3.24 : Graphe en boîte des paramètres de durée et de pause des stimuli dans les classes déterminées par les non experts.

Entre experts et non experts, l'utilisation de la fréquence ne semble pas réalisée de la même façon. Nous pouvons émettre l'hypothèse que pour les experts elle sert plutôt à déterminer les pathologies, pour les non experts elle permet d'identifier le locuteur.

Pour revenir aux différences de construction des arbres entre ces deux groupes d'auditeurs listés au début de cette analyse, les stimuli remarquables de part leur situation frontières entre les classes peuvent être interprétés de la manière suivante :

- Pour les experts, le stimulus s27 (Femme, paralysie laryngée G1R0B1) n'est pas classé en AB car il présente un score G et B différent de 0 mais il intègre la classe A malgré une fréquence fondamentale à 190 Hz (voix de femme sans anomalie ou avec des anomalies légèrement perçues au GRB des non experts). Le stimulus s1 (Femme, paralysie laryngée G3) est non classé par les 2 groupes malgré un grade 3. Il a comme autre particularité d'avoir la durée totale la plus longue de tous les stimuli. La comparaison avec les stimuli proches qui eux changent de statut comme s32 non classé par les non experts mais intégré dans la classe E par les experts montre que ce stimuli est seulement soufflé alors que s32 (également une femme avec une paralysie laryngée) et son voisin s40 présentent également un caractère bitonal.

-Pour les non experts, le stimulus s15, non classé est compris dans la classe AB. Il s'agit d'une femme avec un œdème de Reinke avec une fréquence fondamentale à 209 Hz avec des scores G0R0B0. Il est possible, si l'on retient pour les non experts un critère d'identification du genre prépondérant que cette fréquence relativement aigue porte une ambiguïté sur le genre ne permettant pas son classement entre les classes A et B.

### 3.3.3.5. ANALYSE DES COMMENTAIRES

---

Rappelons que les sujets ont eu à commenter post-catégorisation les classes qu'ils avaient formées. Des extraits de ces commentaires sont rapportés dans les tableaux figurant dans l'annexe n°7.1.5.3.

#### 3.3.3.5.1. LA GRILLE D'ANALYSE

La même grille d'analyse que celle utilisée dans la première étude a été utilisée.

#### 3.3.3.5.2. REFERENCES A DES CRITERES IDENTIQUES A CEUX DU GRBASI

Le GRBASI correspond à l'échelle d'Hirano complété par un paramètre « I » traduisant la notion d'instabilité perceptible à l'écoute du signal (Dejonckere et al 1998).

Les non experts ont peu fait allusion aux critères de cette échelle perceptive, excepté pour la raucité (53%). D'ailleurs, le vocabulaire qui se réfère à cette notion est assez diversifié (raucité, érailement, rocaillage, granulosité). Le souffle est le critère qui est le plus souvent évoqué (78%) par les experts (figure n°3.3.25). La stabilité est évoquée, contrairement aux autres groupes, mais très peu. Le terme de raucité est plus souvent employé que celui d'érailement.

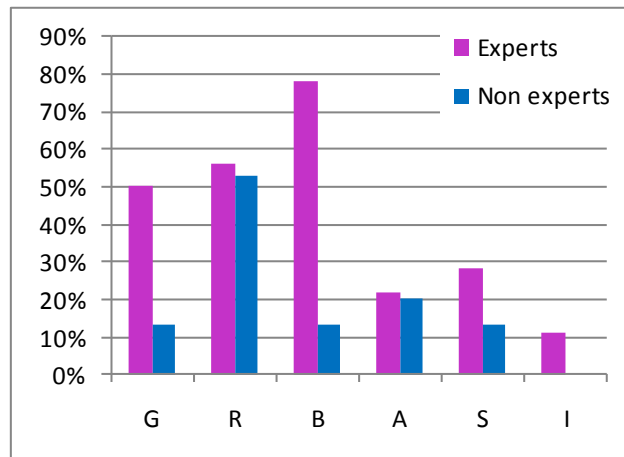


Figure n°3.3.25 : Distribution (en %) du nombre d'auditeurs se référant aux critères du GRBASI

Tous les groupes d'auditeurs comparent fréquemment certains critères d'une classe à l'autre. Le critère le plus souvent comparé est la raucité pour les non experts (30%), la sévérité pour les experts (69%),

Ces données sont en faveur de l'importance de l'adaptation des critères liée à la culture professionnelle.

### 3.3.3.5.3. DEGRE D'ALTERATION DES PARAMETRES ET REFERENCE A LA NORME

Plus de la moitié de chaque groupe d'auditeurs a fait référence à une norme, particulièrement les experts (78% vs 66% pour les non experts). En revanche, 50% des experts font référence à un degré d'altération de la voix contre 13% des non experts alors que cela apparaît clairement à la lecture des arbres quelque soit la catégorie d'auditeur.

### 3.3.3.5.4. SUPPOSITION PAR RAPPORT AU LOCUTEUR

Environ la moitié des non experts a utilisé le genre supposé des locuteurs comme critère de classement et 60% d'entre eux l'âge. Trente trois pourcents seulement des experts ont utilisé ces deux critères (figure n°3.3.26).

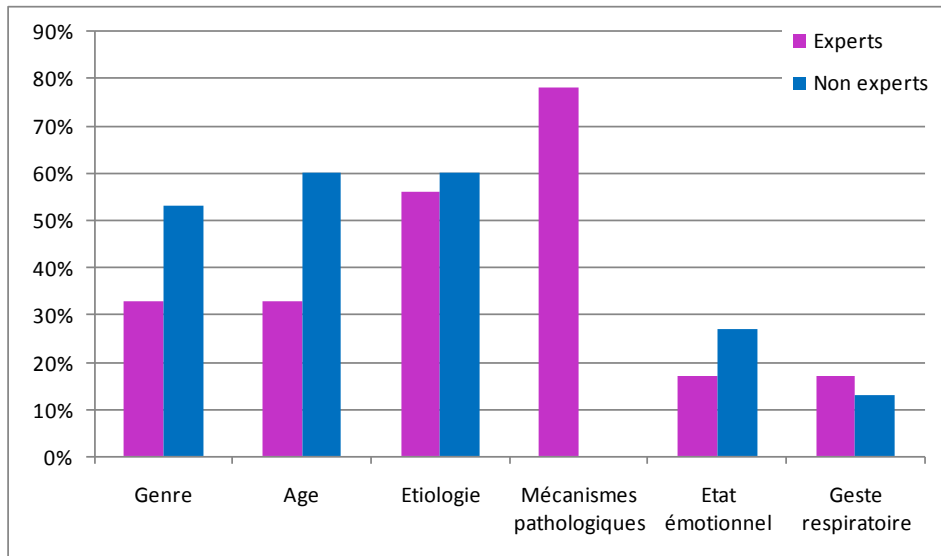


Figure n°3.3.26 : Supposition par rapport aux locuteurs

La notion de recherche étiologique a été utilisée de manière équivalente par les non experts et les experts (60% environ). L'explication des mécanismes pathologiques a intéressé 78% des experts, et très peu les non experts.

Environ 30% des non experts ont fait des suppositions sur l'état émotionnel supposé des locuteurs contre seulement 17% des experts.

Environ 15% de chaque groupe d'auditeurs a pris en compte le geste respiratoire supposé des locuteurs.

On retiendra que quelque soit la nature des auditeurs, la quête de la raison de la déviance par rapport à la normalité est commune mais la manière d'y arriver dépend des connaissances ou de la culture des auditeurs expliquant l'utilisation des mécanismes de manière exclusive par les experts.

### 3.2.3.5.5. SENTIMENTS « RESSENTIS » PAR LE LOCUTEUR A L'ECOUTE DU STIMULUS

Les sentiments générés à l'écoute des voix ont été évoqués par un tiers des auditeurs. Ces résultats sont confortés par la fréquence de recours aux images personnelles pour décrire et classer les voix : 89% pour les experts et 73% pour les non experts.

### 3.3.3.5.6. REFERENCE A CERTAINS MECANISMES ET AUX PARAMETRES

#### ACOUSTIQUES DES STIMULI

Tous les groupes d'auditeurs (figure n°3.3.27) ont fait référence à la nasalité, surtout les non experts (53% vs 39%). En revanche, les experts ont beaucoup plus évoqué les autres types d'altérations comme la bitonalité, (33% vs 13%), la désonorisation (39% vs 7%). Ils sont les seuls à avoir évoqué l'aspect voilé des voix (28% vs 0%). Ainsi la nasalité apparaît une perception traduite par un vocabulaire commun entre les deux groupes alors que les autres termes sont du domaine de l'expertise relevant probablement de l'apprentissage d'un vocabulaire spécifique.

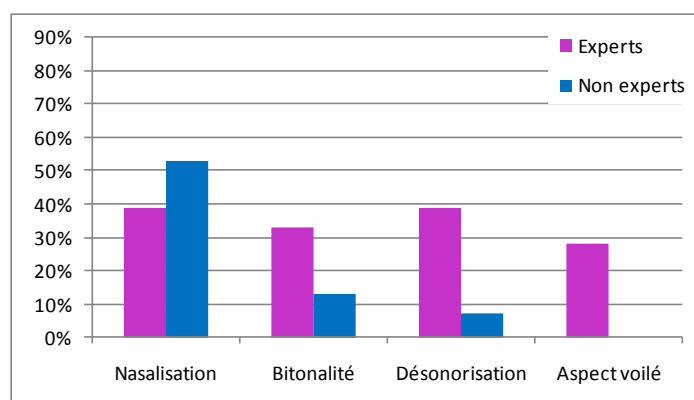


Figure n°3.3.27 : Référence à certains mécanismes

Globalement, les auditeurs n'ont pas souvent fait référence aux paramètres acoustiques, excepté à la hauteur : 56% des experts et 40% des non experts (figure n°3.3.28).

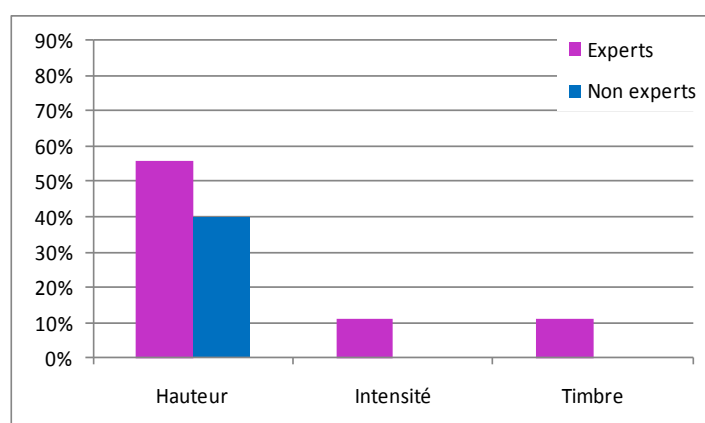


Figure n°3.3.28 : Références aux paramètres acoustiques des stimuli

### 3.3.3.5.7. REFERENCE AU SUPPORT LINGUISTIQUE

Tous les auditeurs ont utilisé la voix comme support principal de classement.



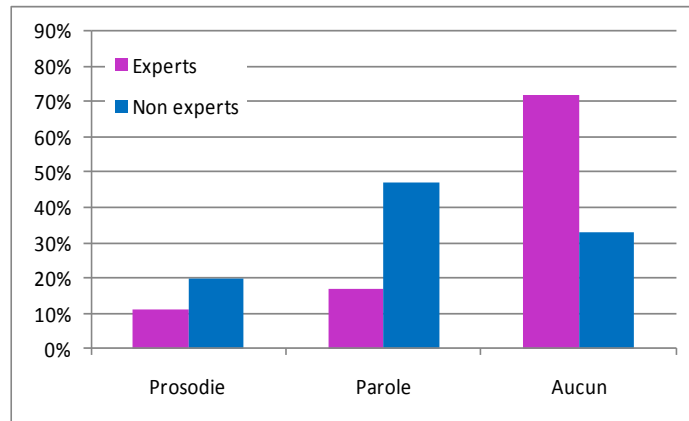


Figure n°3.3.29 : Références à un support linguistique secondaire autre que la voix

La prosodie est un support secondaire pour 20% des non experts et 11% des experts. La parole en est un pour la moitié (47%) des non experts et 17% des experts. Aucun autre support n'est évoqué en dehors de la voix pour 17% des experts et 33% des non experts (figure n°3.3.29). Ces résultats donnent l'impression d'un filtre que les experts appliqueraient pour focaliser leur attention sur l'objet qu'ils ont identifié comme leur cible.

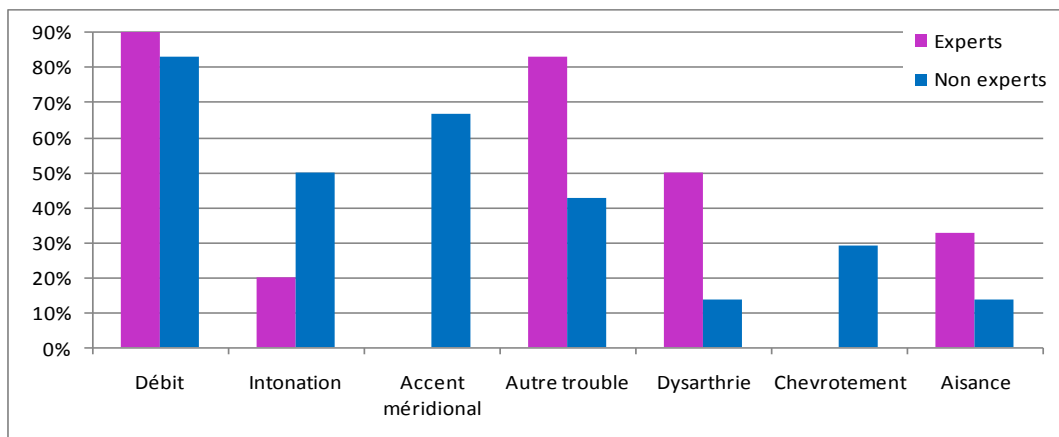


Figure n°3.3.30 : Références évoquées pour les supports secondaires

Lorsque la prosodie est évoquée, 76,5% des auditeurs s'intéressent au débit, surtout les experts (100% vs 83%). La moitié des non experts évoque l'intonation contre 20% des experts, 67% des non experts s'intéressent à l'accent méridional alors qu'aucun expert n'y fait référence.

Lorsque l'articulation est évoquée, la moitié des experts et 14% des non experts parlent de dysarthrie ; 29% des non experts d'un chevrotement ; 83% des experts et 43% des non experts d'un autre trouble. Un tiers des experts et quelques non experts remarquent l'aisance de parole. Le fait que seuls les non experts parlent de l'accent est également en faveur du filtre que peuvent utiliser les

professionnels par rapport à l'identité de l'individu au bénéfice de « l'objet voix ». L'usage du mot dysarthrie est surprenant dans le contexte. Il a été attribué par 2 experts orthophonistes et concerne 3 stimuli de la classe E (deux paralysies laryngées et un polype). A noter que les difficultés de lecture des locuteurs ont parfois été évoquées mais peu (12,5% au total).

L'ensemble donne l'impression d'un continuum voix parole en partie maqué par les experts.

La référence à l'intelligibilité est rare : une fois chez les experts, et une fois chez les non experts. Il s'agit du stimulus s1 déjà décrit. Le commentaire associé à la catégorie créée par l'un des auditeurs est le suivant : « voix sévèrement altérée avec une altération du rythme de la parole gênant la compréhension du message - bruit de prise d'air ». Le terme « d'audible ou d'audibilité » n'a jamais été utilisé.

### 3.3.3.5.8. CRITERE PRINCIPAL ET ANALYSE DE LA STRATEGIE DES AUDITEURS

L'estimation des critères de tri est décomposée en la recherche du critère de tri principal (figure n°3.3.31) et en critères de tri secondaires (figure n°3.3.32)

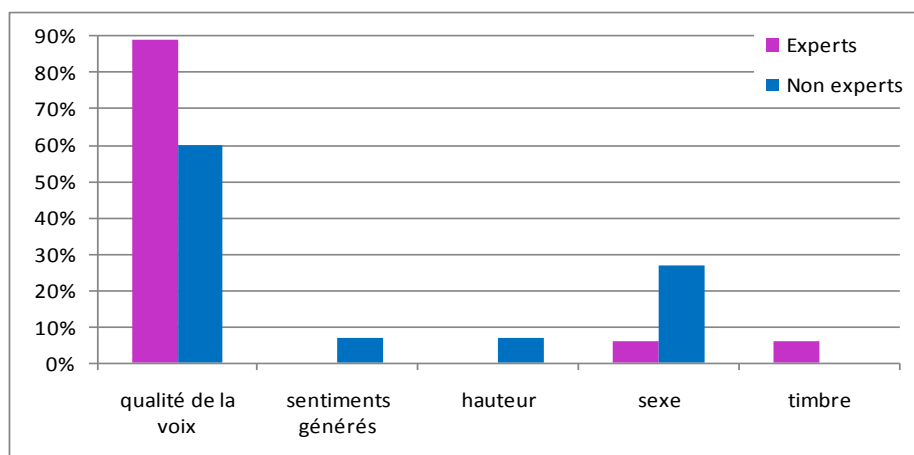


Figure n°3.3.31 : Histogrammes de l'estimation du critère de tri principal

Le critère de tri principal des auditeurs est la qualité de la voix : 60% des non experts et 89% des experts. Vient ensuite le genre surtout pour les non experts à 27%.

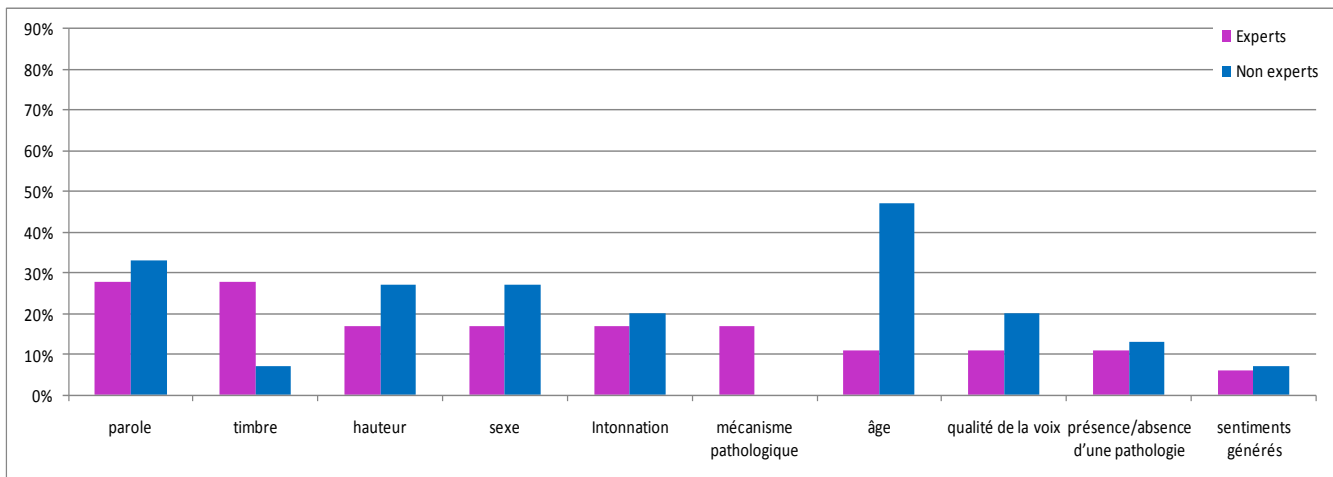


Figure n°3.3.32 : Histogramme de l'estimation des critères de tri secondaire

Douze critères de tri secondaire ont été repérés. Ils sont parfois absents pour les experts (44%). Les plus fréquents pour les non experts sont la parole, le genre et la hauteur. Pour les experts ce sont le timbre et la parole. La gêne respiratoire et l'état du locuteur n'ont jamais été utilisés comme critères de tri secondaire.

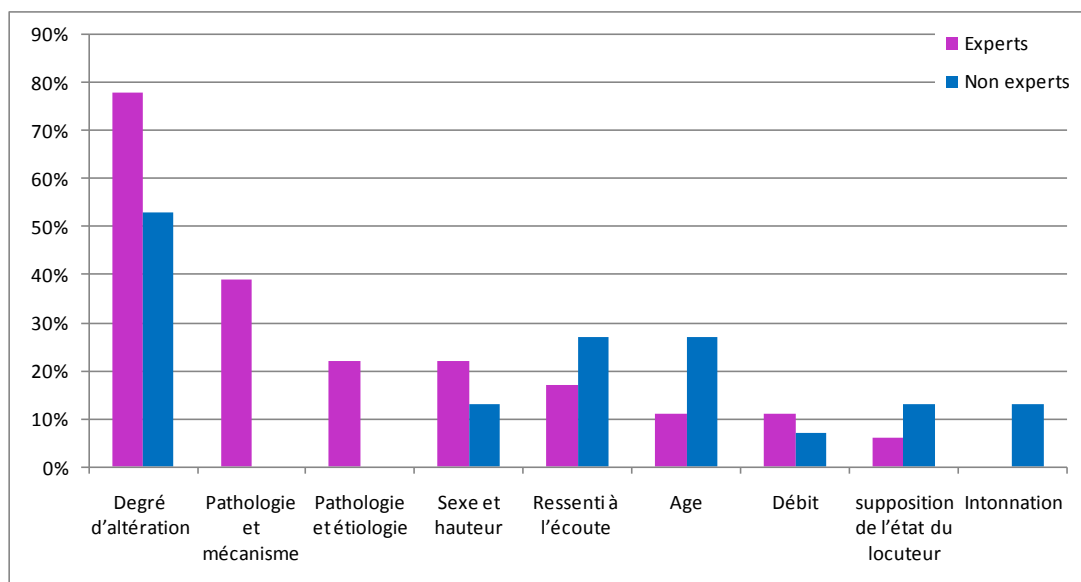


Figure n°3.3.33 : Histogramme de l'estimation de la stratégie de catégorisation

La stratégie de catégorisation la plus fréquente est le degré d'altération (figure n°3.3.33).

### 3.3.3.6. ANALYSE DES CLASSES EN FONCTION DES COMMENTAIRES

#### 3.2.3.6.1. POUR LES EXPERTS

Les 4 classes peuvent être décrites de la manière suivante :

Une classe (AB) composée de voix avec un score à 0 (sauf 1 R1) avec des commentaires ayant trait soit à des suppositions sur le genre (très peu sur l'âge), soit à la description du timbre.

Une classe (E) qui comprend des personnes qui sont G et B 3 ou 2 avec un R variable et une vitesse de parole plus lente. Dans les commentaires les qualificatifs équivalents à « désonorisé, détimbré, serré » et/ou les qualificatifs marquant un degré très important d'altération sont les plus fréquents. Deux experts expriment même le caractère douloureux pour les locuteurs et la souffrance qu'ils ressentent

Une classe (C) où les voix sont tous B1 ou B0 , presque tous G 1 (1 G2) et R 1 ou 2 et dont le timbre reste harmonieux. Les commentaires sont riches en description du timbre.

Une classe D où les voix sont toutes G et R 1 ou 2 alors que les paramètres B varient de 0 à 3.

Comme pour la classe C, les commentaires sont riches en description du timbre sans que l'on puisse faire de différence nette entre les 2 classes sur les qualificatifs émis.

Dans aucun commentaire, la pathologie ne paraît déterminante.

L'hypothèse émise ici, repose sur les oppositions suivantes : Normal ou presque / Pathologie marquée/ altérations avec peu de souffle / altération un peu plus sévère sans prise en considération particulière du souffle.

#### 3.2.3.5.6.2. POUR LES NON EXPERTS

L'analyse des commentaires permet de décrire les classes de la façon suivante.

Une classe (A) déterminée par des voix de timbre présentant un score 0 ou 1 quelque soit le paramètre est exclusivement féminine. La référence au genre est quasi constante dans les commentaires. Les altérations de la voix sont souvent associées à des événements de la vie comme la fatigue, l'âge, le tabac... Une classe (B) déterminée par une fréquence plus grave avec des scores presque toujours à 0 (toujours pour le B) est très majoritairement composée d'hommes. La fréquence aggravée de la voix d'une femme porteuse d'un œdème explique la seule exception. Les commentaires comme pour la classe A explicitent le genre supposé des locuteurs et expliquent les anomalies par des événements « banaux » de la vie.

Une classe (E) dominée par des scores à 3 présente cependant une certaine hétérogénéité car à part pour le grade, des scores 0 ou 1 existent pour B et R. Les commentaires traduisent une « gêne » à

travers le serrage et le forçage présumé chez les locuteurs ou à travers le côté désagréable, douloureux à écouter. Les explications associées sont soit pauvres, soit en lien avec le repérage du caractère pathologique ou « pas banal ».

Les commentaires pour les non classés sont nombreux et hétéroclites. Ils traduisent la prise en compte de l'âge (enfant, adolescent, sujet âgé), de l'intonation (qualité présumée d'un lecteur ou narrateur) et des perturbations audibles mais gardant un caractère « naturel ».

L'hypothèse émise ici, repose sur les oppositions suivantes : Anomalies marquées sans explications « naturelles » / voix de femme « banales »/ voix d'hommes banales

### 3.3.4. RESUME-DISCUSSION

Cette seconde étude confirme deux pôles communs à tous les auditeurs : le degré de sévérité et la référence à une normalité par rapport à la présence de stimuli pathologiques.

Entre les deux si la question est abordée par le côté plutôt « normal », les références concernent l'identité des locuteurs et les informations que l'on peut extraire à propos des personnes parlantes (pour les non experts) sur la qualité de la voix (pour les experts). Abordé par l'entrée « altération sévère» des questions ou des hypothèses sur les causes (non experts) et les mécanismes (experts) semblent intervenir (figure n°3.3.34).

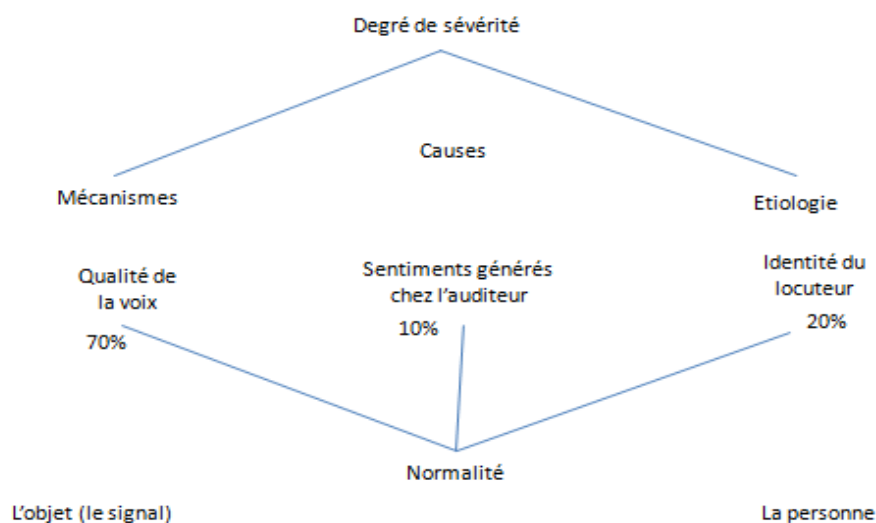


Figure n°3.3.34 : Hypothèse schématique des critères sous-tendant la perception des troubles de la parole pathologique

L'impression générale est celle d'une expertise qui se concentre sur l'objet voix en filtrant les informations sur la personne et une démarche non experte qui reste attachée à la personne en l'absence d'un vocabulaire ou d'une stratégie d'analyse spécifique au signal vocal.

### 3.4. COMPARAISON RESULTAT ETUDE 1 ET ETUDE 2

Pour mener cette comparaison nous ne prendrons en considération que les groupes d'auditeurs comparables des deux études (figure n°3.4.1).

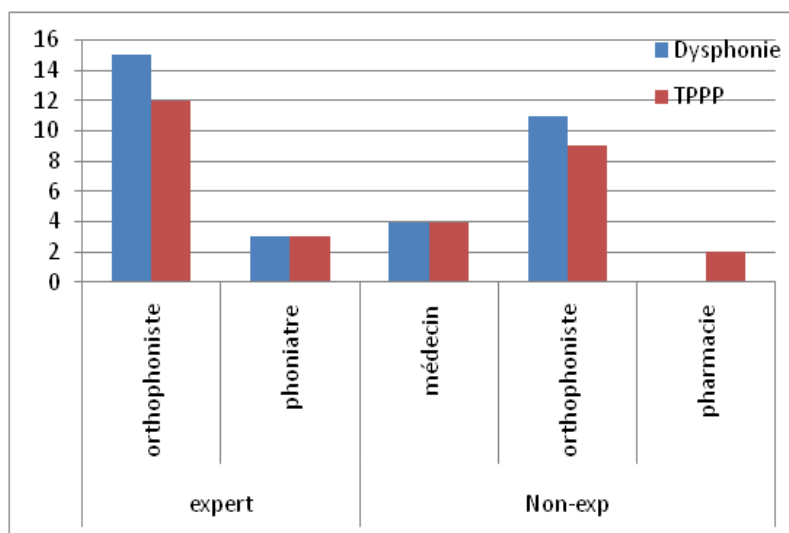


Figure n°3.4.1 : Répartition des auditeurs

La répartition entre les groupes d'auditeurs est comparable en dehors du fait qu'il y a relativement un peu plus d'orthophoniste dans le groupe des experts et un étudiant en pharmacie « à la place » d'un étudiant en orthophonie dans le groupe des non experts.

### 3.4.2. COMPARAISON DES RESULTATS DES CLASSES

Le nombre de catégories constituées est plus élevé pour les dysphonies que pour les TPPP.

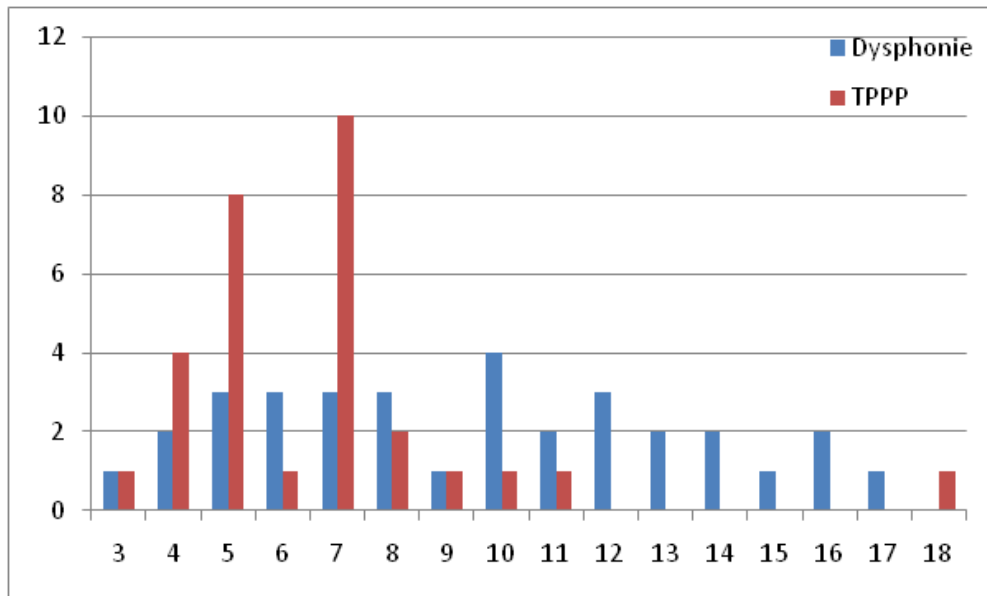


Figure n°3.4.2. : Le nombre de catégories constituées par auditeurs

L'analyse des arbres montre que l'on retrouve dans l'organisation des arbres une répartition des stimuli entre deux pôles identifiables par le degré d'altération du stimulus. Ces 2 pôles laissent supposer un processus de perception commun aux deux types de pathologies et aux différents groupes d'auditeurs étudiés avec à une extrémité de degré d'altération maximal et à l'autre, un pôle d'absence d'altération. Ces 2 pôles se traduisent par des classes clairement identifiables. Entre ces 2 pôles, des classes sont construites de manière différente en fonction des groupes d'auditeurs soit la constitution d'une classe n'a pas été possible. Il est remarquable que le nombre de stimuli non classé est beaucoup plus important dans la 2<sup>ème</sup> étude consacrée à la dysphonie (21 pour les non experts, 13 pour les experts versus 9 pour les non experts et 0 pour les experts).

---

### 3.4.2. COMPARAISON ENTRE L'ANALYSE DES COMMENTAIRES DU CORPUS COMPORTANT DES PATHOLOGIES VOCALES ET DU CORPUS COMPORTANT DES PATHOLOGIES DE LA PRODUCTION DE LA PAROLE.

---

#### 3.4.2.1. LES NOTIONS DE NORMALITE ET LES REFERENCES AU DEGRE D'ALTERATION

---

La notion de normalité est évoquée de manière forte et équivalente dans les deux pathologies (figure n°3.4.3).

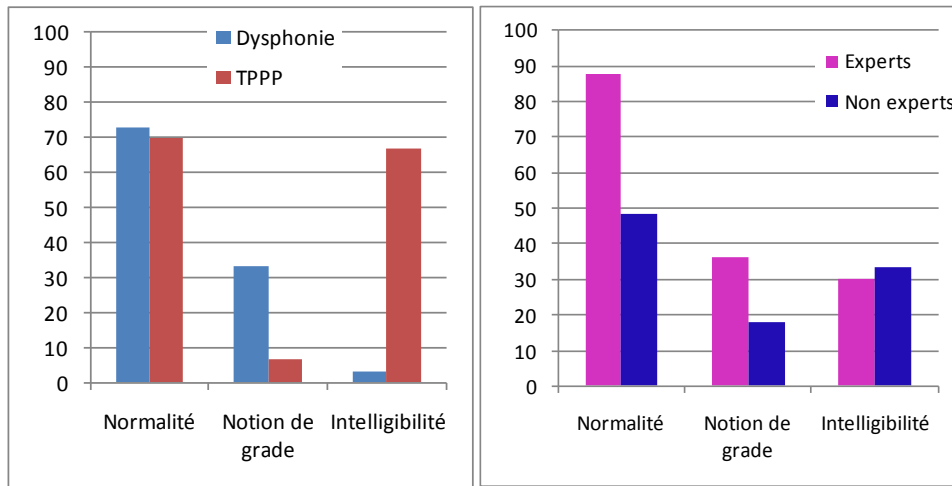


Figure n° 3.4.3 : Notions de normalité et les références au degré d'altération en pourcentage

Elle est probablement renforcée par la présence de stimuli « témoins » dans l'étude 1 et de voix non altérées sur le plan acoustique dans l'étude 2. La notion d'altération est très présente à travers la perte d'intelligibilité dans les TPPP et à un moindre degré dans les dysphonies par la notion de grade alors qu'elle est évidente à la lecture des arbres. En dehors de l'intelligibilité, ces notions sont prépondérantes pour les experts.

### 3.4.2.2. LES SUPPORTS DE REFERENCE

Les supports de référence sont logiquement réparti entre la parole pour les TPPP et la voix pour les dysphonies avec pour 30% des auditeurs une référence aux deux que ce soit à partir des TPPP ou des dysphonies (figure n° 3.4.4 et figure n° 3.4.5). Ceci est indépendant de la qualité des auditeurs (experts ou non experts).

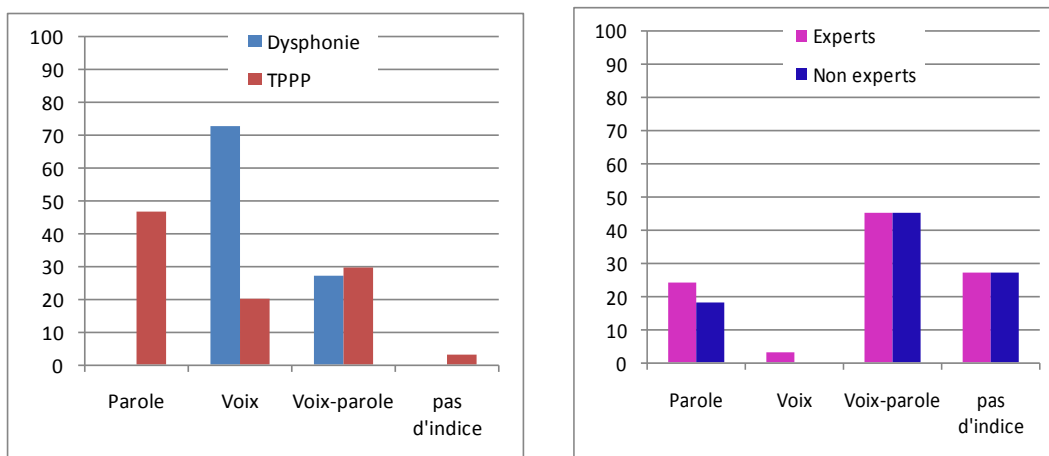


Figure n° 3.4.4 : Les supports en pourcentage



Les références à la qualité de la production, qui est dans tous les cas de la parole, quelque soit le contexte pathologique emploient des qualificatifs relatifs à la voix puis à la prosodie et enfin à l'articulation et ceci indépendamment de la qualité des auditeurs. A noter que la nasalité est également utilisée. La notion de forçage et de stabilité, moins fréquentes sont plus spécifiques aux dysphonies.

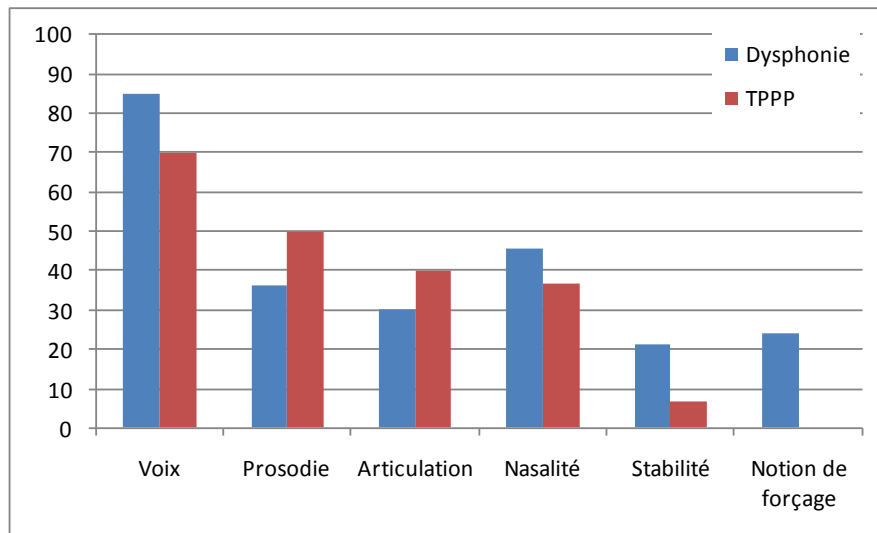


Figure n° n°3.4.5 : Niveaux évoqués par les qualificatifs utilisés pour la description de la production de parole dans les deux études en pourcentage

### 3.4.2.3. LA CLE DE TRI PRINCIPAL

La clé de tri principal (figure n° n°3.4.6 et n°3.4.7) authentifiée à travers les commentaires est de manière prépondérante la qualité de la voix avec un pourcentage de 60% pour les dysphonies et 33% pour les TPPP. Le mécanisme du trouble est aussi fréquent pour les TPPP et au deuxième rang pour les dysphonies. L'intelligibilité intervient ensuite pour les TPPP (16%) alors que ce sont les références au locuteur pour les dysphonies (9%).

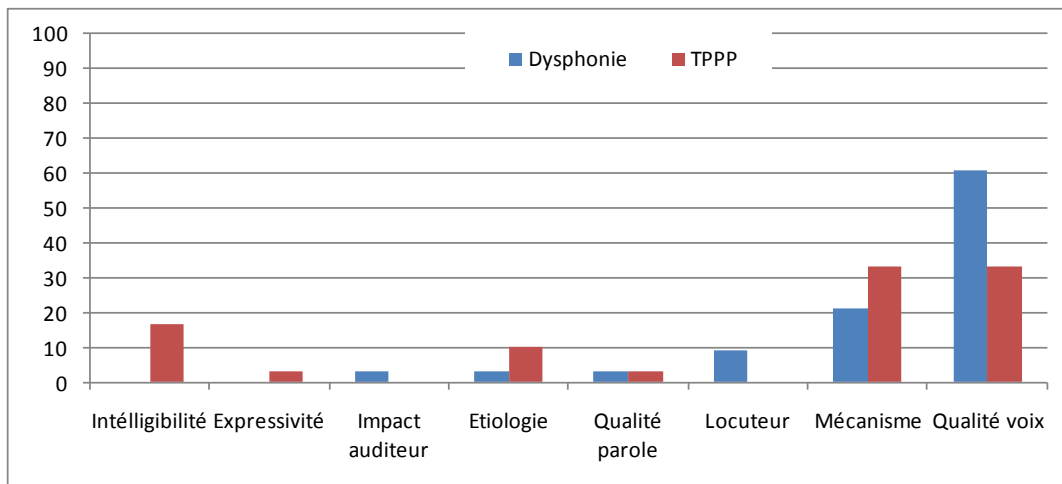


Figure n° n°3.4.6 : Clé de tri principal en fonction de la pathologie en pourcentage

On remarquera sans surprise que l'utilisation en tant que critère principal des notions de mécanisme et d'étiologie relève du domaine de l'expertise contrairement à la qualité de la voix ou à l'intelligibilité.

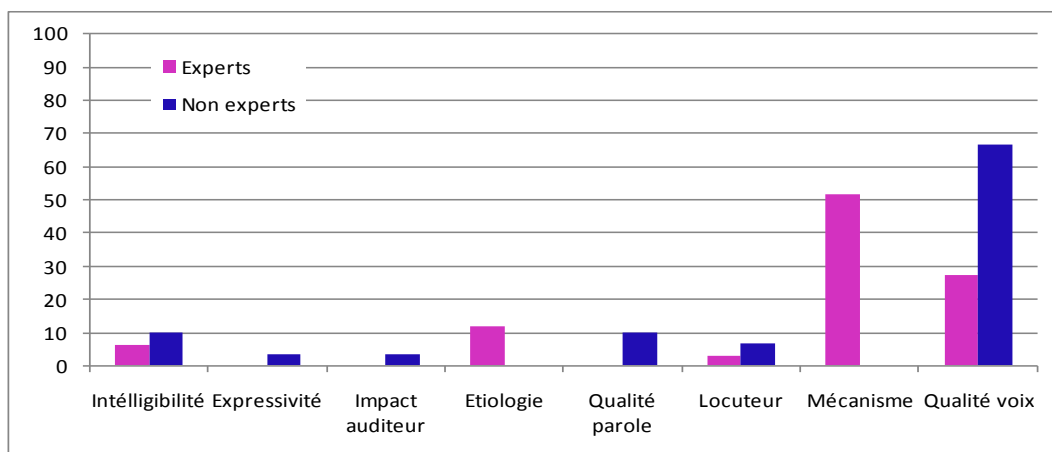


Figure n° n°3.4.7 : Clé de tri principal en fonction de la catégorie des auditeurs en pourcentage

#### 3.4.2.4. LES PRINCIPAUX CRITERES DE TRI RELEVES

Les principaux critères de tri retrouvés (figure n° n°3.4.8 et n°3.4.9) dans les commentaires sont les références à l'état de santé pour les TPPP alors qu'il s'agit de l'identité du locuteur pour les dysphonies. Les autres hypothèses sur l'état du locuteur sont dominées par l'état émotionnel ou l'état de fatigue. Elles sont présentes uniquement dans l'étude des dysphonies. L'état de santé est une préoccupation des experts à travers le mécanisme et l'étiologie alors que ce qui caractérise le locuteur (identité et état physique ou psychique) est partagé par tous les auditeurs même si cela concerne un peu moins les experts. Enfin malgré la professionnalisation, l'écoute des voix conduit les auditeurs à exprimer les sentiments que peut générer l'écoute d'une production de parole.

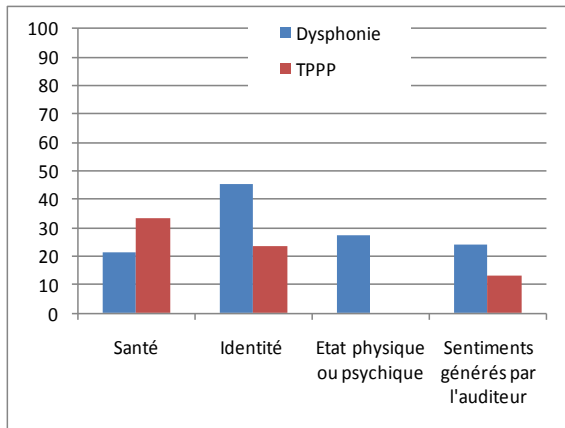


Figure n° n°3.4.8 Critères de tri en fonction de la pathologie (pourcentage)

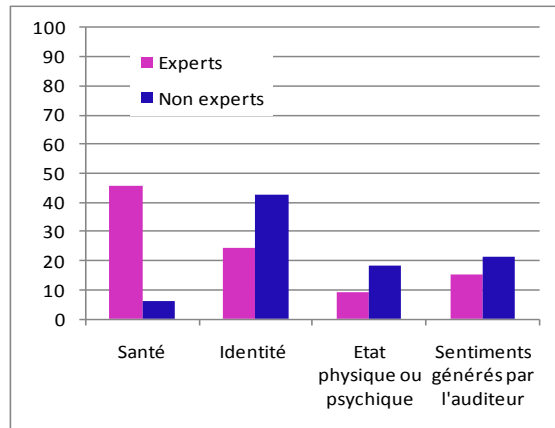


Figure n° n°3.4.9 Critères de tri en fonction des auditeurs (pourcentage)

En cas de référence à la pathologie, le mécanisme est plus fréquemment évoqué pour les TPPP alors que c'est l'étiologie pour les dysphonies (figure n° n°3.4.10).

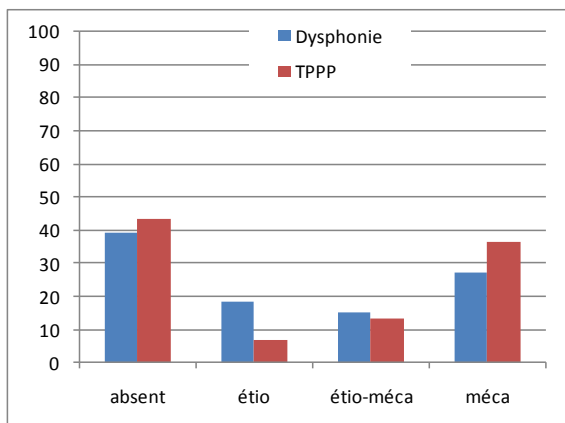


Figure n° n°3.4.10 En présence d'une évocation de la pathologie

En cas de référence aux sentiments que l'écoute génère chez les auditeurs, les sensations d'effort nécessaires pour écouter et le caractère agréable ou désagréable sont partagées dans les deux expériences (15 % dans les 2 cas). En revanche, la notion de souffrance est plus rare mais explicite exclusivement pour les dysphonies (9%).

### 3.5. DISCUSSION GENERALE DE LA CATEGORISATION

#### 3.5.1. PROPOSITION D'UNE HYPOTHESE

Ainsi nous proposons l'hypothèse d'une classification déterminée à une extrémité par une comparaison à un « prototype de normalité » et à l'autre extrémité par le degré d'altération de la production de parole (figure n°3.5.1).

Le prototype de normalité apparaît ici dans un contexte de présentation de différentes pathologies en présence de stimuli produits par des témoins ou acoustiquement non altérés. Il apparaît comme l'expression d'une norme par rapport à une dégradation dans le cadre d'une maladie. Il soulève la question de la spécificité des altérations d'un signal de parole lié à un trouble de la production.

Le degré d'altération de la production peut faire référence à la notion de sévérité. Celle-ci peut se traduire différemment en fonction du type de dégradation. L'intelligibilité en tant qu'indice de la sévérité est un concept retrouvé pour les TPPP aussi bien dans la construction des arbres que dans les commentaires. Intelligibilité et compréhensibilité sont alors interchangeable, ces termes renvoyant à la difficulté ressentie à accéder globalement au sens. La notion de gravité est clairement identifiable sur les arbres pour les dysphonies mais n'est pas explicitée dans les commentaires. Ceci correspond d'ailleurs à une problématique largement débattue dans la littérature; le grade relevant de l'intégration sensorielle de multiples paramètres acoustiques impossibles à traduire par le langage (Kreiman et al 2007, Giovanni et al 1992,1996, Dejonckere 2001, 2003, Wuyts et al 2000).

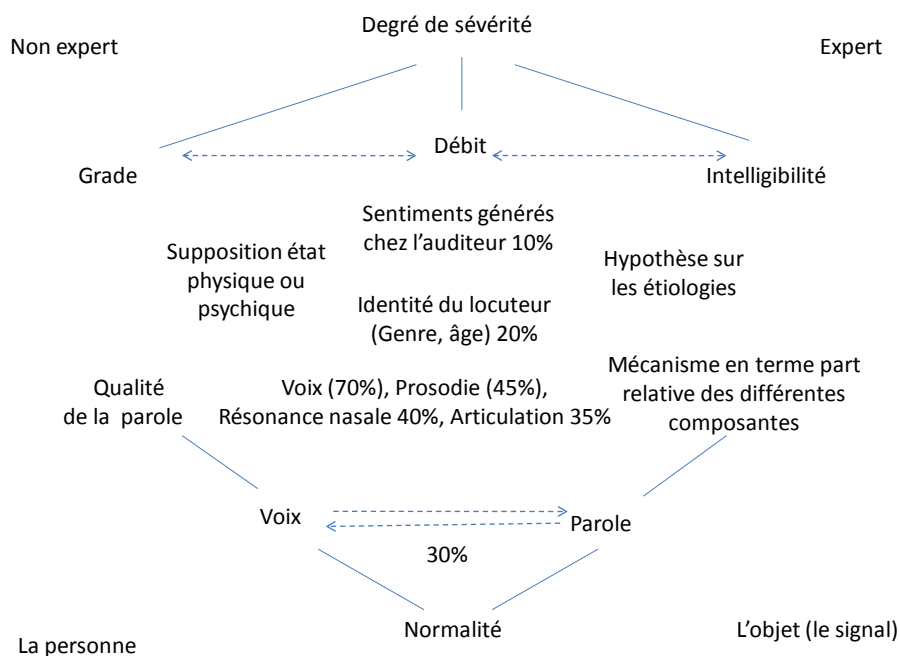


Figure n°3.5.1 :Hypothèse schématique des critères sous-tendant la perception des troubles de la parole pathologique.

En fonction de l'entrée suggérant un prototype de voix ou de parole nous n'utiliserions pas les mêmes critères. Bien que les stimuli soient tous des extraits de parole continue, l'entrée vocale est plus employée d'une part car elle est naturellement le support des troubles identifiés de la production vocale d'autre part car, pour les non experts, et parfois même, les experts, elle est utilisée devant un trouble de la production de la parole. Elle conduit à l'utilisation d'un critère de tri de qualité de parole dominé à 70% par la qualité de la voix, puis par la prosodie et la résonance nasale (45 à 40%) enfin par la qualité de l'articulation (35%). Ces critères sont partagés par tous les groupes d'auditeurs alors que les notions de forçage et de stabilité rarement exprimées, relèvent surtout des dysphonies et des experts. Cette dernière constatation évoque l'effet de la connaissance de l'échelle du GRBASI.

L'entrée parole, plus utilisée par les experts oriente vers une clé de type mécanisme qui peut être interprétée comme la responsabilité relative des différentes composantes dans la qualité de la production. Sont toujours présents quelque soit le type d'entrée ou les auditeurs, l'identité du locuteur (genre et âge) et les sentiments générés chez l'auditeur par l'écoute du stimulus. Ces critères prédominent pour l'entrée vocale et pour les auditeurs moins « professionnels » (proches>non experts>experts). Par l'autre pôle, les clés authentifiées par les paramètres étudiés dans ce travail sont l'intelligibilité subjective pour les TPPP et le grade pour les dysphonies mais également les aspects temporels. Les interactions entre les deux pôles semblent participer à l'élaboration d'hypothèses expliquant le trouble, plutôt avec la notion d'étiologie en réponse à une préoccupation professionnelle pour les experts et avec des suppositions sur l'état psychique ou physique (fatigue, tristesse...) du côté des non experts.

Cette analyse présente plusieurs limites. Nous en retiendrons principalement deux. La première est liée à une interprétation des classes restreinte aux seuls paramètres perceptifs et acoustiques étudiés qui pourraient être complétés. La seconde concerne l'absence de méthode d'analyse explicative des arbres eux mêmes. Cependant les méthodes compatibles avec l'établissement de relations entre les paramètres issus du calcul des arbres et des paramètres caractéristiques des stimuli nécessitent des heuristiques dont nous ne disposons pas.

Ces résultats malgré leurs limites, conduisent à discuter les trois points suivants :

1. La normalité et la spécificité des altérations pathologiques
2. La contribution aux notions d'intelligibilité, compréhensibilité et sévérité

3. La notion d'expert en tant que caractéristique recherchée pour déterminer les groupes d'auditeurs.

---

### 3.5.2. LA NORMALITE ET LA DEVIANCE PATHOLOGIQUE PAR RAPPORT A LA « NORMALE »

---

La notion de norme linguistique est largement débattue (Borel & Billières 1989). Posée comme « moyenne des divers usages d'une langue à une époque donnée, ou usage imposé comme le plus correct ou le plus prestigieux par une partie de la société » par Mounin (1993), nous retiendrons comme Baylon (1991 pp. 161-162), que la norme linguistique n'est qu'un aspect de l'ensemble complexe des normes sociales et qu'elle fonctionne dans une société comme un régulateur du comportement collectif. En phonétique, la prononciation est une construction sociale en permanente évolution de part les sujets parlants. Et face à la diversité des parlers, il n'y a pas de norme prescriptive (ou normative) sur laquelle nous pourrions nous appuyer.

Par contre, même si les critères de normalité sont difficiles à définir, le degré de déviance pathologique de la voix et surtout de la parole est un concept plus consensuel. Bunton & al (2007) dans un travail sur la reproductibilité de l'évaluation perceptive dans les dysarthries, confirme la meilleure reproductibilité des mesures pour tous les paramètres étudiés aux deux extrémités de l'échelle en 7 points utilisée. Ils suggèrent que les auditeurs sont d'accord pour définir ce qui est « normal » et ce qui est le plus déviant par rapport à la normale. Ce processus est donc également celui que l'on observe « spontanément » lors d'une tâche de catégorisation libre quelque soit le groupe d'auditeurs impliqué dans l'expérience. Ceci soulève deux questions sous jacentes.

Cette déviance par rapport à la normale dans le cadre des TPPP est elle spécifique à la pathologie ?

Rares sont les travaux qui comparent les troubles de la production non pathologiques présents dans le discours spontané et les troubles dans le cadre des TPPP. Danielle Duez, du fait de sa contribution à la recherche dans les deux domaines précise que toutes les anomalies ou erreurs produites dans le discours spontané sont celles retrouvées dans les TPPP mais c'est leur fréquence et surtout leur caractère non contrôlé qui est caractéristique (Duez 2007). Il est donc possible que cette spécificité suffise à générer chez les auditeurs une référence de normalité et ne permette pas de généraliser ce

processus à la simple augmentation contrôlée de fréquence des erreurs de production. Le fait que dans les expériences que nous venons de décrire, toutes les catégories d'auditeurs utilisent cette référence en présence de stimuli pathologiques est en faveur du caractère partagé de cette référence. Par contre, les variations quand à la composition des classes entre les groupes montrent que la frontière perceptive par rapport à cette référence est mouvante. L'extrême familiarité des proches paraît augmenter leur tolérance par rapport à cette déviance mais les rend très pertinent sur le repérage de l'altération de l'intelligibilité/compréhensibilité. Les experts qui présentent une autre forme de familiarité avec les locuteurs sont beaucoup plus sélectifs avec un probable « déplacement » de la pertinence pour la détection d'une production anormale ou corrigeable en fonction de la préoccupation inhérente à la sélection des auditeurs dans ce groupe.

La déviance maximale observable est elle de nature différente en fonction du type de trouble ?

Ce travail confirme que le degré de déviance est exprimé différemment à travers les commentaires entre l'étude comportant des TPPP et l'étude des dysphonies. Ainsi pour la parole, la déviance maximale fait référence à l'atteinte de l'intelligibilité au sens large du terme, c'est-à-dire qui gêne l'accès au sens. Alors que pour la voix, il n'y a pas de référence conceptualisée. L'intelligibilité est rarement évoquée. La notion d'audibilité est plus associée au TPPP. Elle n'a jamais été utilisée dans l'expérience avec les dysphonies. L'interprétation de nos résultats ne permet que des hypothèses basées sur les commentaires correspondant à la classe E. Par rapport aux autres classes, l'absence ou l'épuisement des explications relevant des événements banaux de la vie renvoie à la perte du caractère naturel décrit dans les dysarthries mais elle est associée ici à une souffrance supposée (pour le locuteur) ou vécue (par l'auditeur). Il semble de ce point de vue exister un seuil de « tolérance » à l'effort physique présumé nécessaire pour parler.

En simplifiant, nous pourrions opposer :

La difficulté que j'ai ou l'effort que je fais pour comprendre/ La difficulté que j'ai ou l'effort que je fais pour écouter, en faisant respectivement référence à la fonction conative et à la fonction émotive du schéma de communication proposé par Jakobson (figure n°3.5.2).

## Schéma de conversation



## Fonctions correspondantes

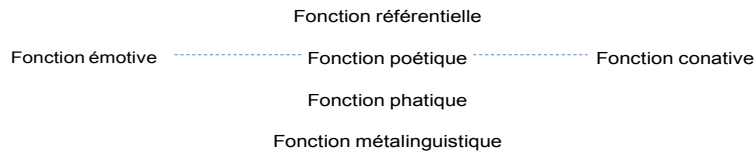


Figure n°3.5.2. Schéma de la communication de Jakobson

Ces résultats apportent des arguments en faveur des hypothèses issues des travaux de Bunton et al (2000) sur le rôle des paramètres prosodiques dans la détermination des TPPP. Ils renforcent également la notion de « sévérité perceptive » définie par Auzou (2007), qui prend en compte l'ensemble des « déviations » de la parole (qualité vocale, prosodie, intelligibilité...) en précisant le cadre conceptuel issu de l'absence de neutralité possible de l'auditeur face à l'écoute de la voix dans la parole. Cette absence de neutralité de la part des juges lors d'une tâche perceptive a été observée initialement face aux troubles de la fluence. Mais cette tendance dominée par l'attribution de traits de personnalité a ensuite été étendue à l'écoute de parole de personnes présentant un trouble de la communication quel qu'il soit (Williams & Dietrich, 1996,2001, Allard & Williams 2008). A noter également que l'expression des émotions et l'âge ont pu être exprimés sous la forme d'algorithme par des systèmes totalement artificiels (Oudeyer P. 2003).

---

### 3.5.3. CONTRIBUTION A LA DEFINITION DE L'INTELLIGIBILITE

---

L'usage des mots construits à partir des morphèmes « intelligible » ou « compréhension » dans cette expérience confirme que même les experts malgré leur niveau de formation emploient de manière interchangeable l'un ou l'autre. Ceci peut s'expliquer par l'absence de tâche explicitée : peut être cela serait –il différent dans un contexte de « test » de l'intelligibilité ? Cependant, même si une partie de la communauté, principalement les phonéticiens, partage une définition de l'intelligibilité en relation avec



les tests d'identification d'unité linguistique et de décodage n'est pas incompatible avec ce point de vue.

En effet comme nous l'avons démontré au chapitre 2, la définition de l'intelligibilité a un sens commun non différentiable de la notion de compréhension. En utilisant comme référence le processus d'accès au sens, il est concevable, en fonction de l'étage du processus auquel on s'intéresse que le niveau d'accès au sens qu'il représente puisse répondre à une terminologie spécifique. Mais ceci reste une terminologie par des groupes de personnes unis par une connaissance particulière du fait des bases théoriques en rapport avec le questionnement associé. Ainsi les frontières entre la notion de perception, d'intelligibilité, de compréhension et d'interprétation de la parole doivent être précisées pour éviter tout malentendu dès que l'on veut introduire une notion plus précise que celle d'un accès au sens du message.

Cette différence dans l'interprétation d'un mot en fonction de ses utilisateurs introduit naturellement la question de la notion d'expert.

---

### 3.5.3. LA NOTION D'EXPERT

---

Dans le domaine clinique des TPPP, la majorité des études convoquant des mesures subjectives de l'intelligibilité de patients font appel à des auditeurs experts : médecins, internes ou étudiants spécialistes des troubles de la communication et/ou du langage, voire les auteurs eux même (Maier A et al 2009, Whitehill 2002). Plusieurs travaux ont mis en évidence l'importance du niveau d'expérience des auditeurs dans la tâche de perception des TPPP (Williams & Watson 1985, Zyski & Weisiger 1987) faisant discuter la part de la familiarité autant que de l'expertise en ce qui concerne les scores d'intelligibilité (Beukelman & Yorkston 1980). Cette différence n'est pas confirmée notamment si un entraînement préalable à l'expérience est réalisé (Doyle et al 1989, Bunton et al 2007). Ceci pose bien la question de la différence entre le niveau d'expérience dans une pratique définie et le statut d'expert. Dans un chapitre intitulé « Expert ? Vous avez dit expert ? » Morange (2010) « essaie de dégager la complexité de cette catégorie en récusant la (trop) simple dichotomie entre expert et novice, associé au statut de connaissance savante du premier qui serait opposée à une connaissance « naïve » du second ». En optant pour le questionnement des connaissances expertes, acquises par les pratiques professionnelles, en particulier dans le champ de l'analyse sensorielle, sa réflexion concorde avec

notre problématique. Pour les besoins d'une étude linguistique et cognitive comprenant une étude acoustique de données sonores, Morange & Caudea (2009) ont pointé l'importance de la prise en compte des types d'expertises potentielles. Chaque sujet se situe dans une série d'ancrages selon qu'il est dans une pratique d'usager ou une pratique d'expert. Elle rejoint le point de vue développée par Dubois et al 1992 sur l'importance de ne pas catégoriser l'expert comme seul détenteur d'un savoir « vrai » dans un discours savant. Car tout sujet peut être expert en fonction de la perspective dans la quelle s'inscrit de même que tout expert ne l'est pas en permanence et doit faire avec « ses états d'âmes » et ses « états d'être ». Dans notre travail, la présence dans les commentaires, y compris des professionnels exprimant des sensations générés par l'écoute de la voix confirme que le « filtre » théoriquement associé au professionnalisme n'est pas infaillible ou absolu. Ceci est probablement lié au support émotionnel que représente la voix et la parole et au fait qu'elles véhiculent en plus des informations sur le locuteur.

Toujours en référence à la réflexion de Morange sur : qui est l'expert de référence ? , le sujet que l'on interroge et dont le ressenti intéresse le chercheur pour connaître et identifier les processus cognitifs qu'il met en œuvre ? Nous sommes effectivement face à une mauvaise utilisation du terme expertise dans la mesure où chaque groupe d'auditeurs a été déterminé par l'existence d'un ressenti sous-tendu par des processus communs distincts de part leur position ou relation avec les personnes souffrants de TPPP.

- Familiarité professionnelle avec une préoccupation diagnostique et/ou thérapeutique pour les professionnels
- Familiarité personnelle avec un statut relationnel régies par les règles sociales pour les « proches »
- Neutre ou vierge de contact en l'absence d'objectif précisé pour les étudiants.

Ainsi, professionnels expérimentés ou proches peuvent être experts avec des savoirs différents.

Comme pour Castellengo & Dubois 2005, sur une étude au sujet de la perception des timbres musicaux ou d'autres études sur le gout du vin, cela renvoie à la relation entre l'expérience perceptive des sujets et leur connaissance et l'usage du lexique qui conduit à opposer l'expérience individuelle d'un usager avec l'expertise négociée d'un groupe qui pourra donner une signification spécifique a des termes partagés avec un meilleur agrément (reproductibilité) pour ce groupe. Cette

mise en commun est peut être traduite dans notre travail par le nombre de catégories nettement moins élevé pour experts par rapport au non experts.

### **3.6. CONCLUSION**

---

Ces expériences de catégorisation libre sont en faveur d'un processus de perception des troubles de la production de la parole pathologique où l'intelligibilité de la parole est un critère déterminant dans la construction des arbres. Mais l'intelligibilité est ici utilisée au sens large et renvoie à la notion d'un degré d'altération ou de sévérité. Cela nous a conduit à exploiter la méthode d'analyse perceptive des stimuli de l'étude sur les TPPP pour étudier les rapports entre l'impression d'intelligibilité et celle de sévérité subjective.

*Les pressions évolutives qui ont façonné la conception et les capacités de nos cerveaux consistaient à trouver « la ligne de conduite qui est la meilleure pour soi », et la « recherche de la vérité » n'était qu'un facteur secondaire, subalterne (d'ailleurs clairement au service du premier)*

*Elkhonon Goldberg 2007, au sujet des mécanismes de cognition prescriptives p185*

## **4. INTELLIGIBILITE ET SEVERITE DANS LES TROUBLES DE LA PRODUCTION DE LA PAROLE PATHOLOGIQUE**

---

4.	Intelligibilité et sévérité dans les troubles de la production de la parole pathologique	167
4.1.	Introduction.....	169
4.1.1.	La sévérité d'un trouble, une notion mal définie en français.	169
4.1.2.	Réflexions à partir du modèle de la surdité et ébauche d'une définition	170
4.1.3.	Qu'en est-il dans les autres domaines ?	171
4.1.4.	Les notions de sévérité et d'intelligibilité dans les TPPP	173
4.1.5.	Objectifs de ce travail	174
4.2.	Population, Matériel et Méthode.....	174
4.2.1.	Les locuteurs	174
4.2.2.	Le corpus	175
4.2.3.	Les auditeurs	176
4.2.4.	La procédure	176
4.2.5.	Méthodologie	178
4.3.	Résultats.....	179
4.3.1.	Expérience 1	179
4.3.1.1.	Les scores de l'évaluation perceptive	179
4.3.1.2.	Les scores d'altération de l'intelligibilité et de sévérité	181
4.3.1.3.	Les corrélations entre les scores	185
4.3.2.	Expérience 2	188
4.3.2.1.	La distribution des scores de l'analyse perceptive	188
4.3.2.2.	La distribution des scores de sévérité et d'altération de l'intelligibilité	189
4.3.2.3.	Les corrélations entre les scores	189
4.3.3.	Expérience 3	190
4.4.	Discussion.....	194
4.4.1.	Les faits établis par cette expérience	194
4.4.2.	Quelles peuvent être les raisons de cette hétérogénéité ?	195
4.4.3.	En ce qui concerne la différence de score entre les deux notions, quelles peuvent être les explications ?	196
4.5.	Conclusion.....	198

## 4.1. INTRODUCTION

---

L'absence de consensus quant à la définition de l'intelligibilité par rapport à la notion de sévérité ou de compréhensibilité dans le domaine des troubles de la production pathologique de la parole est source de confusion dans la pratique Clinique. Ce défaut est entretenu par le découpage du domaine médicale par organe. Ainsi, l'intelligibilité du point de vue de l'audiologiste face à la surdité ne soulève pas le même questionnement Clinique que celui du neurologue face à l'aphasie ou à la dysarthrie ou celui de l'otorhinolaryngologiste face aux séquelles de cancer ou encore celui du pédiatre face aux difficultés d'apprentissage de la parole dans les séquelles de fentes faciales.

Nous avons déjà présenté les différents aspects de la notion d'intelligibilité et développerons ici la notion de sévérité d'un trouble.

---

### 4.1.1. LA SEVERITE D'UN TROUBLE, UNE NOTION MAL DEFINIE EN FRANÇAIS.

---

Labastrou (2000) a réalisé au Canada une recherche documentaire sur le terme sévère dans le domaine médical. Elle commence son article par la remarque figurant à la fin de la définition du mot sévérité du Grand Robert quant au sens de gravité, de caractère dangereux : « ce sens comme celui de l'adjectif qui lui correspond, est critiqué comme un anglicisme abusif ; il est néanmoins ancien dans l'usage médical ». Elle conclut que « les traducteurs ont souvent raison de traduire *severe* par *grave*, adjectif qui reste d'usage courant en médecine, mais la tendance actuelle semble en faveur de l'emploi de sévère pour désigner le plus haut degré d'intensité d'une affection, d'une atteinte ou d'une manifestation indésirable. Cet adjectif fait aussi partie de syntagmes désignant des formes particulières de maladies... Enfin, il existe d'autres adjectifs, moins usités, qui peuvent parfois servir de solution de rechange : intense, sérieux, important, violent »

Lors de la consultation du « MeSH » (*Medical Subject Headings*, thésaurus biomédical de référence et outil d'indexation, de catalogage et d'interrogation des bases de données de la *National Library of Medicine*, Bethesda, USA), qui comprend MEDLINE/PubMed), le terme « *Severity* » renvoie à « *Severity of Illness Index* ». Ce terme, introduit dans la base en 1989 est explicité comme suit sans

réelle définition, insistant sur la nécessité d'une mesure: "*Levels of severity of illness within a diagnostic group which are established by various measurement criteria*".

Cependant, la consultation en ligne du MeSH bilangue français anglais vient pondérer les conclusions précédentes de Mme Laboustra. En effet, L'Inserm, partenaire français de la NLM, a traduit le MeSH à l'intention des utilisateurs francophones en 1986 et le met à jour chaque année sur <http://mesh.inserm.fr/mesh/index.htm>. L'entrée « gravité » envoie au descripteur français « indice de gravité médicale » correspondant au descripteur américain « Severity of Illness Index », en revanche, l'entrée « sévérité » reste caduque !

Quoiqu'il en soit, la traduction de la définition de « Severity of Illness Index » du MeSH en tant que niveaux de gravité d'une maladie dans un groupe diagnostic établis par différents critères de mesure » correspond à la pratique et sous entend bien que les critères déterminant la gravité sont spécifiques à chaque maladie.

---

#### 4.1.2. REFLEXIONS A PARTIE DU MODELE DE LA SURDITE ET EBAUCHE D'UNE DEFINITION

---

Le domaine de la surdité peut servir comme base de réflexion. Le degré de gravité ou sévérité est suffisamment consensuel pour être utilisé dans le domaine public. Ainsi la Haute autorité de Santé ([www.has-santé.fr](http://www.has-santé.fr)) dans les recommandations sur « surdité de l'enfant : accompagnement des familles et suivi de l'enfant de 0 à 6 ans » utilise un critère de sévérité proposée par le BIAP (bureau international d'Audiophonologie). Le degré de sévérité est défini par le seuil minimal de perception des sons par la meilleure oreille exprimée en dB HL (décibel Hearing Level). Elle est qualifiée de légère entre 21 et 40 dB, moyenne entre 41 et 70 dB, sévère entre 71 et 90 dB, profonde au-delà de 90 dB. Ainsi, la mesure du déficit de perception des sons purs en termes d'intensité obtenue par la moyenne sur 4 fréquences a été classée en 4 degrés en fonction des conséquences sur la perception de la parole. Cette classification est utilisée car elle a des applications pratiques notamment sur les modalités de prises en charge (indications d'interventions thérapeutiques comme l'indication de geste chirurgical simple de mise en place de drains transtympaniques dans les otites séromuqueuses de l'enfant dans les surdités légères et moyennes, des dispositifs plus complexes comme la pose d'implant cochléaires dans les surdités de perception sévères et profondes). En effet, depuis les débuts de l'audiométrie, les cliniciens et chercheurs se sont attachés à déterminer le pronostic sur

l'intelligibilité de la parole à partir de la mesure tonale conduisant à cette classification (Fournier 1951, Lafon 1964). Enfin, les implications industrielles dans le domaine diagnostique et thérapeutique ont favorisés puis imposés des travaux de normalisation des tests. Ceci n'est pas le cas dans les TPPP. Ce rappel sur le modèle de la surdité et la recherche terminologique préalable conduit à faire l'hypothèse :

1. que la sévérité d'un trouble fait référence :

- a. Au niveau d'intensité du déficit qu'il soit fonctionnel (ex surdité) ou lésionnel (ex traumatisme du nerf de l'acoustique)
- b. Au niveau de risque que représente ce déficit sur la ou les différentes composantes du système dans lequel s'inscrit la fonction ou la lésion en cause

2. que la classification qui en découle aura une utilité dans le sens où elle aura des applications pratiques pour des groupes d'utilisateurs.

En synthèse pour la surdité, la mesure du niveau de perception des sons classée par rapport au risque de ne pas pouvoir percevoir la parole, en référence à la fonction de communication dans un système modélisé selon la théorie de l'information de Shannon, permet de proposer des algorithmes de prise en charge.

---

#### 4.1.3. QU'EN EST-IL DANS LES AUTRES DOMAINES ?

---

Pour les troubles de la voix, la recherche d'une mesure du niveau de déficit reste problématique. La plainte ne se réduit pas à l'altération de la qualité de la voix mais en une restriction dans l'usage de la voix. Ainsi, utiliser l'impact sur le signal acoustique, potentiellement lié aux difficultés d'émission ne reflète pas l'étendue du déficit.

De plus, l'aspect multidimensionnel de la perception de la voix n'a pas permis de trouver un système satisfaisant pour mesurer la qualité de la voix (Kreiman & Gerratt 2009, Giovanni et al 1996, Dejonckere et al 2000,2003, Wuyts et al 2000). Une des raisons, comme le souligne Ron Baken (2000) dans l'introduction à son ouvrage sur la mesure clinique de la parole et de la voix, est l'absence de modélisation des combinaisons des paramètres répondant à la perception : « a major problem with simple listening as a diagnosis method is that the auditory system is inherently configured for dealing with the speech signal as a whole entity and for detecting linguistically relevant features in it. It is this highly evolved aptitude that makes the auditory system so efficient in speech communication.



Unfortunately, from the speech pathologist's point of view, auditory processing often does not leave the listener with a conscious awareness of the acoustic details that have combined to generate a given perception".

Enfin, le risque par rapport au quel doit être confronté un paramètre pour qu'il devienne un indice de sévérité n'est pas évident. En référence à la fonction de communication, les auteurs (ref Dejonckere...) proposent le terme d'audibilité en opposition à celui d'intelligibilité bien que l'on reste dans le domaine de la perception de la parole, mais cette fois ci, perçue par autrui. Peut être est ce simplement que l'impact sur la perception de la parole est trop « faible » et que le problème reste conscrit à la personne en terme de limitation des performances vocales. De ce fait, le degré d'altération du signal sonore pourrait simplement ne pas être le bon indice.

Dans tous les cas, en pratique sont utilisées des échelles subjectives de sévérité dont la validité a été jugé satisfaisante pour assurer l'évaluation de la qualité de la voix en pratique clinique (Dejonckere et al 1993, De Bodt et al 1997). Ces échelles souvent ordinales, reposent sur plusieurs paramètres jugeant subjectivement la qualité de la voix, le paramètre global de gravité étant le plus reproductible. Cependant, l'importante variabilité observée dans les études scientifiques incite à poursuivre les recherches avec la création d'indices composites (Wuyts et al 2000, Yu et al 2001) ou d'autres méthodes telle que des tâches d'ajustement synthétique (Kreiman & Gerrat 2007). Cependant, pour l'instant, aucun de ces indices ne débouche sur des applications pratiques, probablement par absence de prédiction sur un risque ou un handicap qui supporterait un algorithme décisionnel.

En ce qui concerne les TPPP, bien que l'indexation de l'intelligibilité pour mesurer la sévérité du trouble soit reconnue de manière consensuelle, aucune norme n'a été éditée. Pourquoi n'a-t-on pas, comme dans la surdité, de classification en pratique clinique ?

En France, le score d'intelligibilité (SI), a été proposé en 1998 par Auzou (1998) à partir de l'adaptation française du Frenchay Dysarthria Assessment (Enderby 1983) en faisant la somme obtenue à 3 épreuves : la lecture de 10 mots, la lecture de 10 phrases et de la parole spontanée. Chacune de ses épreuves étant coté de 0 à 8, un sujet est considéré comme dysarthrique si son score d'intelligibilité est inférieur ou égal à 23, la sévérité de la dysarthrie peut être classée en fonction du score en légère entre 23 et 18, modérée entre 17 et 12, sévère entre 11 et 7 et très sévère en dessous de 6. Cependant, sa validité n'a pas fait l'objet de travaux spécifiques et elle n'a pas d'application pratique formalisée dans la prise en charge.

Est utilisé principalement par l'équipe créatrice du test, le SIT « *Sentence intelligibility test* » (SIT) (Yorkston & Beukelman 1996) avec dans des travaux de recherche la création de degré de gravité en fonction des scores de léger entre 76% à 95%, modéré de 46% à 75%, de sévère de 25% à 45% de profond de 5% à 24% (Hustad 2008). Utilisé comme indice de sévérité, ce test a servi à démontrer qu'il n'y avait pas de relation significative entre intelligibilité et compréhensibilité quand l'effet de la sévérité était supprimé. Ceci conduit à aborder de manière plus approfondie la notion de sévérité dans les TPPP par rapport au concept d'intelligibilité.

---

#### 4.1.4. LES NOTIONS DE SEVERITE ET D'INTELLIGIBILITE DANS LES TPPP

---

En se référant au niveau d'intensité de l'atteinte, il apparaît qu'une partie de l'étendue de la gêne induite par l'atteinte ne soit pas couverte par le défaut d'intelligibilité.

Ainsi, le concept de sévérité est plus étendu que celui d'intelligibilité. Van Riper et Emerick (), au sujet de locuteurs malentendants, expliquaient que "*speech is defective when it is conspicuous, unintelligible or unpleasant* ». Pour d'autres auteurs impliqués dans la dysarthrie, un trouble de la parole peut aussi être défini avec des termes comme : normalité, naturel, acceptable, bizarre, le terme de bizarre étant issu de travaux fondateurs de Darley sur la dysarthrie. Cependant le travail de Southwood and Weismer (1993) a démontré une forte corrélation entre les 4 notions, et l'expérience de Yorkston (1992) a mis en évidence la préférence d'utilisation par les personnes impliquées dans les TPPP du "caractère naturel". Ceci a conduit à conserver ce dernier terme pour une partie de la composante non segmentale de la sévérité. L'équipe de Yorkstone préconise également pour le suivi des patients l'association au test d'intelligibilité « SIT » la mesure du débit de parole ajoutant un indice temporel métrique à la notion de sévérité dans les TPPP.

Du point de vue du niveau de gravité, les TPPP sont mieux lotis que les troubles de la voix. Mais cela ne suffit pas à obtenir un indice de gravité utilisé en pratique. Peut-être que le manque de conceptualisation des prises en charge thérapeutique n'a pas stimulé la clarification de ce problème ou tout simplement est ce le manque de connaissance scientifique et donc de moyen de modélisation ? Des arguments concernant cette dernière hypothèse émergent des travaux de Hustad. Ces derniers mettent en évidence le rôle des différents contextes théoriques dans l'approche des TPPP chez les Infirmes moteurs d'origine cérébrale avec une évolution entre les modélisations selon la théorie de l'information de Shanon (1948) et les théories de compréhension du discours (1983) (cf

section 1.3.). Malgré la complexité des approches, il paraît important de clarifier certaines bases qui paraissent acquises et les faire évoluer pour élargir les perspectives.

---

#### 4.1.5. OBJECTIFS DE CE TRAVAIL

---

C'est dans cet esprit que, à la suite de l'expérience menée pour classer les stimuli de l'expérience de catégorisation libre, objet du chapitre précédent, j'ai essayé de déterminer la meilleure mesure de référence à utiliser. Partant de la routine clinique, j'ai demandé à mes collègues avec qui nous avons validé un test d'intelligibilité (Dittner et al 2010) de créer un jury d'expert. Comme nous travaillons ensemble depuis des années et que nous avons déjà suspecté certaines discordances, je n'ai pas été étonnée mais fortement intriguée par nos différences de comportement par rapport à la tâche que j'avais élaborée. De ce fait, la simple manipulation complémentaire par rapport à l'expérience de catégorisation libre est devenue une expérience à part entière.

L'objectif de ce travail est d'étudier comment un groupe d'expert réalise une tâche de jugement de sévérité par rapport à une tâche de jugement d'altération de l'intelligibilité, l'hypothèse émise étant que les experts sont capables de faire une différence après « calibration » par une démarche perceptive analytique, entre un jugement d'altération du signal acoustique de parole et un jugement de degré de compréhensibilité du message sur un segment de parole connu (corpus de même contenu sémantique) et sur un segment de parole inconnu (corpus de contenu sémantique différent).

## 4.2. POPULATION, MATERIEL ET METHODE

---

---

### 4.2.1. LES LOCUTEURS

---

Trente trois locuteurs adultes ont été enregistrés, 13 témoins et 20 patients présentant un trouble de la production pathologique de la parole. Ces patients présentaient des pathologies diverses dont la répartition est la suivante :

- Séquelles de traitement d'une tumeur ou d'une malformation des voies aérodigestives (9)
  - o Séquelles de traitement de cancers (8)

- Séquelles d'une fente faciale (1)
- Séquelles de pathologies neurologiques (3)
  - Infirmité d'origine cérébrale (1)
  - Maladie de Kearns dominé par un syndrome cérébelleux (1)
  - Paralysie des paires crâniennes (1)
- Pathologies neurologiques dégénératives (8)
  - Sclérose en Plaques (1)
  - Syndromes extrapyramidaux (7).

Les enregistrements ont été réalisés consécutivement lors des consultations pour bilan d'un trouble de la production de la parole à l'unité de la voix et de la déglutition dans le service ORL de l'hôpital universitaire Larrey Ranguel à Toulouse. L'entourage des patients et le personnels soignants ayant pendant la période de l'étude accepté d'être enregistrés constituent le groupe de témoin. Les enregistrements ont été effectués après information et obtention du consentement des patients.

---

#### 4.2.2. LE CORPUS

---

Deux corpus ont été constitués à partir des productions des 33 locuteurs :

Le premier est le même que celui utilisé au chapitre 3. Il comprend la dernière phrase du texte « Pierrot » de Maupassant, utilisée en pratique dans la version de 1998 de l'ECD (Evaluation clinique de la dysarthrie, Auzou Pet al. 1998.) : « Et elles s'épouvantaient pour l'avenir, comment dormir tranquillement maintenant ? ». Le choix de cette phrase a été déterminé par le fait que cette partie du texte a fait l'objet de moins d'hésitations liées à la lecture.

Le second est composé de phrases issues du corpus utilisé dans les tests HINT (Hearing in noise Test, Nilsson et al 1994). Il a été constitué secondairement pour disposer chez les mêmes locuteurs de stimuli sémantiquement différents.

Le texte original et les phrases retenues parmi les listes du HINT sont présentées en annexe 1.

Ce corpus a été enregistré avec un micro serre-tête Shure®, une carte d'acquisition Edirol® et un ordinateur Mac Intosh®. Les segments choisis ont été extraits avec le logiciel PRAAT (Paul Boersma & David Weenick 2009) : Praat : doing phonetics by computer version 5.1.0 computer program] Retrieved May 1, 2009, from <http://www.praat.org/>) sous la forme d'un fichier au format .wav. Les fichiers obtenus, ont été utilisés comme stimuli pour cette expérience. Aucun traitement ultérieur n'a

été effectué sur ces enregistrements, en particulier, aucune normalisation en intensité n'a été faite. En effet, la différence d'intensité entre les locuteurs fait entièrement partie des caractéristiques de la parole pathologique.

---

#### 4.2.3. LES AUDITEURS

---

Les auditeurs correspondent à un jury d'experts féminins constitué de 2 phoniatres et 3 orthophonistes habituées à juger des productions de parole pathologique. Travaillant dans le même hôpital depuis plus de cinq ans et ayant déjà participé à un jury d'écoute ensemble, ces experts ont cependant tous eu une formation initiale dans des universités différentes et des villes de France différentes. L'âge moyen est de 42 ans avec des extrêmes de 30 et 52 ans.

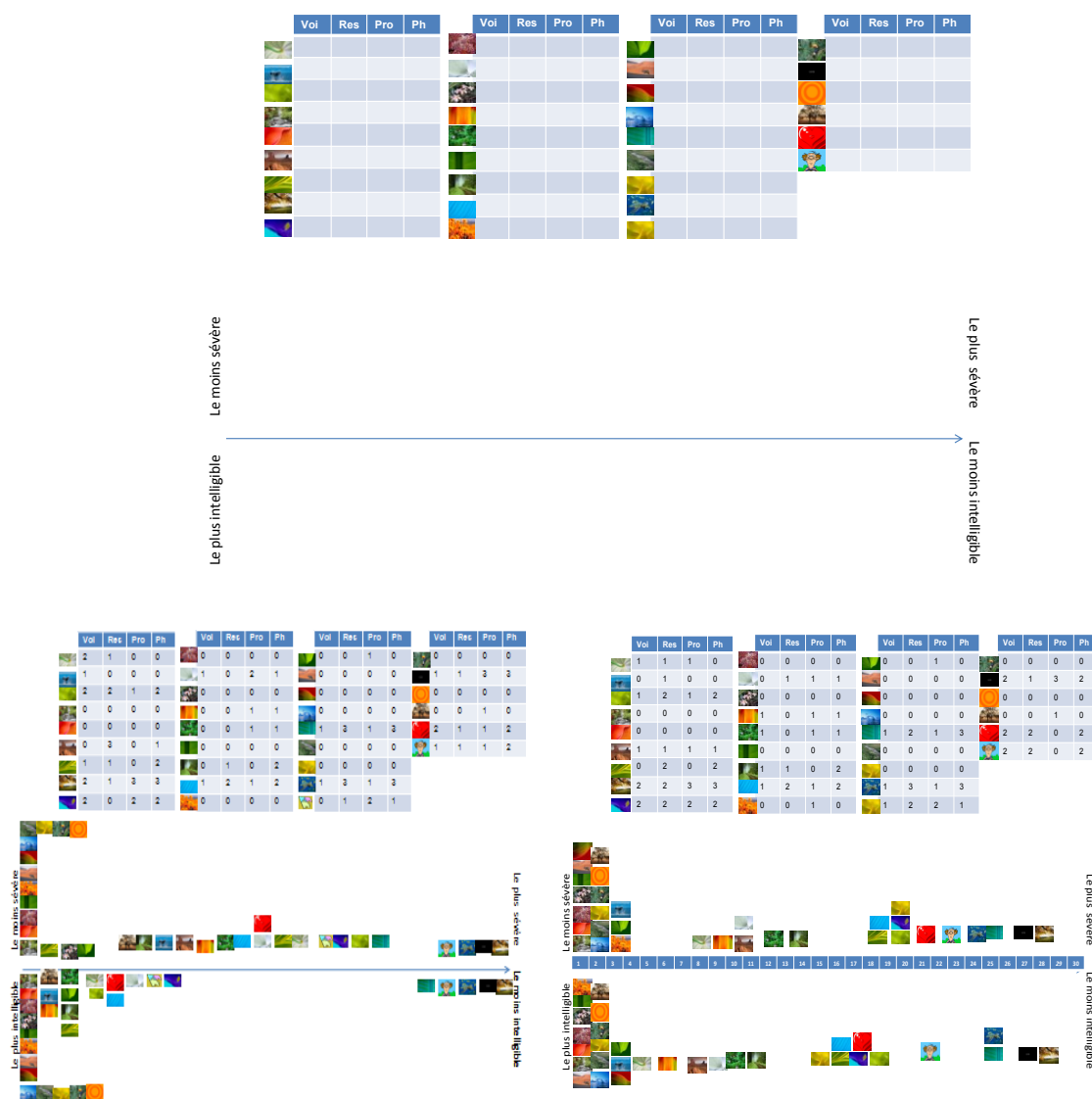
---

#### 4.2.4. LA PROCEDURE

---

Les 33 stimuli obtenus après l'extraction de la phrase choisie, ont été présentés sous la forme d'icone sur une diapositive du logiciel power point en regard d'une ligne de tableau permettant de mettre la cotation correspondant au son audible par un simple clic. A la partie inférieure de la diapositive une droite correspondant à 30 fois la taille d'une icône a été tracée. Orientée vers la droite, le côté gauche était annoté en haut « Le moins sévère » et en bas « Le plus intelligible », le côté droit était annoté en haut « Le plus sévère » et en bas « Le moins intelligible ». Une copie d'écrans de l'expérience est présentée figure n°4.2.1.

A : avant le début de l'expérience.



B : 3<sup>ème</sup> expérience du juge 3 (st. identiques) C : 3<sup>ème</sup> expérience du juge 4 (st. différents)

Figure n°4.2.1 = Copie de l'écran au cours de l'expérience.

A l'aide d'un « copier-coller », l'icône du son après avoir fait l'objet d'une analyse perceptive pouvait être copiée et placée au dessus (pour la sévérité), en dessous (pour l'intelligibilité) de cette flèche pour quantifier le degré d'atteinte.

La tâche demandée consistait donc en l'écoute autant de fois que nécessaire des 33 enregistrements et de les scorer à l'aide d'une échelle catégorielle à 4 niveaux de 0 à 3 (0 pas d'altération, 3 altération importante) sur la qualité de la voix, la qualité de la résonance, la prosodie et la qualité de la production phonémique. Cette méthode d'évaluation a fait l'objet de publications et a permis à De

Bodt MS et al (2002) d'identifier l'intelligibilité comme une combinaison linéaire de ces 4 dimensions. Cette tâche renvoie à une pratique professionnelle hebdomadaire voire quotidienne pour tous les membres de ce jury.

Puis, une fois cette tâche terminée, de placer sur une échelle visuelle analogique l'icône de l'enregistrement en fonction de l'intelligibilité de la parole d'un côté et de la sévérité de l'atteinte de la parole de l'autre. Les définitions utilisées pour cette expérience figuraient dans les consignes lues au préalable par l'expert et disponibles à tout instant. « Vous disposez de 33 stimuli dans la partie supérieure de l'écran. Vous pouvez les écouter tant que vous voulez. Votre tâche consiste à les classer en terme d'intelligibilité et de sévérité après avoir défini pour chacun entre 0 (normal) et 3 (atteinte sévère) le degré d'altération des quatre paramètres suivants : la qualité de la voix, la résonance, la prosodie, les altérations phonémiques. L'intelligibilité est définie comme la compréhensibilité du message véhiculé par le signal, alors que la sévérité est définie comme le degré d'altération globale du signal sonore. En pratique, copier-coller l'icône du son et glisser la copie le long de l'échelle en haut pour la sévérité, en bas pour l'intelligibilité, après avoir rempli les cases en rapport avec la gradation des quatre paramètres préalablement décrits. »

L'analyse préalable de la qualité des stimuli avait pour but de mettre les juges dans une démarche analytique et de les « calibrer » dans une approche similaire avant d'aborder la tâche suivante.

---

#### 4.2.5. METHODOLOGIE

---

Ce travail a été réalisé en 3 étapes justifiées par les résultats obtenus à chacune d'entre elles.

Expérience1 :

Le corpus est composé des stimuli de l'expérience de catégorisation libre c'est-à-dire des stimuli identiques composés de la phrase. :« Et elles s'épouvantaient pour l'avenir, comment dormir tranquillement maintenant ? »

L'analyse est conduite juge par juge avec une comparaison de la distribution sur des histogrammes entre sévérité et intelligibilité puis une comparaison entre les juges des comportements de jugement sur un graphe de cellule. Enfin des graphes bivariés permettent de visualiser la différence de répartition des scores entre sévérité et altération de l'intelligibilité. En raison du faible nombre d'experts ces données sont traitées de manières exclusivement descriptives.

La recherche de facteurs explicatifs à partir de l'analyse de la qualité des stimuli avec la tâche de

jugement a été réalisée par l'étude des coefficients de corrélation. Les valeurs présentées sont au risque 5%.

Expérience 2 :

La même expérience est reconduite 3 mois après avec des stimuli tous différents, chaque patient disant une phase différente.

L'analyse a été conduite de la même manière sans renouveler la recherche de facteurs explicatifs.

Expérience 3 :

Les deux expériences précédentes ont été reproduites sur une même session 6 mois après sans renouveler la recherche de facteurs explicatifs.

Le traitement des données est fait à partir d'une échelle correspondant à 30 unités, l'unité correspondant à la taille de l'icône utilisée pour écouter les sons. 1 correspond à une bonne intelligibilité ou une faible sévérité, 30 à la sévérité maximale ou à la plus mauvaise intelligibilité.

### 4.3. RESULTATS

---

---

#### 4.3.1. EXPERIENCE 1

---

##### 4.3.1.1. LES SCORES DE L'EVALUATION PERCEPTIVE

---

La distribution des scores de l'analyse préalable sur la qualité des stimuli concernant la qualité de la voix, de la résonance, de la prosodie et de l'articulation en terme de production phonémique est représentée pour chaque juge figure n°4.3.1.

De grandes variations demeurent dans la cotation des caractéristiques des stimuli. De manière décroissante, les juges 2, 3 et 5 ont moins de jugements anormaux sur la voix par rapport aux autres paramètres, le juge 3 moins de jugement anormaux sur la résonance et les juges 4 et 5 moins de jugement anormaux pour la prosodie.



Stimuli identiques-Expérience 1

Stimuli différents-Expérience 2

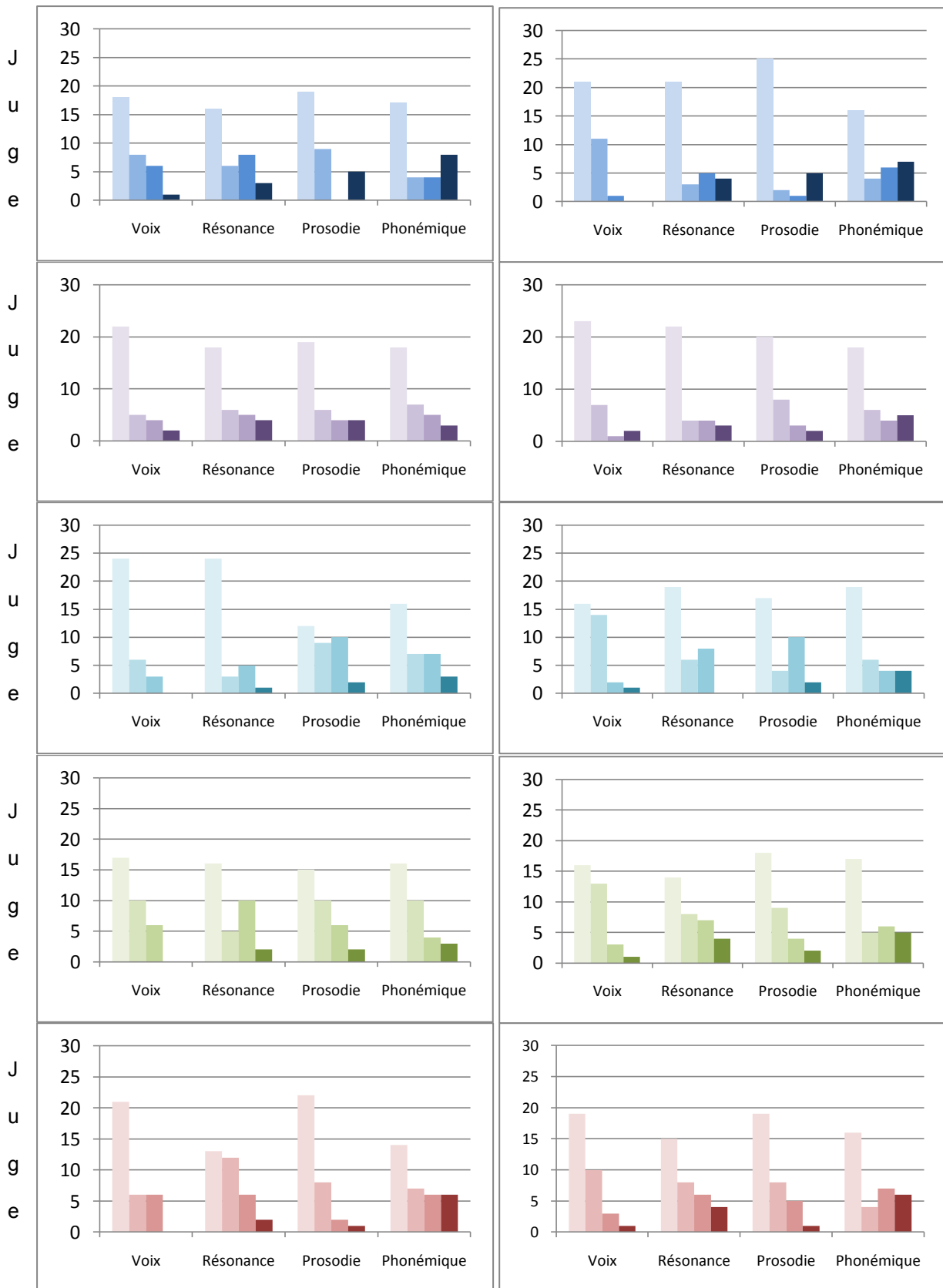


Figure n°4.3.1 : Profil pour chaque juge de la tâche d'analyse perceptive précédant le jugement A gauche, expérience 1 (stimuli identiques), à droite expérience 2(stimuli différents).La barre la plus claire correspond au score 0, la barre la plus foncée au score 3.

#### 4.3.1.2. LES SCORES D'ALTERATION DE L'INTELLIGIBILITE ET DE SEVERITE

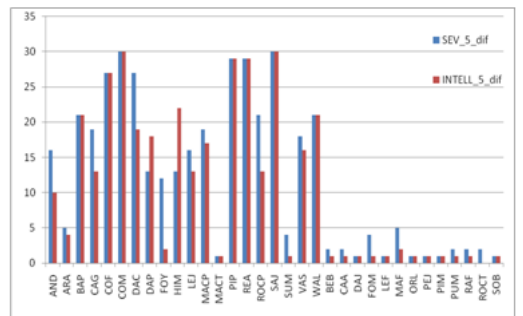
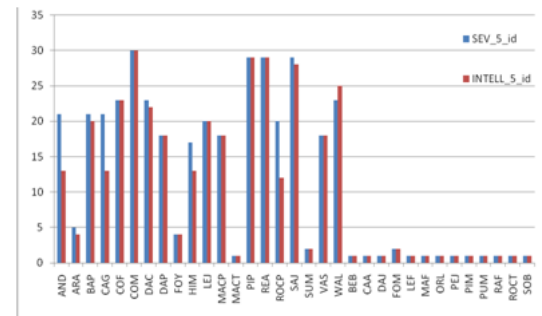
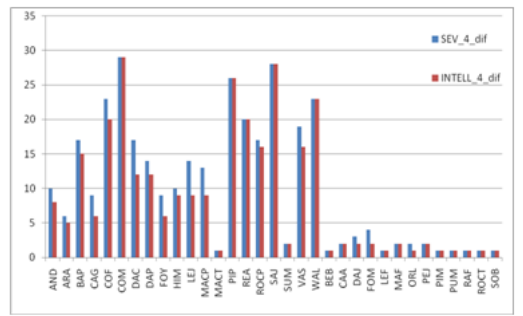
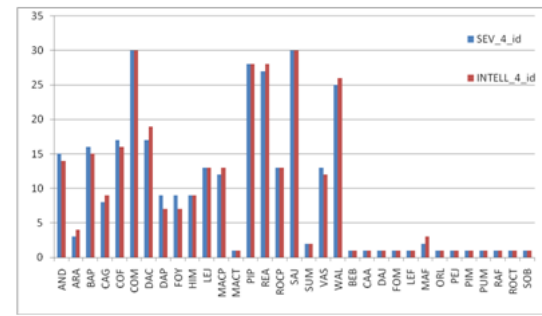
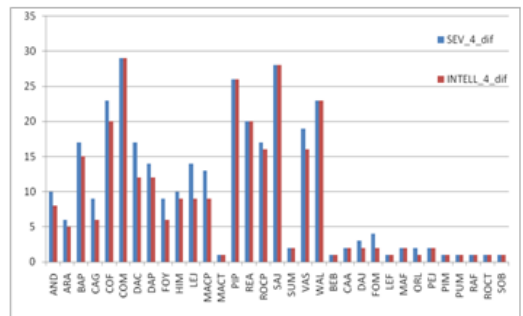
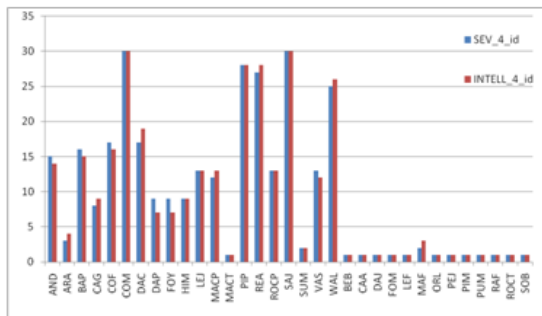
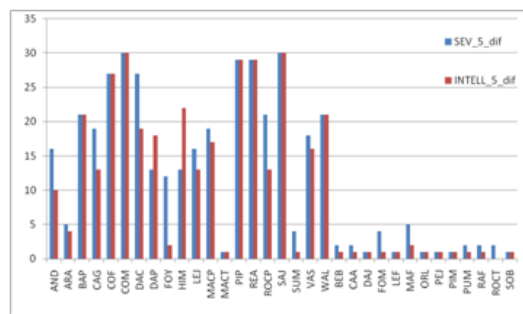
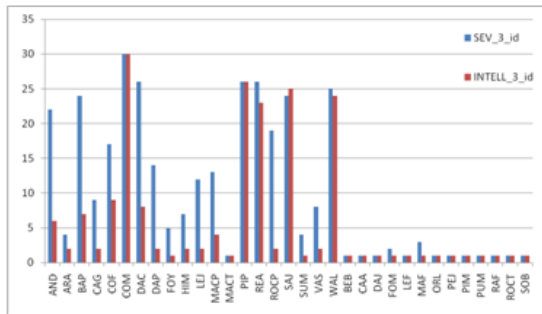
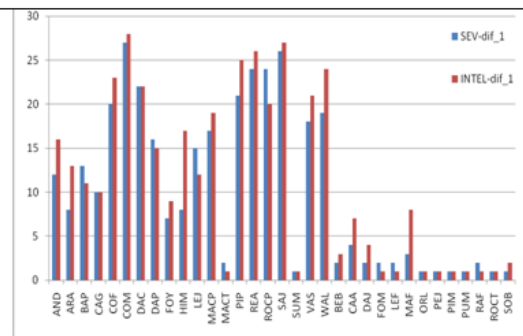
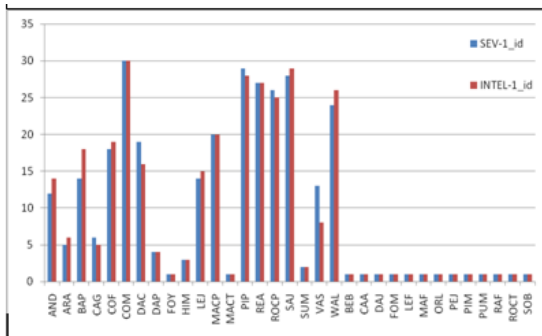
---

Les moyennes des scores de sévérité et d'altération de l'intelligibilité sont respectivement 9,88 (+/- 9,88) et 8,20 (+/-9,17) tout juge confondu. Les coefficients de corrélation intraclasse pour l'ensemble des locuteurs sont satisfaisant, meilleurs pour la sévérité que pour l'altération de l'intelligibilité, respectivement : 0.913, (95% CI: 0.860, 0.952) et 0.832, (95% CI: 0.707, 0.910). Si on ne regarde que dans la population des patients (le plus utile), ils sont moins bons et deviennent discutables pour l'intelligibilité avec des chiffres de 0.749, (95% CI: 0.494, 0.890) pour l'altération de l'intelligibilité et 0.834, (95% CI: 0.708, 0.921) pour la sévérité.

La distribution des scores d'altération de l'intelligibilité et de sévérité pour l'ensemble des stimuli est représentée sur les histogrammes 1 à 5 de la figure n°4.3.2 pour chaque juge.

L'observation de ces distributions met en évidence une différence entre les scores d'altération de l'intelligibilité et de sévérité pour les juges 2 et 3. L'étude de la distribution des différences entre les deux scores (score de sévérité – score d'intelligibilité = SEV-INTEL) permet une meilleure visualisation du phénomène figure°4.3.3.

La différence de distribution entre les scores se traduit pour les juges 2 et 3 par des scores différents pour plus de la moitié des stimuli et un écart de plus de 5 points pour au moins un quart des stimuli. Dans la majorité des cas, le score de sévérité est alors supérieur au score d'altération de l'intelligibilité. Pour les juges 1 et 4 il y a peu de différence (quand elle existe, elle reste toujours inférieure à 5 points) et il n'y a pas de tendance de supériorité pour l'un ou l'autre des critères de classement. Le juge 5 a coté majoritairement de manière similaire la sévérité et l'intelligibilité mais lorsque les stimuli ont été cotés différemment, c'est avec des différences pouvant dépasser 5 points et avec la même tendance que les juges 2 et 3, c'est-à-dire un score supérieur pour la sévérité.

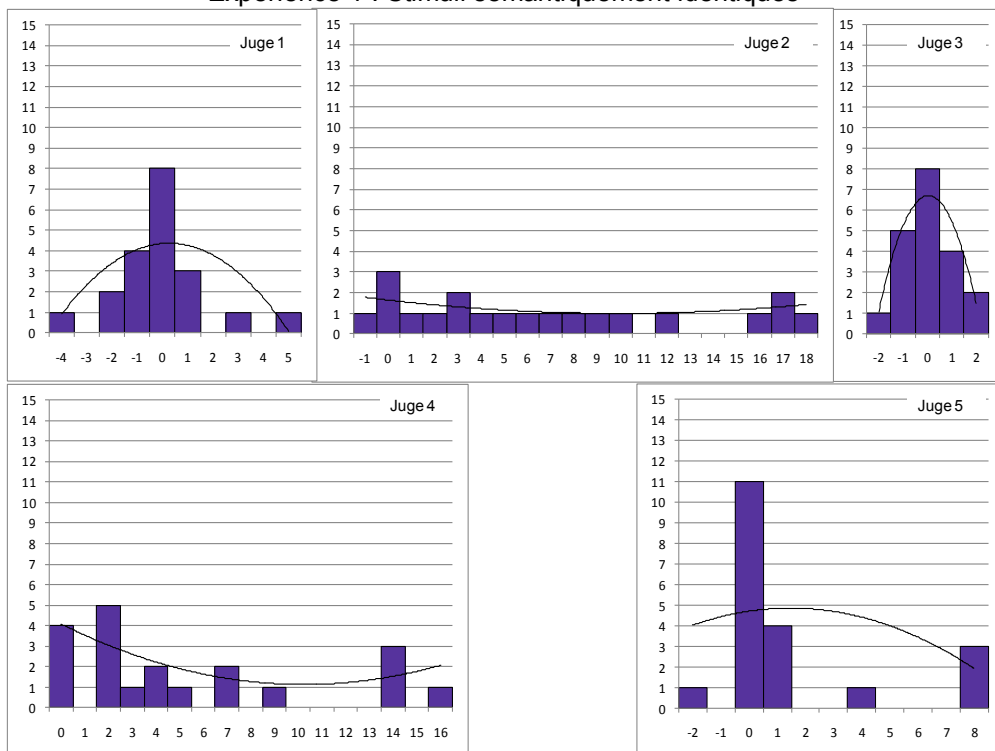


Expérience 1

Expérience 2

4.3.2 : Distribution des scores d'altération de l'intelligibilité et de sévérité pour chaque juge

### Expérience 1 : Stimuli sémantiquement identiques



### Expérience 2 : Stimuli sémantiquement différents

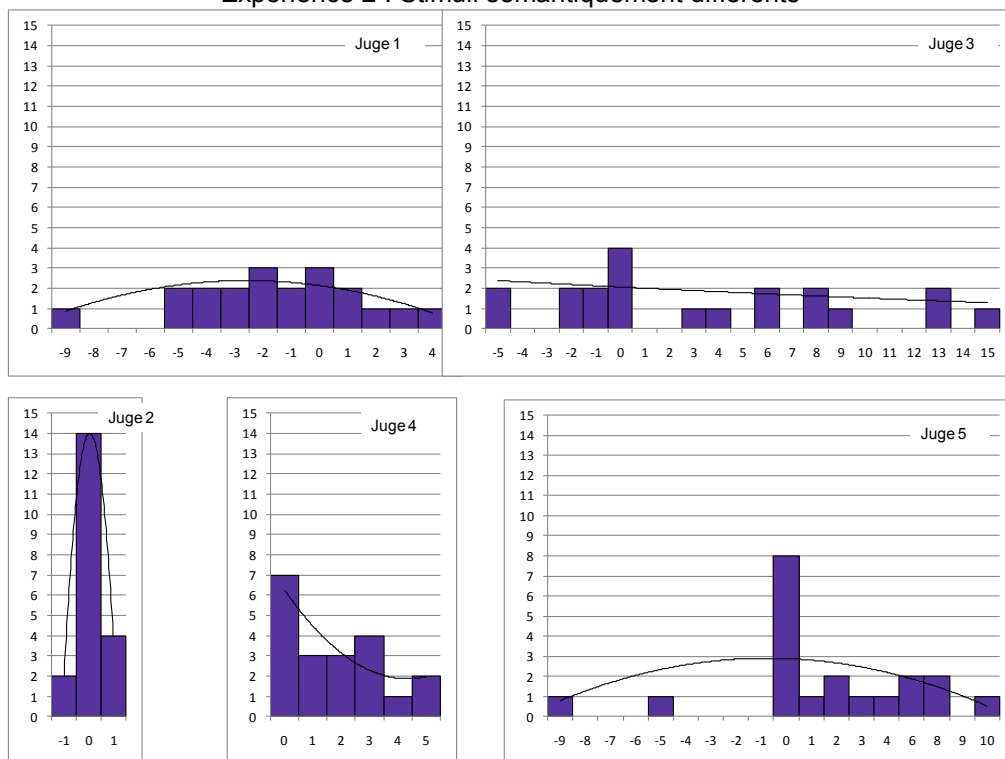


Figure n°4.3.3. : Histogramme de la différence entre sévérité et degré d'altération de l'intelligibilité pour chaque juge avec une courbe de tendance polynomiale d'ordre 2 pour les expériences 1 & 2..

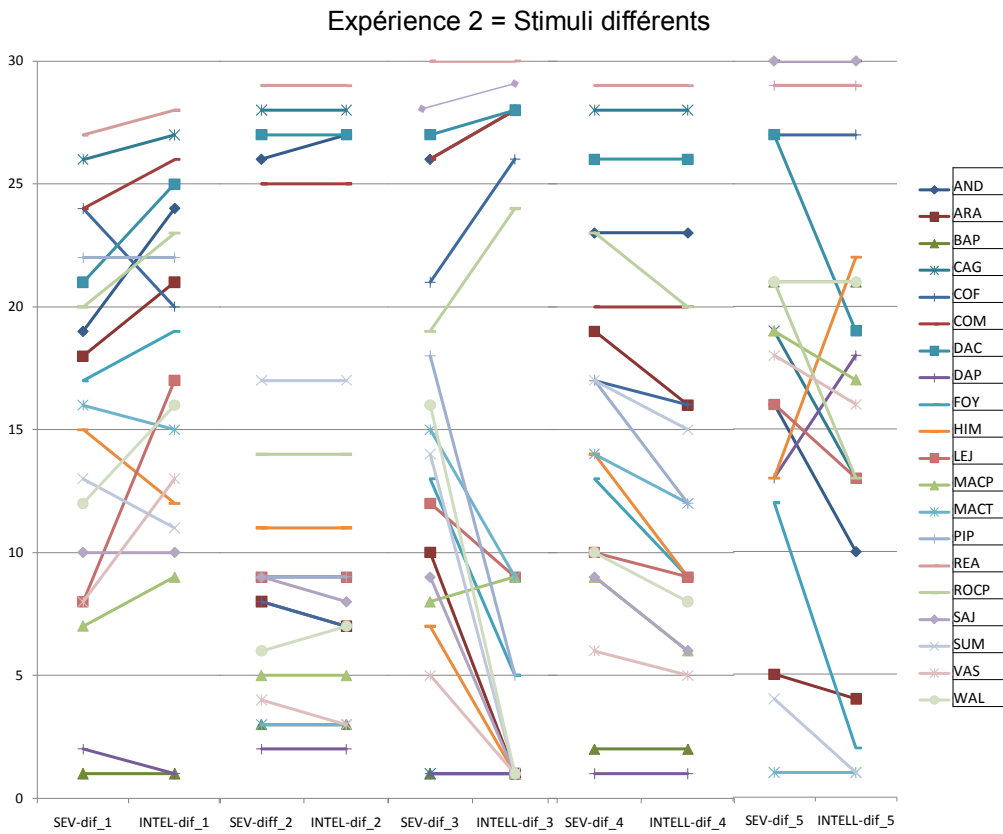
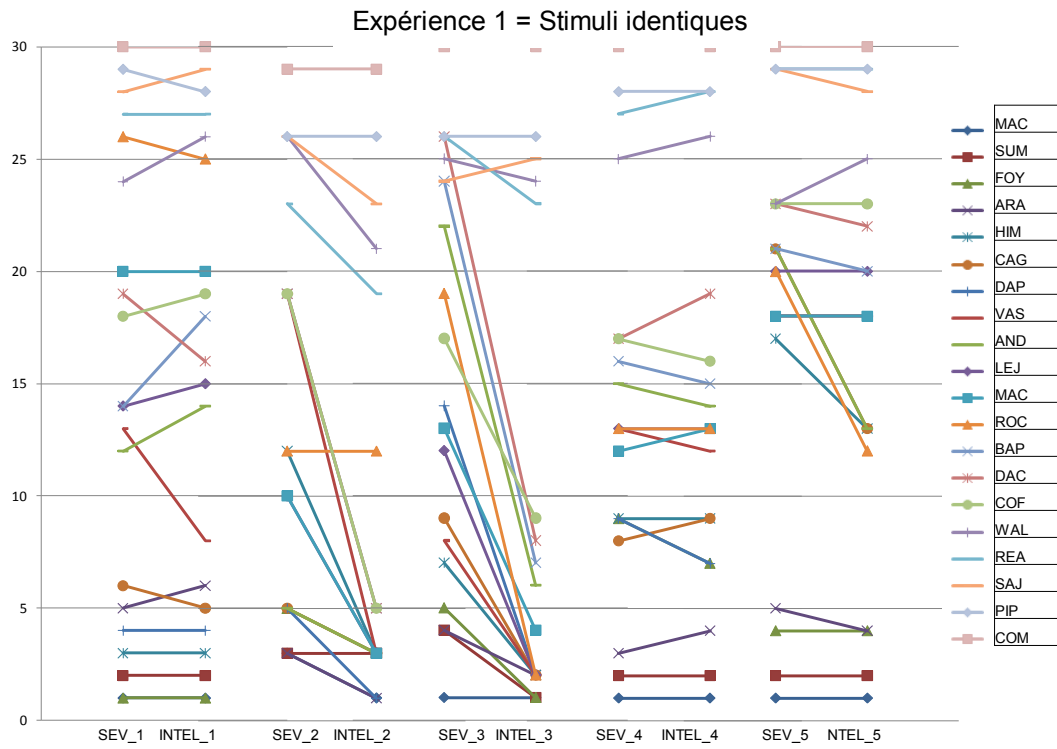


Figure n°4.3.4. : graphe de cellules pour les variables sévérité (SEV) et altération de l'intelligibilité (INTEL) par juge

Ainsi apparaît l'absence de comportement uniforme dans ce jury avec 2 tendances :

- L'absence de différence entre les 2 critères de classement illustré par les juges 1 et 4
- Une différence d'application des 2 critères de jugement avec un score souvent supérieur pour la sévérité illustré par les juges 2 et 3.
- Le juge 5 semble présenter un comportement intermédiaire.

Ces résultats sont illustrés par les graphiques de cellules présentés pour chaque juge sur la figure n°4.3.4

#### 4.3.1.3. LES CORRELATIONS ENTRE LES SCORES

Les corrélations entre les scores de sévérité et d'intelligibilité sont fortes pour chaque juge (>0,9).

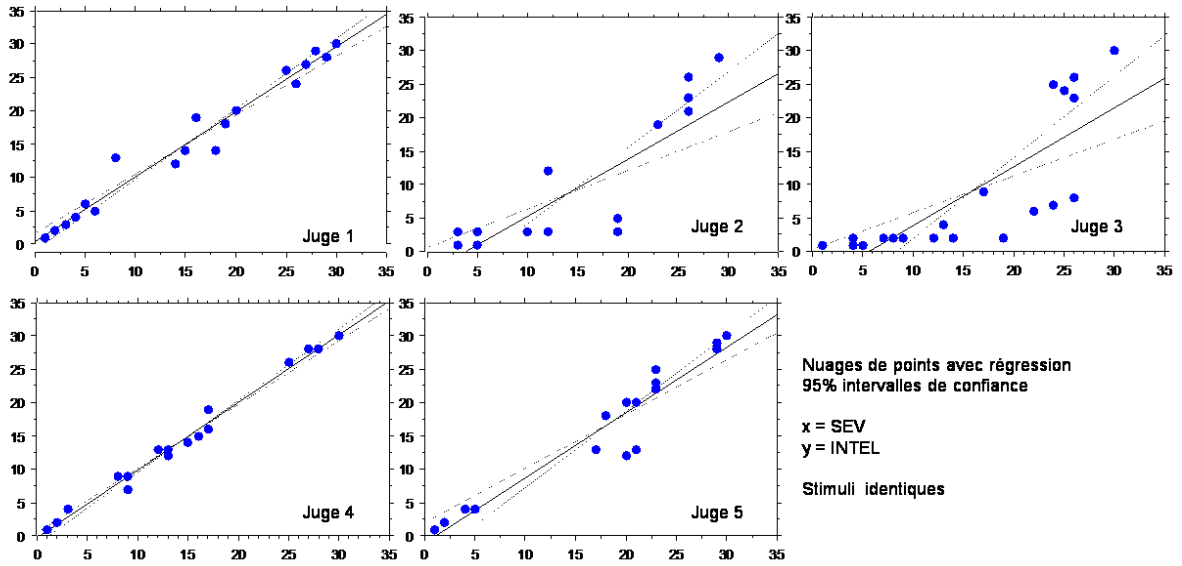
D'une manière générale un score élevé d'altération de l'intelligibilité était associé à un score élevé de sévérité (tableau n°4.1).

Expérience 1	r	p
juge 1	0,99	<0,001
juge 2	0,90	<0,001
juge 3	0,91	<0,001
juge 4	0,98	<0,001
juge 5	0,91	<0,001
Expérience 2	r	p
juge 1	0,96	<0,001
juge 2	0,99	<0,001
juge 3	0,90	<0,001
juge 4	0,98	<0,001
juge 5	0,94	<0,001

Tableau n° 4.1 : scores de sévérité et d'intelligibilité pour chaque juge.

Graphiquement (figure n°4.3.5), les scores d'intelligibilité et de sévérité étaient proches lorsque les scores étaient faibles ou élevés quelque soit le juge. Pour les juges 2 et 3 (et dans une moindre mesure pour le juge 5), lorsque les scores étaient intermédiaires, l'intelligibilité avait un score plus faible que la sévérité. La distribution des scores n'apparaît alors plus linéaire mais concentrés aux deux extrémités de l'échelle des scores pour l'intelligibilité.

### Expérience 1



### Expérience 2

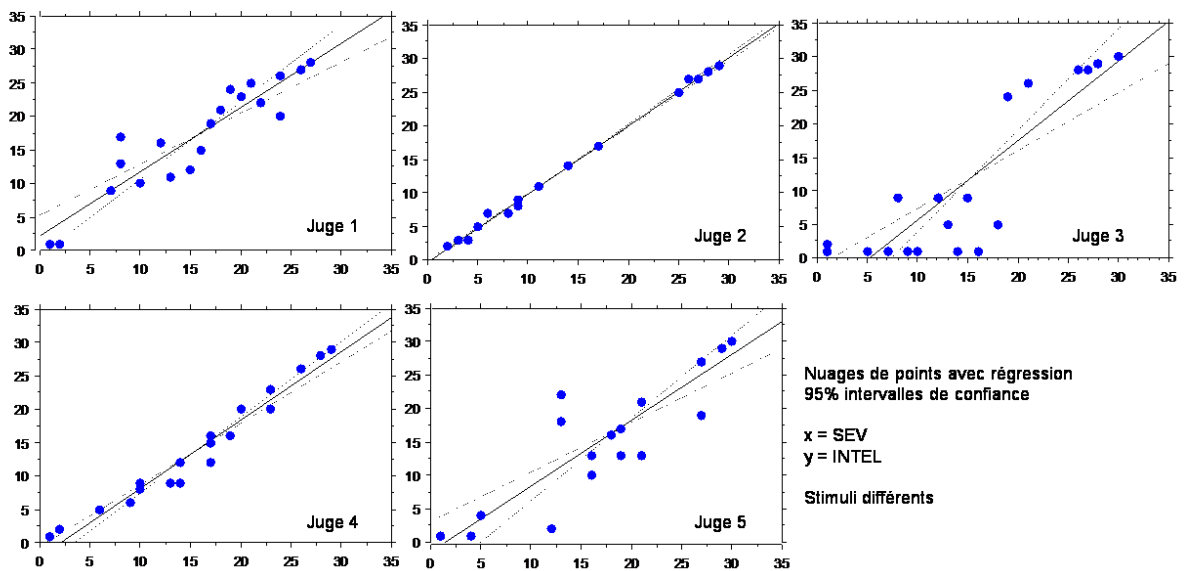


Figure n°4.3.5 : Nuages de points sur un graphe bivarié des scores de sévérité et d'altération de l'intelligibilité

Les corrélations entre les tâches d'analyse préalable et le jugement de sévérité ou d'intelligibilité pour chaque juge, chez les patients sont rapportées sur la figure 4.3.6. Le tableau des valeurs correspondant est dans l'annexe 7.2.2.

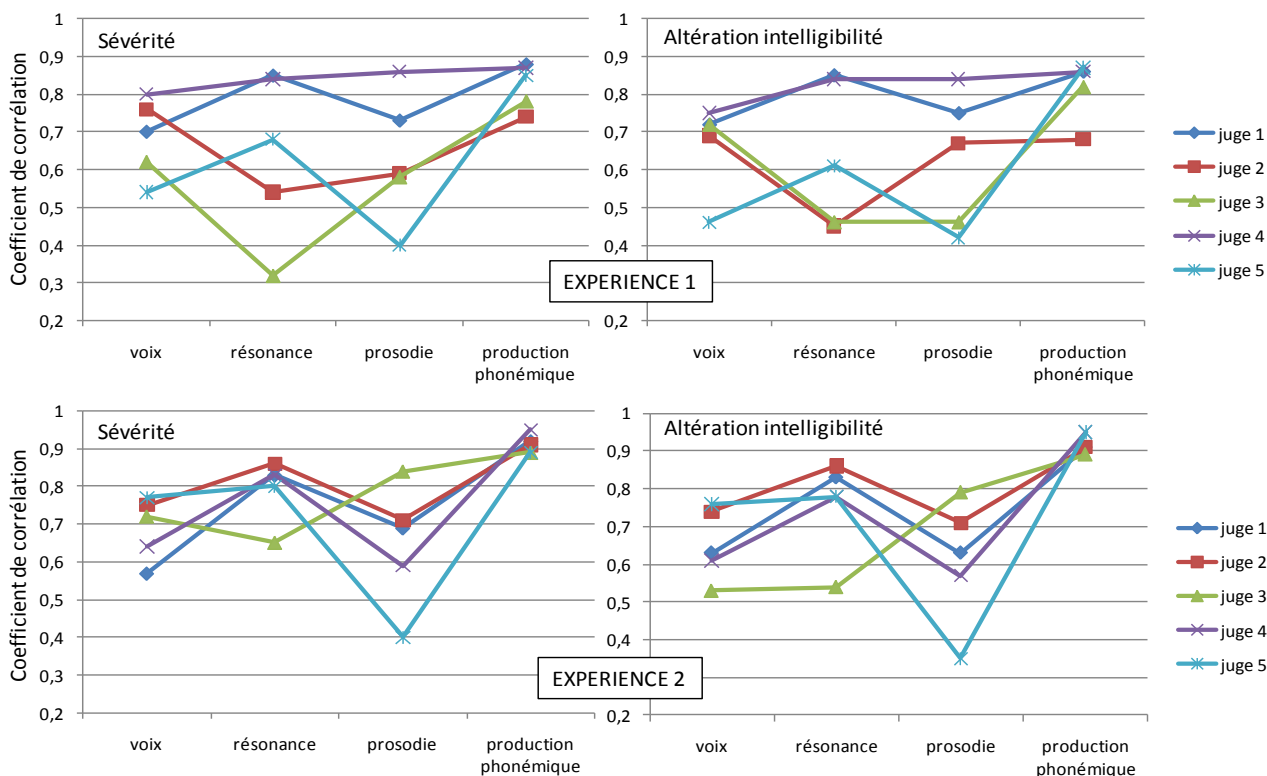


Figure n°4.3.6 : Corrélations entre la Sévérité et l'intelligibilité d'une part, et les tâches de jugement préalable, d'autre part, pour les expériences 1 et 2.

On observe une corrélation élevée entre chaque tâche et la sévérité ou l'altération de l'intelligibilité pour les juges 1 et 4. Pour les juges 2 et 3, les corrélations étaient plus élevées entre sévérité et qualité de la voix ou production phonémique qu'avec la qualité de la résonance (ou la prosodie pour le juge 2). Il n'apparaît pas de différence de corrélation nette entre le jugement d'intelligibilité ou de sévérité par rapport à la tâche préalable.

Si on s'intéresse à la différence entre le score d'intelligibilité et le score de sévérité (sévérité – intelligibilité, figure n° 4.3.7), les corrélations entre cette différence et les différents scores d'analyse étaient faibles pour l'ensemble des juges et on n'observait pas de signe systématiquement positif ou négatif, y compris pour les juges qui avaient tendance à coter différemment la sévérité de l'intelligibilité pour les scores intermédiaires. Le tableau des valeurs correspondant est dans l'annexe 7.2.2.



Cette analyse ne permet pas de répondre à la différence de comportement des juges entre intelligibilité et sévérité. En revanche, elle confirme l'existence de différence dans l'analyse faite par les juges. Ainsi, les altérations phonémiques sont déterminantes même si un des juges a des corrélations plus importantes pour les caractéristiques vocales (que soit pour la sévérité où l'intelligibilité).

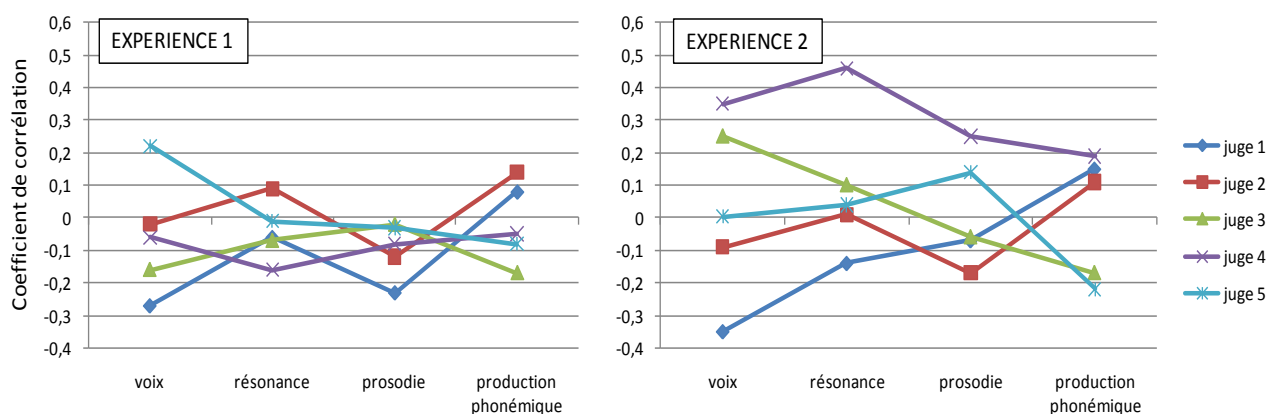


Figure 4.3.7 : Corrélations entre les scores de l'analyse préalable et la différence de score entre sévérité et altération de l'intelligibilité (sévérité – intelligibilité) pour les expériences 1 et 2.

Deux juges semblent utiliser dans leur jugement toutes les caractéristiques alors que 3 autres semblent fonder leur jugement sur 2 caractéristiques principales. Cependant, ils ne s'agit pas des mêmes ou pas dans le même ordre.

#### 4.3.2. EXPERIENCE 2

L'objectif de l'expérience 2 était de vérifier la validité des résultats de l'expérience 1 avec des stimuli dont le contenu sémantique était inconnu et donc différent pour chaque locuteur.

##### 4.3.2.1. LA DISTRIBUTION DES SCORES DE L'ANALYSE PERCEPTIVE

La distribution des scores de l'analyse préalable sur la qualité des stimuli est représentée pour chaque juge figure n°4.3.1 de manière à pouvoir être comparé visuellement avec la distribution de l'expérience 1. Il demeure de grandes variations dans la cotation des caractéristiques des stimuli avec des changements entre des prépondérances de cotation. De manière décroissante, les juges 3, 4, 5 ont moins de jugements anormaux sur la voix par rapport aux autres paramètres (changement pour le juge 2 avec une augmentation relative et le juge 4 avec une diminution relative), les juges 1, 2, 3 moins de jugement anormaux sur la résonance (changement pour les juges 1 et 2) et le juges 5 moins

de jugement anormaux pour la prosodie (au lieu du juge 1). Globalement on assiste à une diminution des scores attribués sauf pour le juge 5.

#### 4.3.2.2. LA DISTRIBUTION DES SCORES DE SEVERITE ET D'ALTERATION DE L'INTELLIGIBILITE

---

Les moyennes des scores de sévérité et d'altération de l'intelligibilité sont respectivement 10,18 (+/- 9,26) et 9,26 (+/-9,36). La moyenne des scores reste moins élevée pour l'altération de l'intelligibilité par rapport à la sévérité mais toujours avec des écarts types élevés.

La distribution des scores d'altération de l'intelligibilité et de sévérité pour l'ensemble des stimuli est représentée sur les histogrammes 1 à 5 de la figure n°4.3.2 pour chaque juge, la distribution des différences entre les deux scores sur la figure°4.3.3. L'observation de ces distributions met en évidence une différence entre les scores d'altération de l'intelligibilité et de sévérité pour les juges 3 et 5 alors que le juge 2 paraît avoir changé de comportement. A part pour ce juge, qui ne fait plus aucune différence, comme si il ne faisait pas la tâche, les profils des histogrammes se sont « harmonisés » avec une augmentation de l'étendue de la différence entre les scores et une asymétrie de distribution par rapport au zéro traduisant la tendance à coter de manière plus forte la sévérité par rapport à l'altération de l'intelligibilité.

Cependant l'analyse des graphiques de cellules présentés pour chaque juge sur la figure n°4.3.4 révèle la persistance d'une tendance plus marquée à coté la sévérité de manière différente en accordant des scores plus proche des extrémités de l'échelle. Le juge 2 est remarquable par son changement de comportement. A noter qu'à la fin de l'expérience, ce juge a exprimé des hypothèses sur la différence existante entre les 2 expériences traduisant la recherche pendant l'expérience de la nature de cette différence. En effet, aucun juge n'a été prévenu de la raison de cette nouvelle manipulation et seul le juge 2 s'est interrogé sur ce point pendant l'expérience.

#### 4.3.2.3. LES CORRELATIONS ENTRE LES SCORES

---

Les corrélations entre les scores de sévérité et d'intelligibilité restent fortes pour chaque juge (> à 0,9). (tableau n°4.1). Graphiquement (figure n°4.3.5), la tendance se confirme, les scores d'intelligibilité et de sévérité sont proches lorsque les scores sont faibles ou élevés quelque soit le juge. Dans la partie

moyenne de l'échelle la variabilité devient importante mais la tendance à concentrer aux deux extrémités de l'échelle des scores pour l'intelligibilité reste net que pour le juge 3.

En ce qui concerne les corrélations entre les taches d'analyse préalable et le jugement de sévérité ou d'intelligibilité (figure n° 4.3.6), un renforcement de la corrélation avec la qualité de la production phonémique est observé pour les deux taches (corrélations  $> 0,9$ ). Il n'apparaît toujours pas de différence de corrélation entre le jugement d'intelligibilité ou de sévérité par rapport à la tache préalable. Les juges qui font le moins de différences tendent toujours à établir des corrélations similaires avec tous les paramètres. Les corrélations avec la différence mesurée entre le score d'intelligibilité et le score de sévérité (sévérité – intelligibilité, figure n° 4.3.7) et les différents scores d'analyse restent faibles pour l'ensemble des juges. On note simplement une corrélation plus forte mais qui reste légère pour le juge 4 avec la qualité de résonance et à moindre degré de la voix.

Cette analyse ne permet toujours pas de répondre à la différence de comportement des juges entre intelligibilité et sévérité.

---

#### 4.3.3. EXPERIENCE 3

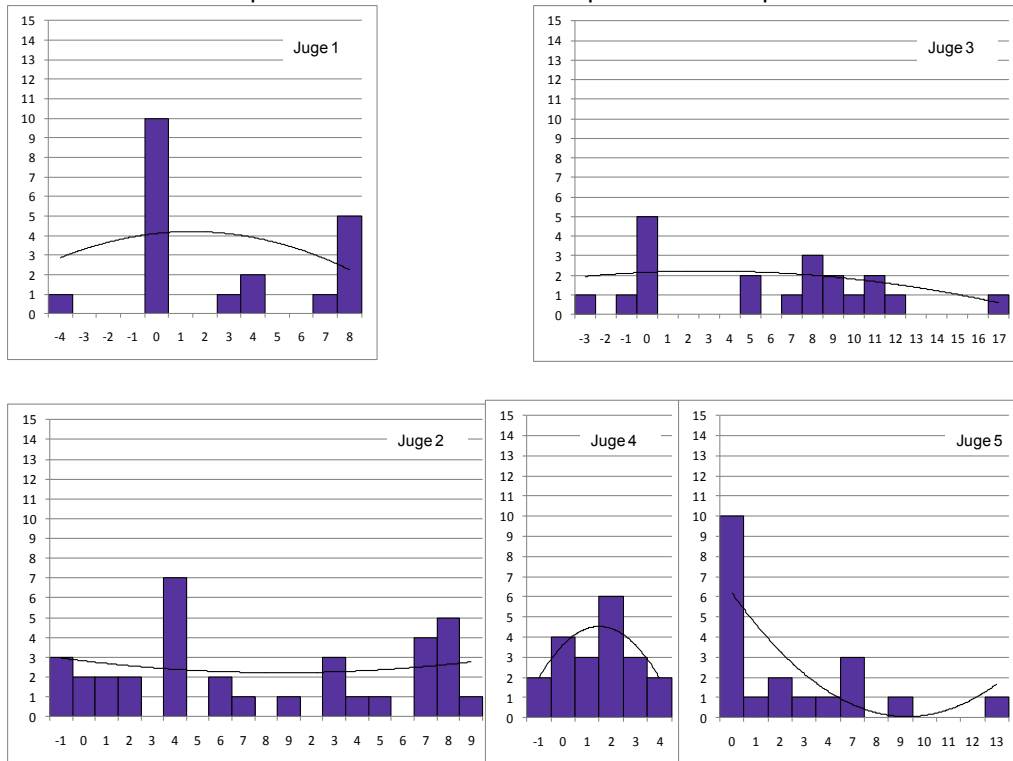
---

L'objectif de l'expérience est de vérifier si la reproductibilité des expériences précédentes est suffisante pour la crédibilité des conclusions générales de l'étude en raison du faible nombre d'experts. Malgré le risque d'évolution liée à la répétition de la tache, nous avons reproduit les mêmes conditions en enchainant les deux expériences dans l'ordre initial. Seule la description de la distribution entre les 2 taches de jugements est rapportée. L'analyse de l'analyse préalable n'apportant pas d'éléments supplémentaires.

La distribution des différences entre les deux scores d'altération de l'intelligibilité et de sévérité pour l'ensemble des stimuli est représentée sur les histogrammes de la figure n°4.3.8. Il confirme 9 mois après la première expérience la tendance des juges 2,3 et 5 à réaliser des écarts de scores plus grands entre intelligibilité et sévérité pour les stimuli identiques avec une tendance à sous coter l'intelligibilité par rapport à la sévérité. Cette tendance est atténuée avec un recentrage autour du 0 pour les stimuli différents. A noter que le juge 2 n'a pas reproduit le changement de comportement majeur de l'expérience 2. La représentation sous forme de graphe de cellules (figure n°4.3.9) précise que si la tendance est à la baisse de score pour la tache avec stimuli identiques pour tous les juges, il

y a bien une tendance à attribuer un score rapproché d'une des extrémités de l'échelle quand les stimuli sont différents. Les graphes bi variés (figure n°4.3.10) confirme la plus grande variabilité sur la partie médiane de l'échelle quand elle n'est pas « déshabillée ».

### Expérience 3 : Stimuli sémantiquement identiques



### Expérience 3 : Stimuli sémantiquement différents

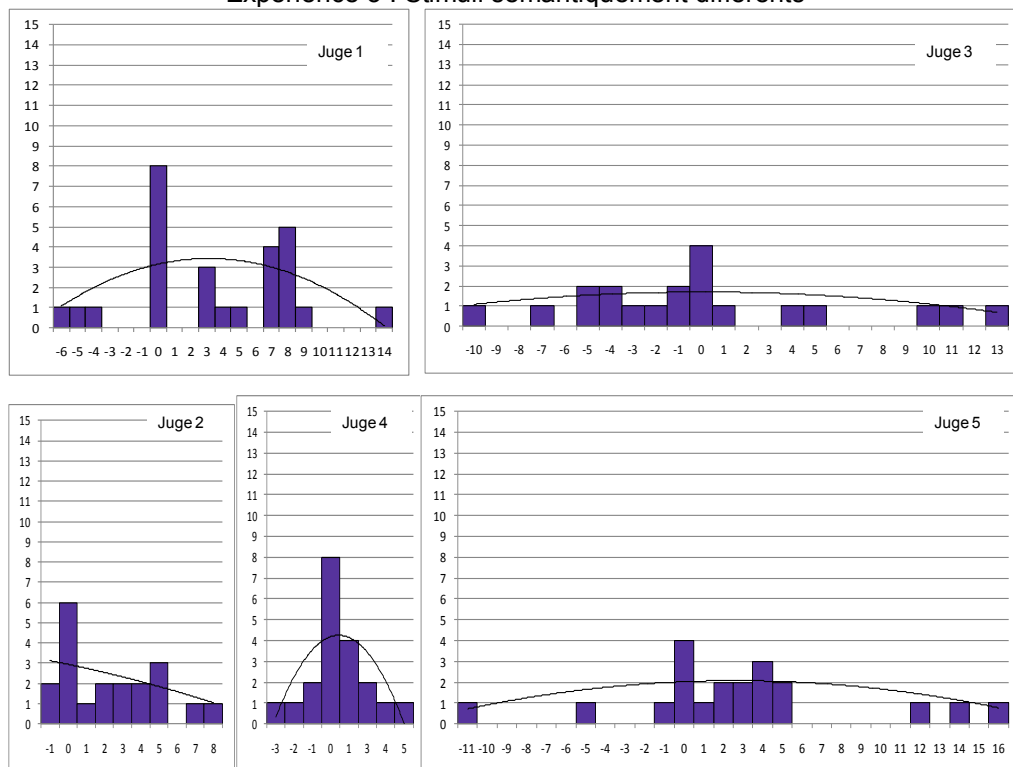


Figure 4.3.8 : Histogramme de la différence entre sévérité et degré d'altération de l'intelligibilité pour chaque juge avec une courbe de tendance polynomiale d'ordre 2 <sup>pir</sup> l'expérience 3.

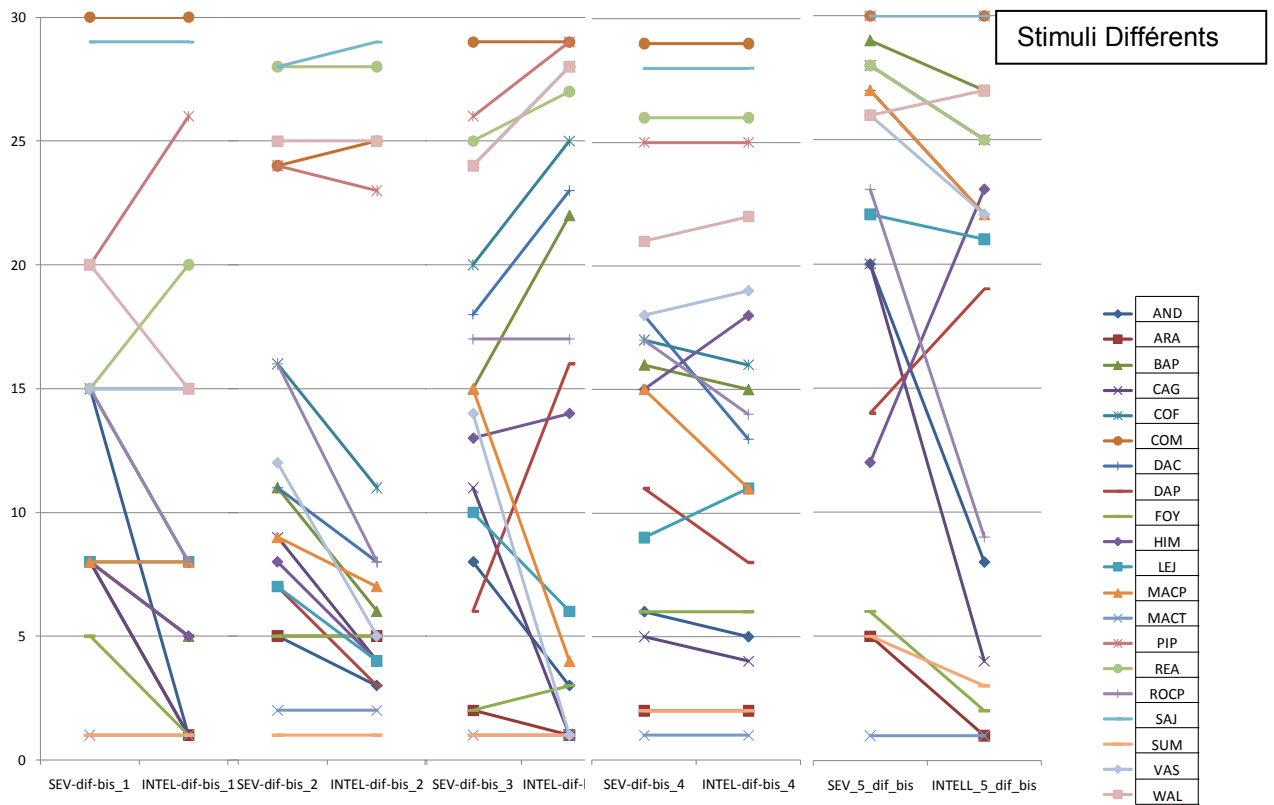
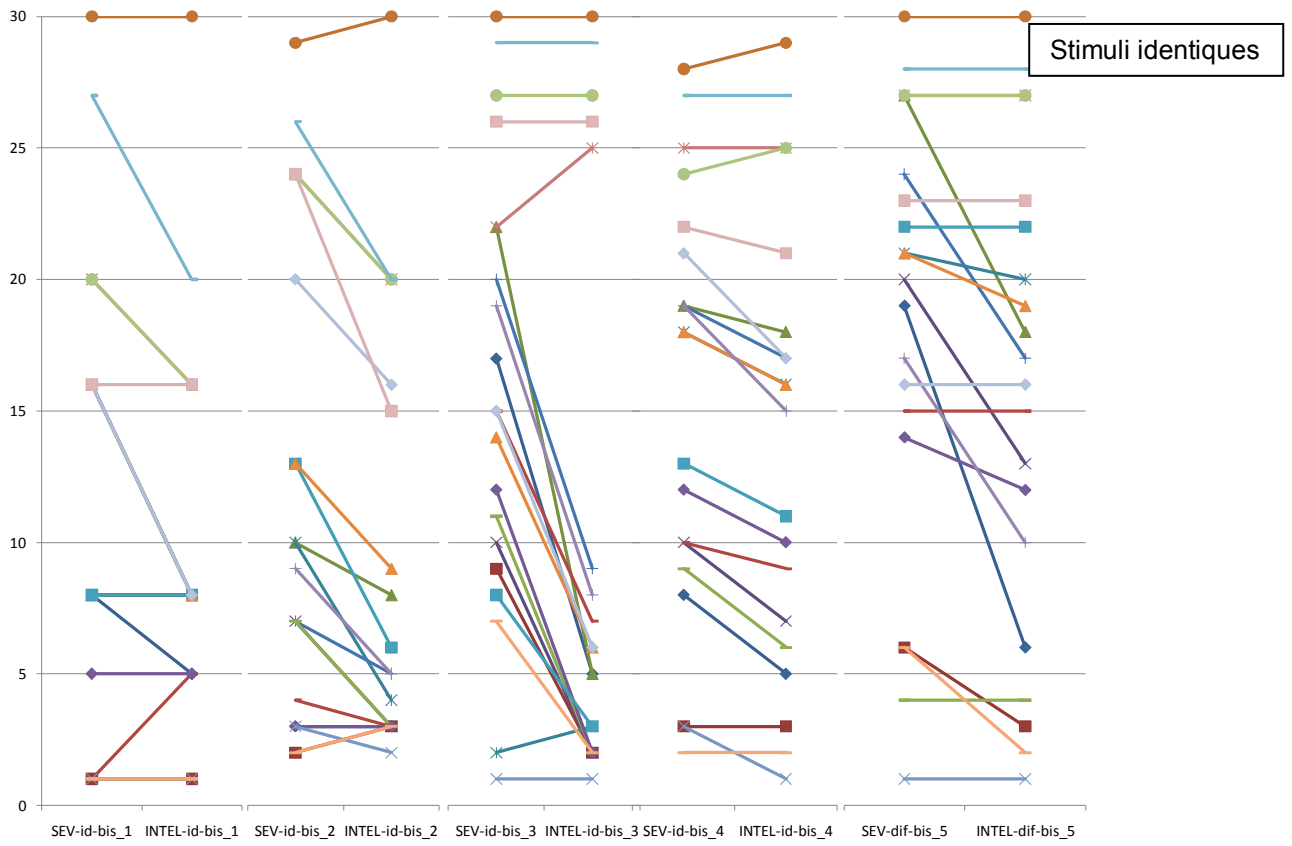


Figure n°4.3.9 : graphe de cellules pour les variables sévérité (SEV) et altération de l'intelligibilité (INTEL) par juge pour l'expérience 3 (en haut, stimuli identiques, en bas, stimuli différents)

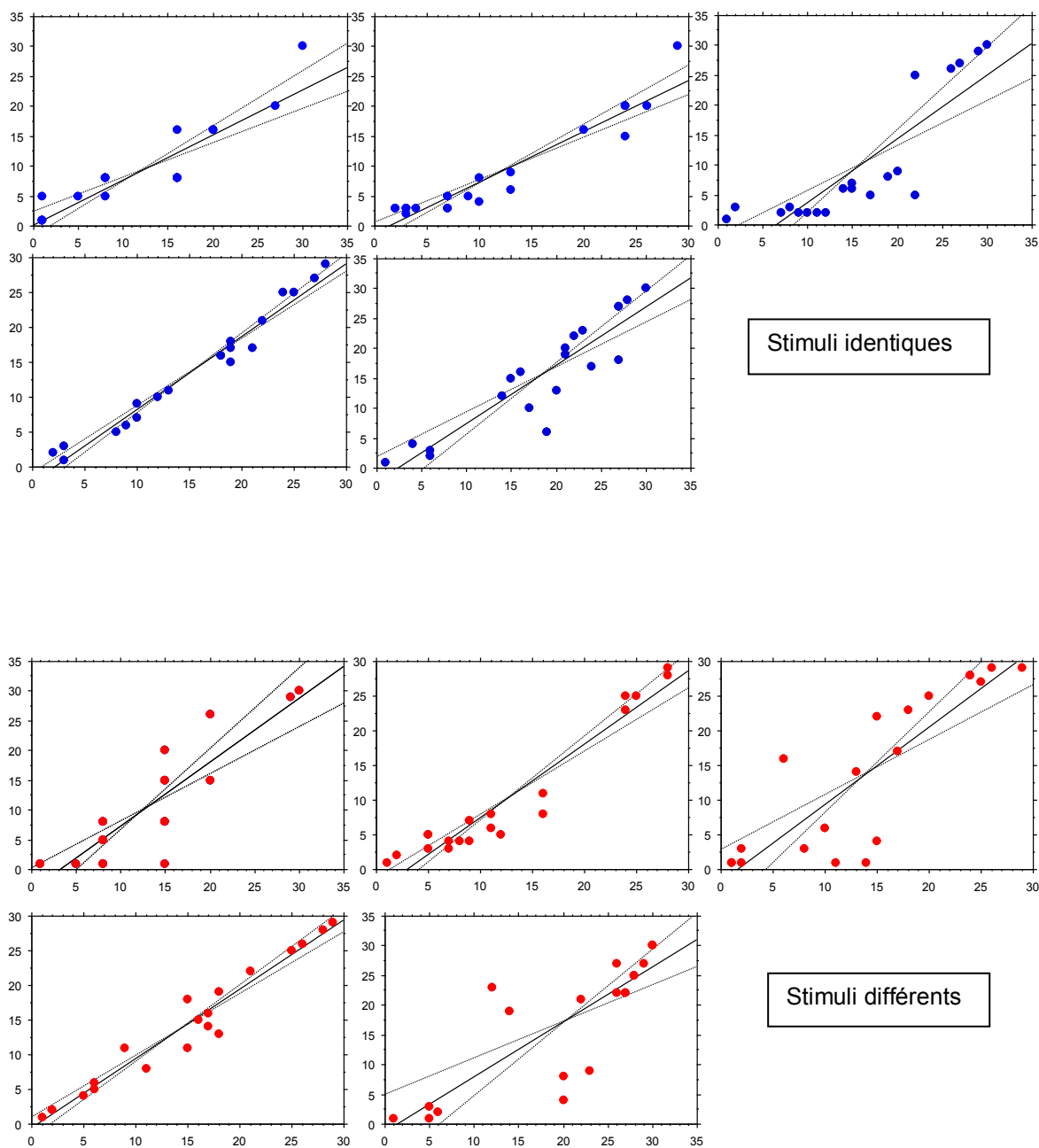


Figure 4.3.10 : Nuages de points sur un graphe bivarié des scores de sévérité (x) et d'altération de l'intelligibilité (y) avec régression à intervalle de confiance 95% (en haut, stimuli identiques, en bas, stimuli différents)

## 4.4. DISCUSSION

---

### 4.4.1. LES FAITS ETABLIS PAR CETTE EXPERIENCE

---

L'expérience menée avec ces 5 experts met en évidence des comportements différents dans le jugement de l'intelligibilité par rapport à la sévérité du trouble pathologique de la production de la parole. Les conditions expérimentales réduisent le champ des conclusions à :

- Des troubles de la production d'origine neurologiques ou structurelles
- A un groupe de cinq experts d'éducation différente mais travaillant ensemble au moins 1 jour par semaine depuis 5 ans.

Les faits établis sont une différence nette de distribution des réponses pour 2 juges, une distribution des réponses identiques pour 2 des juges et un comportement que l'on pourrait qualifier d'intermédiaire pour le dernier juge dans la première expérience réalisé avec des stimuli de contenu sémantique identique. Et ceci, bien que les corrélations entre les deux mesures soient élevées. Cette différence dans la manière d'attribuer des scores aux 2 notions peut être reproduite avec des changements de comportement des juges au cours des différentes expériences. Le phénomène n'est pas lié à l'utilisation des stimuli identiques, même si les stimuli identiques favorisent une tendance à abaisser les scores d'altération de l'intelligibilité. Il est par ailleurs probable qu'un des juges n'est pas réalisé la tâche demandée sur la 2<sup>ème</sup> expérience la distribution du score étant totalement parfaitement centré sur le 0, en contradiction avec les résultats des 2 autres temps de l'expérimentation.

Ces différences de distributions sont associées à des différences dans la contribution des différentes analyses des caractéristiques des stimuli sans qu'un lien de causalité puisse être établi. Les scores obtenus lors de cette tâche préalable faisant également l'objet de différences individuelles.

Ces résultats conduisent à discuter l'hétérogénéité des juges et la différence observée entre le jugement de sévérité et le jugement d'intelligibilité pour certains juges.

Dans la pratique de la recherche clinique, l'étude des jurys a pour principal finalité de démontrer une cohérence suffisante de ses membres pour valider les hypothèses auxquelles il est soumis. C'était le

cas de ce groupe d'expert qui a participé un an avant cette expérience à la validation d'un test d'intelligibilité avec une bonne cohérence et une bonne reproductibilité intra et inter-juge (Dittner et al 2010). Le besoin ou la recherche d'une homogénéité pour valider des outils diagnostiques ou thérapeutiques explique l'absence d'intérêt pour les causes de variabilité et peut ainsi masquer l'existence de processus perceptif différents mettant en jeu la conception des tests. Schiavetti (1992, p27) avait déjà noté dans un chapitre consacré aux méthodes de mesure de l'intelligibilité : "In others words, external validity may have been sacrificed in many studies to improve internal validity »."

---

#### 4.4.2. QUELLES PEUVENT ETRE LES RAISONS DE CETTE HETEROGENEITE ?

---

La manière dont les instructions ont été éditées pour la réalisation de cette tâche peut avoir influencé les processus cognitifs associés aux 2 concepts de sévérité et d'intelligibilité.

Ainsi, le jugement de l'intelligibilité renvoie à la perception de la parole dont les théories de la reconnaissance des mots (cf section 1.3.2.1) mais elle peut aussi du fait de la notion de « compréhensibilité », renvoyer aux théories de la compréhension discursive (cf section 1.3.2.2).

Ainsi, plusieurs niveaux d'analyse peuvent être envisagés pour l'intelligibilité, comme par exemple les trois niveaux de représentation du discours proposé par Van Dijk et Kintsch (1983) comme le propose Hustad (2008) dans son travail sur la compréhensibilité. Par contre, le concept de sévérité ne bénéficie pas d'une réflexion théorique spécifique et évoque plutôt la perception d'un degré de déviance par rapport à l'idée que se fait chaque individu de la parole « normale ». Cette notion dépasse le « simple » accès au sens et inclut la perte du caractère naturel et le degré de difficulté à produire la voix que projette l'interlocuteur sur le locuteur. Le fait d'avoir utilisé dans la première expérience un corpus constitué par le même texte produit par différents locuteurs dont 1/3 de sujets normaux a pu orienter la tâche de sévérité en donnant une forte référence de normalité aux juges.

Il est également possible que cela est induit, un biais sur la tâche d'intelligibilité car le texte étant connu, résoudre la tâche peut avoir généré plusieurs stratégies explicitées sous la forme des interrogations suivantes : Est-ce que je comprendrais le message si je ne connaissais pas le texte ? ou par rapport à la normale, qu'elle est le degré de déviance sur la transmission de l'information ? Ainsi la prédictibilité du sens étant maximum, cela peut mettre en jeu une capacité d'abstraction - est ce cette capacité d'abstraction qui a été mesurée ou une différence de comportement ? A noter



également, que La connaissance du contenu des stimuli équivaut à une familiarisation maximale avec le corpus. L'effet de cette familiarisation a été bien étudié (Beukelman & Yorkston 1980). Elle induit une diminution significative des scores d'intelligibilité. Ceci est retrouvé dans cette étude avec une moyenne des scores tout juge confondu dans l'expérience 2, plus importante pour l'altération de l'intelligibilité (9,26 (+/-9,36) dans l'expérience1 vs 8,20 +/-9,17 dans l'expérience 2) sans que cela soit significatif sur ce petit nombre de stimuli.

La suppression de cette tendance avec l'utilisation de stimuli de contenu différents « démasque » le comportement de « polarisation » des scores d'intelligibilité vers les deux extrémités de l'échelle mais les 2 types de comportements persistent à un moindre degré, même si cela n'est pas pour les mêmes juges. L'hypothèse d'un processus cognitif relevant d'un phénomène d'abstraction ne paraît donc être l'explication (exclusive) à cette différence de comportement. En revanche, un processus cognitif faisant appel à un phénomène d'ordre « top-down » impliqué dans les processus d'identification des mots avec est une hypothèse plausible.

---

#### 4.4.3. EN CE QUI CONCERNE LA DIFFERENCE DE SCORE ENTRE LES DEUX NOTIONS, QUELLES PEUVENT ETRE LES EXPLICATIONS ?

---

Analysé du point de vue de la psychophysique, la différence peut être dans « l'objet » de la mesure elle-même. Steven et Galander en 1957 décrivent deux sortes de continuum perceptif : prothétique versus methathétique et écrivent que *“a prothetic continuum is additive, whereas a metathetic is substitutive but the criteria that define those two classes of continua reside wholly in how they behave in psychophysical experiments”*. Ainsi une explication à cette différence pourrait être que ces 2 notions ne correspondent pas au même type de continuum. Mais, il y a peu d'étude apportant des arguments à cette hypothèse. Comme dans notre étude, Samar et Metz (1988) dans un travail sur la validation d'échelles à intervalle pour mesurer l'intelligibilité, signalent un problème dans le milieu de l'étendue de l'intelligibilité. Ils suggèrent que *“apparently high correlations seen between word identification tests and rating scale measures of intelligibility may be spurious indicators of the validity of rating scales because linear correlational , often do not take into account the possibility of sigmoidal relationship along the entire intelligibility range”*. Placé dans la perspective des théories de la perception de la parole, cette fonction sigmoïdale pourrait être interprétée comme la conséquence des processus cognitifs de d'identification des mots.

En fait, cet effet est maintenant bien connu et n'intéresse pas seulement l'intelligibilité mais tous les paramètres perceptifs utilisés dans les TPPP y compris vocaux (Kreiman et Gerrat 2009, Bunton et al 2007). En revanche, ils n'ont pas décrit de regroupement vers les extrémités de l'échelle des scores attribués. Leur majoration pour la tâche d'intelligibilité chez certains membres du jury peut être interprété comme le phénomène de transduction entre un phénomène acoustique et un phénomène linguistique. Dans une perspective connexioniste de la construction du sens, Coleman postulait en 2002 l'existence d' « espaces psychophysiques » dans lesquels l'association entre l'espaces auditif et l'espace sémantique serait rendu possible «in order to encode the semantically contrastive function of differences between the sounds of different words, the auditory space must be associated with semantic space (...). The phonetic space may be quasi continuous, but in places it is effectively carved up by virtue of these associations to discrete semantic categories". Ainsi, les tâches de jugement de sévérité et d'altération de l'intelligibilité dans les TPPP peuvent non seulement répondre à des processus cognititifs différents mais aussi trouver leur origine dans des unités différentes du signal de parole (en tant qu'indices acoustiques de base).

Schiavetti (1992), investi dans le domaine de la surdité et particulièrement dans la parole des sourds, a travaillé sur les procédures de mesure de l'intelligibilité de la parole et concluait qu'il y avait de nombreux arguments contre l'utilisation d'échelles pour cette mesure. Parmi ceux-ci, il signale l'absence de corrélation et la grande variabilité dans la partie moyenne de l'étendue des scores d'intelligibilité. De plus, il insiste sur les avantages d'une procédure d'identification de mot pour l'intelligibilité mais précise que si une échelle de mesure était utilisée, une échelle ordinale serait plus adaptée qu'une échelle à intervalle.

Mais comme la tendance dans les TPPP est à utiliser l'intelligibilité comme indice, il n'y a pas de travail antérieur sur la mesure directe de la sévérité. Ce travail apporte des arguments pour une relation plus linéaire sur toute l'étendue de l'altération du signal sonore en opposition aux données sur l'intelligibilité mais cela n'est pas suffisant pour savoir si une procédure d'échelle à intervalle est valide.

#### 4.5. CONCLUSION

---

Les notions de sévérité, d'intelligibilité et de compréhensibilité ont besoin d'être clairement définies dans le domaine des TPPP. Avoir des limites bien définies et consensuelles de ces définitions permettrait de déterminer des critères de validité externe et choisir les meilleures mesures pour chaque concept. En pratique ce travail apporte en complément de l'étude précédente (chapitre 3) des arguments pour considérer que l'impression générale d'intelligibilité peut correspondre à une estimation d'amplitude de la sévérité si elle est définie comme le degré d'altération de la production ou comme le degré de déviance par rapport à la normalité

Il est donc fondamental de standardiser la consigne pour réduire au maximum les effets inhérents à la tâche elle-même. Ceci paraît primordial avant de déterminer si des échelles à intervalle ou simplement ordinal pourront être utilisées dans la mesure où une meilleure connaissance de l'objet de la mesure est déterminant pour répondre à cette question. Le fait que les coefficients de corrélation intraclass pour l'ensemble des locuteurs sont meilleurs pour la sévérité que pour l'altération de l'intelligibilité est un argument supplémentaire.

Il est difficile de promouvoir l'utilisation restreinte du mot intelligibilité au sens des phonéticiens à son rapport à la perception de la parole. Non pas car cela n'est pas « justifié » mais simplement car il est difficile d'aller contre l'usage commun des mots. Or il est important de ne pas entretenir cette source d'ambiguïté. En revanche, associé systématiquement à la notion de test, la définition centrée sur la mesure du décodage dans une dimension acoustico-phonétique est plus largement partagée dans un contexte d'évaluation. Cela va dans le sens de la conclusion du seul ouvrage consacré à l'intelligibilité de la parole dans les TPPP avec une mesure par proportion des unités correctement perçues.

Reste à déterminer les relations entre les 3 notions pour élaborer un indice de sévérité qui pourra guider la prise en charge thérapeutique. Mais nous reprendrons ce point de discussion dans la discussion générale.

*Que notre Univers serait mutilé si nous le réduisions au Très grand et au Très petit, c'est-à-dire aux deux seuls abîmes de Pascal [...] un troisième abîme existe : celui de la complexité [...]. C'est sur ces trois infinis qu'est bâti spatialement le Monde. L'Infime et l'Immense, sans doute, mais aussi l'immensément compliqué [...] Quel est l'effet spécifique du très grand complexe ? C'est ce que nous appelons la Vie.*

*P. Teilhard de Chardin, Le Groupe zoologique humain. Structures et directions évolutives, 1956, p. 21*

## **5. ROBUSTESSE ET CONTRIBUTION A L'INTELLIGIBILITE**

---

5. Robustesse et contribution a l'intelligibilité	199
5.1. Introduction.....	203
5.1.1. La notion de robustesse	203
5.1.2. Hypothèses de recherche	205
5.1.2.1. Les limites du cadre expérimental usuel	205
5.1.2.2. Les hypothèses proposées	206
5.1.2.3. L'objectif de ce travail	210
5.1.2.4. Les résultats attendus	210
5.2. choix méthodologiques.....	212
5.2.1. Les locuteurs	212
5.2.2. Le Corpus	212
5.2.2.1 Choix du corpus	212
5.2.2.2 Modalités d'enregistrement.	214
5.2.2.3 Préparation des stimuli	214
5.2.2.4 Présentation des résultats	215
5.2.3. La Procédure de la tâche de perception	218
5.2.3.1. Description de la tâche	218
5.2.3.2. Les auditeurs	219
5.2.3.3. La procédure	219
5.2.3.3.1 Préparation de la présentation des stimuli sur le logiciel PRECEVAL	219
5.2.3.3.2 Passation	220
5.2.3. La méthode de la reconnaissance automatique de la parole.	221
5.2.3.1 Le processus de paramétrisation	222
5.2.3.2 La modélisation	222
5.2.3.3. Le système de reconnaissance	224
5.2.3.4. Les variables issues de cette méthode et introduites dans la base de données	225
5.2.4. Le Plan d'analyse	226
5.2.4.1. Gestion des données	226
5.2.4.2. Plan de l'analyse	226
5.3. Résultats .....	227
5.3.1. Description des locuteurs	227
5.3.1.1 Les témoins	227
5.3.1.2 Les patients	227
5.3.1.2.1 Données générales et carcinologiques (tableau n°5.3.1.1)	227
5.3.1.2.2 Description du traitement	228
5.3.1.2.3 Description du traitement	231
5.3.2. Résultats du point de vue des auditeurs	232
5.3.2.1. Etude du taux de bonne perception	232
5.3.2.1.1 Quelques vérifications	232
5.3.2.1.1.1 Le taux de perception juste par auditeurs	232
5.3.2.1.1.2 Les locuteurs témoins	233
5.3.2.1.1.3 Influence de la fréquence du segment sonore	236
5.3.2.1.1.4. A retenir avant d'aller plus loin....	237
5.3.2.1.2 Taux de perception pour le groupe patients	237
5.3.2.1.2.1. Résultat par locuteur	237
5.3.2.1.2.2. Les caractéristiques des patients influençant le taux de perception des consonnes	237
5.3.2.1.2.3. A retenir	240
5.3.2.1.3 Taux de perception par consonne et par trait pour les patients	240
5.3.2.1.3.1. Taux de perception par consonne	240
5.3.2.1.3.2. Taux de perception par traits	242
5.3.2.1.3.3. A retenir	243
5.3.2.2 Etude des confusions	244
5.3.2.2.1. Matrice de confusion	244
5.3.2.2.2. Calcul d'un taux de confusion	244
5.3.2.2.2.1. Par consonnes	244
5.3.2.2.2.2. Par traits	245
5.3.2.2.3. Description des confusions	246

5.3.2.2.4. A retenir	248
5.3.2.3 Pondération du taux de perception par le taux de confusion	248
5.3.2.3.1. La F mesure	248
5.3.2.3.2. A retenir par rapport aux confusions	250
5.3.2.4. Approche de la robustesse par les auditeurs	251
5.3.2.4.1. Etude par rapport à l'impression globale de sévérité	251
5.3.2.4.1. Etude par rapport aux taux de perception par locuteur	252
5.3.2.5. A retenir	252
5.3.3. Résultats du système automatique de reconnaissance vocale	253
5.3.3.1. Quelques vérifications préalables	253
5.3.3.1.1. Etude des résultats des locuteurs contrôles (au total 10 locuteurs, 5 femmes et 5 hommes)	253
5.3.3.1.1.2. Que nous apprend l'analyse des résultats obtenus avec les sujets contrôles sur la reconnaissance des syllabes du corpus ?	253
5.3.3.1.1.2. Vérification de l'alignement	254
5.3.3.1.2. Résultats des patients et des deux témoins dont le corpus a été utilisé lors du test de perception	256
5.3.3.2.1. Pour les syllabes	256
5.3.3.2.2. Description des résultats pour les consonnes	256
5.3.3.1.3. En résumé	258
5.3.3.2. Détermination de l'indice de vraisemblance de la consonne cible à partir du score de vraisemblance syllabique.	259
5.3.3.3. Analyse des résultats avec le rang médian	260
5.3.3.3.1. Pour les témoins	260
5.3.3.3.2. Pour les patients	261
5.3.3.3.3. Comparaison des résultats pour les deux groupes	262
5.3.3.3.4. A retenir	263
5.3.4. Test des hypothèses	263
5.3.4.1. Rappel des hypothèses	263
5.3.4.2. La comparaison de l'impact du degré de dégradation (obtenu par le système) sur l'intelligibilité (obtenu par le test de perception par les auditeurs).pour les consonnes.	265
5.3.4.2.1. La F mesure en fonction du rang médian pour les consonnes.	265
5.3.4.2.2. Le rappel ou taux de perception juste et la précision en fonction du rang médian pour les consonnes.	268
5.3.4.2.3. En synthèse	270
5.3.4.2 Interprétation des diagrammes $r_{cible}$ en fonction de $r_{percu}$	270
5.3.4.2.1. Recherche des unités linguistiques particulièrement robuste à la perception	271
5.3.4.2.2. Recherche des unités linguistiques particulièrement fragiles	272
5.3.4.2.3. Analyse de l'autre zone « extreme » à la recherche de consonnes source de confusion	273
5.3.4.2.4. En synthèse	274
5.4. Discussion .....	275
5.4.1. Synthèse des résultats	275
5.4.2. La robustesse aux distorsions	276
5.4.2.1. Les troubles de la production de la parole dans les cancers de la cavité buccale et de l'oropharynx	276
5.4.2.2. Comparaison avec les autres sources d'altération du signal	278
5.4.2.2.1. La comparaison par rapport aux autres sources d'altération d'origine pathologique	278
5.4.2.2.2. La comparaison avec des conditions de transmission altérant la qualité du signal de parole	279
5.4.2. Les méthodes pour évaluer l'impact du déficit de production sur la perception de la parole	280
5.4.2.1. Les limites des taux de perception juste	280
5.4.2.2. Les perspectives pour des évaluations plus pertinentes	282
5.4.3. La robustesse à la perception	283
5.4.3.1. Les limites d'interprétation de nos résultats	283
5.4.3.2. La recherche d'informations pouvant corroborer nos résultats	285
5.5. Conclusion.....	286



## 5.1. INTRODUCTION

---

### 5.1.1. LA NOTION DE ROBUSTESSE

---

Les tests d'intelligibilité pratiqués en clinique étudient tous les segments de la parole dans la mesure où ils apportent tous une information. Ainsi, les listes sont soit exhaustives, soit restreintes sans critères de sélection validés. Pour limiter le ratio temps de passation par rapport à la quantité d'information utile dans les bilans, pour déterminer les cibles les plus efficaces dans la prise en charge en rééducation ou dans les techniques augmentatives, la connaissance du degré de conservation d'une bonne perception d'un phonème ou d'une syllabe malgré leur distorsion pathologique est un des éléments qui permettraient d'établir des tests plus performants et/ou plus « économiques ». Cette connaissance en fonction des pathologies, pourrait être la base de travaux sur la restauration synthétique d'un signal de parole pathologique.

Cette question renvoie au problème de la « robustesse » des phonèmes à certaines contraintes ou expositions. Elle est plus particulièrement étudiée dans le domaine de la perception auditive en raison de l'aggravation des difficultés de compréhension dans le bruit des sujets présentant une perte auditive. Facteur limitant l'efficacité des appareils auditifs, les recherches sur ce sujet font l'objet d'un soutien industriel particulièrement actif.

A noter que la conception de robustesse par rapport à l'intelligibilité est différente de celle plus générale de « robustesse » utilisée en linguistique évoquée par :

- La fréquence des phonèmes dans les langues du monde (se réfèrent plutôt au système phonologique des langues)
- Le nombre d'allophone pour un même phonème (réfèrent plutôt au système phonétique des langues).



Il s'agirait là de la contribution « proportionnelle » pour l'intelligibilité, attribuée à chaque unité phonémique, celle-ci n'étant valable que dans un système linguistique donné dans la mesure où chaque langue possède son propre système phonémique.

Le bruit expose le signal de parole à des altérations d'un type bien précis en fonction du type de bruit, altérations différentes de celles liées aux TPPP. Si la première situation pose la question de la « résistance » du contenu du signal acoustique à l'exposition à des bruits environnementaux, la seconde comprend la résistance à la distorsion produite par une altération de la production corrigée au mieux par le locuteur par des adaptations de traits. La frontière avec des changements de catégories phonémiques induisant des confusions dépendant probablement du système phonologique de chaque langue. Il est également possible pour le locuteur d'utiliser des stratégies de niveau plus élevé, comme des changements de catégorie phonémique, anticipant ainsi d'éventuelles confusions (quand cela est possible sur le plan moteur). Ce dernier type d'adaptation est étroitement lié au système phonologique de la langue.

Il est possible que la robustesse perceptive d'un phonème, définie comme la contribution relative à l'intelligibilité de chaque unité phonémique dans son système linguistique soit la même dans ces deux situations si l'on fait l'hypothèse que l'intelligibilité dépend exclusivement des capacités de l'auditeur à extraire du signal qu'il perçoit les indices déterminant la compréhensibilité du message. Mais ceci restera à vérifier. En effet, le type d'altération du signal découlant de ces deux situations étant différent du fait qu'elles soumettent les unités phonémiques à des altérations différentes, les travaux dans ce domaine nécessitent des démarches parallèles avant d'envisager leur complémentarité. Cette question de la robustesse dans les TPPP et les méthodes pour y répondre demeurent un champ de recherche inexploré qui nous ont conduit à élaborer certaines hypothèses de recherche.

### 5.1.2.1. LES LIMITES DU CADRE EXPERIMENTAL USUEL

---

La manière d'appréhender le problème est plus complexe dans la situation des TPPP. En effet, dans le contexte de l'exposition au bruit, l'intention d'émission du message correspond généralement à la production réalisée par les locuteurs et donc (en dehors de quelques erreurs de production) au signal produit et enregistré. Alors que, comme la difficulté se situe entre l'intention et la production dans le cas des TPPP, il y a une différence entre l'intention et le signal enregistré et donc pas de connaissance réelle de ce qui est produit sur le plan phonétique. Ainsi, même si souvent dans les TPPP le signal produit par les locuteurs est identifié par la consigne et donc l'intention de produire, il est fondamental pour mener à bien cette recherche de savoir s'affranchir de cette approximation. Ainsi, nous allons d'abord mener une réflexion quant aux données recueillies par les expériences menées couramment pour l'étude de l'intelligibilité dans les TPPP.

Dans les situations expérimentales usuelles (figure n°5.1.1), il est demandé à un locuteur de lire le corpus, le signal enregistré est ensuite l'objet d'analyse perceptive par un jury déterminant l'intelligibilité. Nous connaissons donc l'intention du locuteur (le corpus ou une cible dans le corpus). Nous appellerons cette donnée identifiée en termes de contenu linguistique : C (donnée « consigne » ou « cible »). Le patient articule pour produire ce corpus. Mais ce temps ne fait pas l'objet d'une analyse, il n'y a pas de données s'y rapportant. Cette production génère le signal acoustique qui est enregistré, l'hypothèse courante étant simplement que le signal acoustique est le résultat direct de la production par le locuteur et que la qualité de la production est traduite par la qualité du signal. Nous disposons donc de sa nature physique mais ne connaissons pas sans analyse spécifique ses caractéristiques et notamment sa valeur en terme d'intelligibilité. Le contenu linguistique inconnu de ce signal sera appelé X. Ensuite, un groupe d'auditeur écoute le signal déterminant le contenu linguistique perçu de celui-ci. Les données issues de cette écoute sont connues. Nous l'appellerons P (donnée perçue).

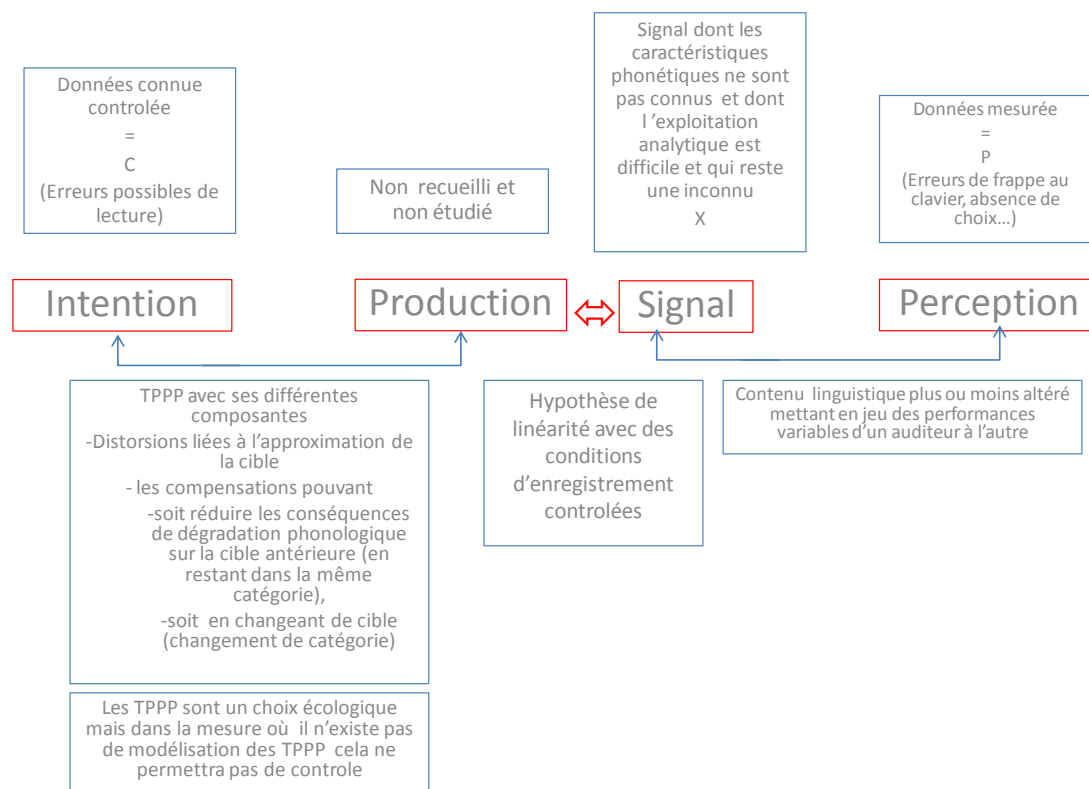


Figure n°5.1.1 : Schéma des expériences menées couramment pour l'étude de l'intelligibilité dans les TPPP.

Connaître le contenu phonétique du signal reste problématique. Les paramètres sont très nombreux, ils ne sont pas les mêmes en fonction de la catégorie phonétique. Dans l'état actuel des connaissances, il est très difficile, par une analyse acoustique d'obtenir une mesure unique ou globale de la distorsion d'un phonème par rapport à une norme. Seul les systèmes de reconnaissance vocale, du fait de leur système d'apprentissage offre la possibilité de comparer le signal enregistré à des unités génériques obtenues à partir de corpus de sujet témoin.

### 5.1.2.2. LES HYPOTHESES PROPOSEES

Introduire un système de reconnaissance automatique de la parole dans l'expérience permet, non pas de connaître directement « la nature » du signal mais son degré de vraisemblance par rapport à la cible et par rapport à la perception qui en a été faite par les auditeurs. Ainsi, si nous appelons R la donnée fournie par un système de reconnaissance automatique, cette donnée ne sera qu'une nouvelle estimation du signal X et ne contribuera pas à résoudre le problème. Par contre, la mesure

de vraisemblance fournie par le système pour la cible et pour la perception correspondante, dans la mesure où elle repose sur une paramétrisation unique, fiable et reproductible sera pertinente (figure n°5.1.2).

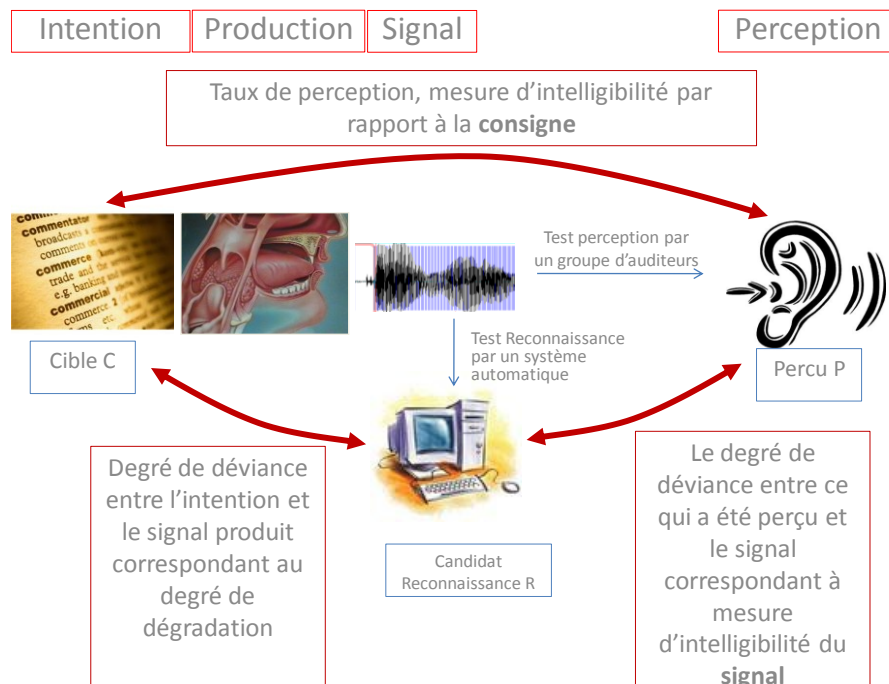


Figure n°5.1.2 : Les hypothèses de l'étude (C : donnée cible ou consigne, P Donnée perçue, X signal résultant de la production).

A partir de cette analyse, les hypothèses suivantes ont été émises.

Etant donné que :

1. La comparaison du signal X par rapport au générique de l'intention du locuteur, C (cible) évalue le degré de déviance entre l'intention et le signal produit et peut correspondre au degré de dégradation.
2. La comparaison du signal X par rapport au générique de l'unité linguistique perçue par l'auditeur évalue le degré de déviance entre ce qui a été perçu et le signal et peut correspondre à une mesure d'intelligibilité du signal, un degré d'intelligibilité.
3. La comparaison de la cible C par rapport à la perception par l'auditeur P correspond à une mesure classique de l'intelligibilité sous la forme d'un taux de bonne ou mauvaise perception.

Peut être proposé que :

1. L'étude du taux de perception par les auditeurs en fonction du degré de déviance par rapport à la cible C issu de l'analyse par le système de reconnaissance vocale permettra de déterminer s'il

existe des différences de robustesse perceptive en cas de distorsion entre les unités linguistiques étudiées.

2. L'étude de la répartition des unités linguistiques étudiées en fonction des distances obtenus par le système de reconnaissance contribuera également à l'étude de la robustesse à partir du raisonnement illustré par la figure n°5.1.3.

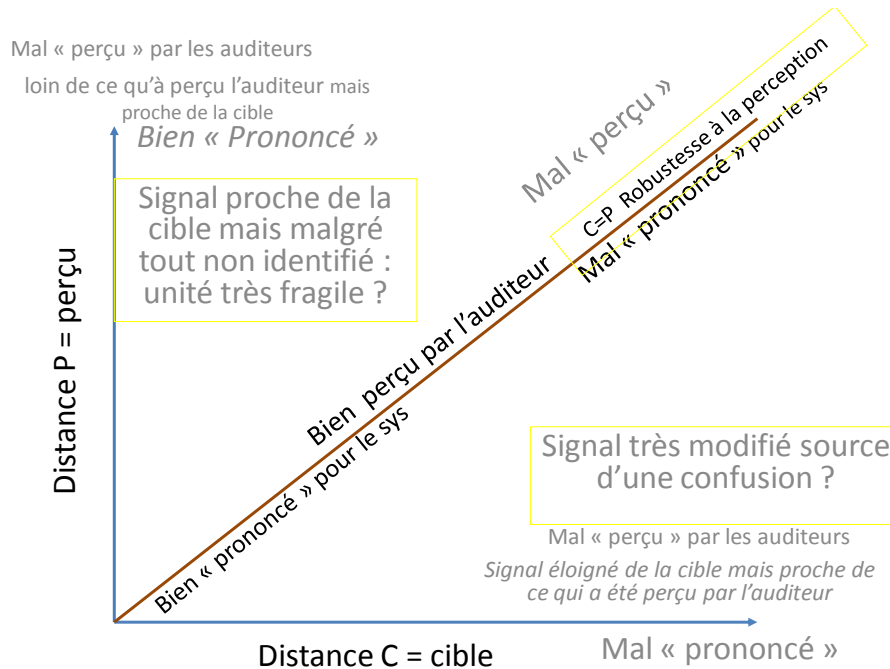


Figure n°5.1.3 : Hypothèses d'interprétation de la distance par rapport à la cible ou à la perception.

Ainsi, lorsque la cible est correctement perçue par les auditeurs, les points produiront une droite (C=P). La partie distale de cette droite correspond à des unités distantes de C, donc « mal prononcé » et distante de P donc loin du générique de l'unité perçue par les auditeurs et malgré cela bien perçu par les auditeurs. Cela peut donc être interprété comme une robustesse à la perception malgré une déviance importante dans la production. Ces sons pourront ensuite faire l'objet d'une analyse acoustique spécifique si ils sont repérables (soit par trait, soit par locuteur, soit par auditeur) pour étudier la nature du processus de « compensation ». Il pourrait s'agir d'une substitution permettant une bonne identification c'est-à-dire compatible avec une restauration par les auditeurs.

Lorsque la cible est mal perçue par les auditeurs, deux cas de figures « extrêmes » peuvent être décrits.

Si la distance par rapport à P (perçu) est grande alors que la distance par rapport à C (cible) est faible, alors l'explication serait que l'unité mal perçue correspond effectivement à un signal loin de ce qu'a perçu l'auditeur alors qu'il est plutôt bien prononcé. Cette unité serait donc « fragile » dans mesure où « peu » de distorsion suffirait à la rendre non « reconnaissable » et donc peu robuste.

Si la distance par rapport à C (cible) est grande alors que la distance par rapport à P (perçu) est faible, le signal est éloigné de la cible mais proche de ce qui a été perçu par l'auditeur. Dans la mesure où l'auditeur n'a pas perçu la « bonne » unité, mais une autre unité différente dont le signal est proche, il peut s'agir d'une substitution à l'origine d'une confusion (ne permettant pas une restauration sur la cible). Une vérification pourrait être obtenue par l'analyse des confusions sur ces groupes s'il existe.

Ainsi, l'utilisation d'un système automatique de reconnaissance vocale peut nous permettre

1. D'estimer un degré de dégradation du signal et de comparer l'impact de ce degré de dégradation sur l'intelligibilité en termes de taux de perception en fonction des phonèmes ou autres unités linguistiques (des traits à la syllabe) ;
2. De rechercher s'il existe des unités linguistiques particulièrement fragiles,
3. De rechercher s'il existe des unités linguistiques particulièrement exposés aux confusions ;
4. De rechercher s'il existe des unités linguistiques dont la distorsion donne lieu à des compensations par substitution.

A noter que l'analyse de la fréquence de la mauvaise reconnaissance entre l'intention et le perçu pourraient être suffisante pour déterminer la robustesse du phonème si il était possible d'étudier simplement un grand nombre de sujet. Or les contraintes de jury d'écoute et de vaste corpus pour approcher les situations usuelles de parole sont difficilement résolubles en pratique. De plus, il n'y a aucune certitude quand à l'universalité des distorsions pathologiques. Enfin la grande variabilité dans la qualité des productions reste dans tous les cas un écueil. C'est pourquoi l'analyse de la mesure d'intelligibilité par rapport au degré de déviance par unité phonétique étudiée (syllabe ou phonème) de la consonne dans des groupes de pathologie ciblé est la méthode retenue ici basée sur le raisonnement ci-dessus.

### 5.1.2.3. L'OBJECTIF DE CE TRAVAIL

---

. Au départ de ce travail de thèse, l'objectif était de vérifier si les phonèmes de la langue française présentait ou non une robustesse différente à la distorsion, et si oui, d'identifier le degré de robustesse de chacun d'entre eux. La difficulté que représentait la quantification du degré de distorsion des phonemes nous a paru résolvable par l'utilisation d'un système automatique de reconnaissance de la parole. Cependant, les hypothèses exposées ci-dessus, invitent à concentrer notre objectif sur la validation de ces hypothèses, avant de pouvoir assurer la fiabilité des résultats.

Ainsi, nous faisons l'hypothèse que l'utilisation d'un système de reconnaissance automatique de la parole associé à une tâche d'identification perceptive permet de déterminer la robustesse des phonèmes d'une langue et l'objectif de ce travail devient la validation de la méthode décrite ci-dessus

Cette décision détermine les choix méthodologiques exposés dans la section suivante.

### 5.1.2.4. LES RESULTATS ATTENDUS

---

1. La comparaison de l'impact du degré de dégradation sur l'intelligibilité en termes de taux de perception en fonction des phonèmes ou autres unités linguistiques (des traits à la syllabe) permettra de déterminer la robustesse des unités linguistiques. (figure n°5.1.4). Nous attendons trouver des consonnes ou des traits qui malgré la distance par rapport au générique de la cible maintiennent leur taux de perception. Nous espérons démontrer qu'il existe des unités linguistiques dont la distorsion donne lieu à des compensations, soit par ce que elle résiste à la distorsion (pente faible de la figure) soit parce qu'il existe des substitutions phonétiques donnant lieu à des restaurations phonologiques (présence de second pic sur la figure).

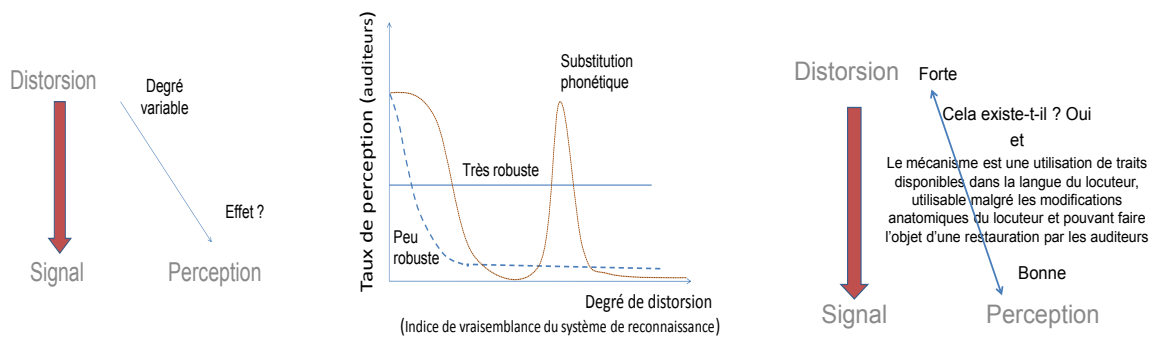


Figure n°5.1.4 : La comparaison de l'impact du degré de dégradation sur l'intelligibilité en termes de taux de perception. De gauche à droite :

1. Le degré de distorsion a un impact direct sur le signal produit avec à degré variable un impact théoriquement proportionnellement croissant sur la perception
2. En fonction des caractéristiques intrinsèques des unités linguistiques, cette relation n'est probablement pas linéaire et divers phénomènes correspondant à des configurations particulières sur un diagramme du taux de perception en fonction du taux de distorsion peuvent être décelés.
3. Nous espérons repérer des sons présentant un fort taux de perception malgré un fort taux de distorsion

Nous attendons également confirmer

- Qu'il existe des unités linguistiques particulièrement fragile qui correspondent à une courbe avec une pente rapidement décroissante du diagramme précédent et correspondant à des groupes identifiés sur le diagramme de la figure n°5.1.3. à une distance Perçu >> à la distance Cible soit une localisation dans la partie supérieure gauche.
- Qu'il existe des unités linguistiques particulièrement exposés aux confusions. Ces groupes seront identifiés par leur localisation en bas à droite sur le diagramme de la figure n°5.1.3 et cela sera contrôlé par un fort taux de confusion pour ces groupes par les auditeurs.

Ainsi, nous attendons valider nos hypothèses par l'étude de la concordance des résultats du test de perception et des résultats de vraisemblance du système de reconnaissance.

Cependant la généralisation des résultats ne sera pas possible du fait d'un profil de degré de déviance spécifique à la population de locuteur étudié par catégorie de trait ou de consonne. Le contrôle de ce biais demandera si les hypothèses sont validés de comparer les résultats avec d'autres groupes de TPPP.



## 5.2. CHOIX METHODOLOGIQUES

---

### 5.2.1. LES LOCUTEURS

---

Les troubles de la production pathologiques de la parole sont d'origines très diverses et conduisent à une grande variété de distorsion.

Pour cette expérience, l'objectif étant de valider une méthode de mesure de la robustesse, il a semblé judicieux de choisir un corpus de production pathologique de parole homogène quand au type de TPPP. Une origine structurelle par rapport à une origine neurologique nous est apparue la meilleure solution en raison :

1. Du moindre risque d'atteinte des fonctions supérieures associées,
2. De la stabilité des distorsions liés à la lésion des structures anatomiques
3. Au profil relativement homogène du type de distorsion résultant des étiologies.

Parmi les TPPP d'origine structurelle, nous avons recherché un groupe suffisamment homogène mais chez qui le nombre de distorsion soit suffisamment représenté. Les pathologies conduisant à un déficit lingual appariassent plus adapté à cet objectif par opposition a un déficit vélaire ou laryngé. Dans ce contexte, les cancers du fait des multiples mécanismes des séquelles thérapeutiques nous sont apparu particulièrement approprié comme modèle.

---

### 5.2.2. LE CORPUS

---

#### 5.2.2.1 CHOIX DU CORPUS

---

Le corpus utilisé pour cette expérience est composé de syllabes de structure ouverte construites avec les consonnes du français présentée sous la forme si possible d'un mot monosyllabique correspondant en une structure CV dans une phrase porteuse.

Nous avons choisi de cibler les consonnes car ce sont les unités les plus nombreuses, pour laquelle la notion de robustesse est la moins étudiée alors que ces phonèmes supportent une grande partie de l'intelligibilité de la parole (Trouver les références pour préciser ce rôle). La syllabe ouverte (terminée par une voyelle) est la plus fréquente de la langue française avec une majorité de syllabe (54.8% selon Delattre) de type CV. De plus les anomalies prédominent sur les phonèmes en position initiales dans les productions pathologiques (Kent).

Pour avoir une production se rapprochant de la situation naturelle de production, ces syllabes sont présentées au locuteur sous la forme de mots. Ceci n'a pas été possible pour 10 syllabes correspondant alors à un logatome (cf tableau n°5.2.2). Le choix d'une structure monosyllabique biphonémique est justifié par la réduction de la variabilité de la durée de la syllabe du fait de l'absence d'accentuation. Cela permet d'éviter la perte du naturel de lecture de logatome au niveau de la prosodie.

Les caractéristiques de la consonne étant déterminées par la voyelle suivante, chaque consonne sera étudiée en association à 5 voyelles : 3 universelles /a,u,i/ et 2 plus spécifiques à la langue française /y,an/.

Pour contrôler l'environnement de la syllabe, celle-ci est présentée dans une phrase porteuse « le mot\_\_\_\_\_convient ». Elle est ainsi précédée d'une même voyelle et suivi d'une même syllabe commençant par une occlusive pour faciliter la segmentation.

La liste des cibles figurent sur le tableau n°5.2.2. Ils ont fait l'objet d'un pré-test car la présentation orthographique de certains d'entre eux risquait d'induire des erreurs de prononciation. Ceci conduisant alors à les présenter sous une forme plus simple, « phonétique ». Sont ainsi produits 5 fois la même consonne dans un contexte vocalique différent. Le corpus est donc constitué de  $16 \times 5 = 80$  phrases.

Deux corpus complémentaires ont été constitués.

Le premier pour servir de leurre pour la tâche de perception. Il est constitué de mots monosyllabiques diphonémiques comprenant les voyelles manquantes. En effet, la tâche de perception des voyelles de la syllabe risque d'être faussée par la conscience des auditeurs de la réduction à 5 du nombre de voyelles ( « dé » « beau » « son » « feu » « main »).

Le deuxième par 5 locutrices et 3 locuteurs témoins supplémentaires pour la phase d'apprentissage du système de reconnaissance automatique de la parole.

#### 5.2.2.2 MODALITES D'ENREGISTREMENT.

---

En pratique, les syllabes cibles ont été présentées dans une phrase porteuse sous une présentation écrite correspondant à un mot monosyllabique quand cela est possible. La phrase porteuse est affichée sur un écran sans caractère ou ponctuation particulière (exemple : le mot chat convient) à l'aide du logiciel Viewer élaboré au Laboratoire Parole et langage d'Aix. L'ordre de présentation du corpus est randomisé.

L'enregistrement a été effectué avec un micro serre-tête Shure®, une carte d'acquisition Edirol® et un ordinateur Mac Intosh® dans le salle assourdie.

#### 5.2.2.3 PREPARATION DES STIMULI

---

Aucune modification n'a été réalisée sur le signal. Notamment aucune normalisation du niveau sonore dans la mesure où ce paramètre est souvent une caractéristique du TPPP.

La segmentation du signal a été effectuée avec le logiciel PRAAT (Paul Boersma & David Weenick 2009).

Les phrases des 20 locuteurs (10 témoins et 10 patients) ont été segmentées après annotation de leur contenu afin que le contenu puisse servir de nom de fichier. Toutes les pauses trop marquées ou les répétitions de phonèmes ou les absences de phonèmes ont été listés sur un fichier à part pour l'étape de la reconnaissance par le système informatique.

Le repérage et l'extraction de la syllabe cible a été réalisée pour les 2 locuteurs témoins masculin dont les stimuli ont été utilisé lors de la tache d'identification par le groupe d'auditeurs et les 10 patients. En raison des difficultés de repérage dans ce type de production pathologique et après plusieurs essais, les structures les plus stables ont été choisies pour l'extraction. Il s'agit de la fin de la voyelle précédent la syllabe et de la fin de la voyelle composant la syllabe cible. Au sein de ce segment le début de la syllabe est marquée par un « point tier ».

Idéalement, nous voulions extraire la syllabe à partir du milieu de la voyelle précédente pour disposer de toutes les informations contenues dans la structure du phonème précédent la consonne et la

transition. Mais les difficultés de repérage sont telles que cela était trop aléatoire. Nous avons donc privilégié une extraction identique de tous les phonèmes malgré la perte d'information que cela peut représenter. Nous avons ainsi priorisé la reproductibilité.

Ont été ainsi « préparé » sous la forme de fichiers au format « wave » 80x10 (patients) + 85x2 (témoins) stimuli correspondant à des syllabes et 80x20 stimuli correspondant aux phrases porteuses de la syllabe.

#### 5.2.2.4 PRESENTATION DES RESULTATS

---

Le choix des mots dans leur présentation orthographique est issu de l'analyse de la fréquence des mots relevé dans la base de données du site <http://www.lexique.org> ( Boris New & Christophe Pallier).

Pour injecter dans le tableau de données de l'étude, nous avons recueilli pour tous les syllabes :

1. La freqfilm du mot tel qu'il a été orthographié pour la lecture du corpus, c'est-à-dire la fréquence selon le corpus de sous-titres (par million d'occurrences) pour chaque intention d'émission des mots de notre corpus.
2. La freqlemfilms c'est-à-dire la fréquence du lemme selon le corpus de sous-titres (par million d'occurrences) pour chacun des mots de notre corpus. Lorsque plusieurs lemme existe pour la même forme phonétique correspond au mot demandé, nous avons additionné l'ensemble des lemmes correspondant phonétiquement
3. La fréquence de token des syllabes. (la fréquence token est la fréquence de la syllabe dans le corpus des mots de lexique.org c'est-à-dire par million d'occurrence).

Toutes ces données sont colligées sur le tableau n°5.2.2.

	a	i	u	ou	an
P	Pas	Pie	pu	Poux	Pan
Freq mots même ortho	18315,21	3,47	22208,67	8,49	7,05
Freq même phon	18338,13	37,62	5524,47	26,43	12,13
F Syll token	11965,4	1571,51	1358,64	1675,02	1774,02
B	Bas	Bi	Bu	Bout	Banc
Freq mots même ortho	171,82	1,27	55,52	123,9	8,96
Freq même phon	235,79	13,76	130,45	146,8	11,67
F Syll token	1986,31	1610,4	884,98	1144,41	342,44
M	Ma	Mi	Mue	Mou	ment
Freq mots même ortho	2347,09	8,97	0,42	7,9	24,08
Freq même phon	2350,97	259,1	1,61	10,89	109,67
F Syll token	6348,07	5727,18	1105,01	452,7	14223,3
T	Ta	Ti	Tu	Tout	tant
Freq mots même ortho	1265,97	0	14669,79	3993,93	380,34
Freq même phon	1331,25	0,93	14820,85	3997,94	1419,93
F Syll token	5844,84	7411,37	3084,76	4384,56	6297,13
D	Da	Dit	Dut	Doux	Dent
Freq mots même ortho	10,6	2162,97	1,91	53,97	13,27
Freq même phon	10,6	3190,63	4775,61	53,97	8423,45
F Syll token	1169,54	7877,07	8765,23	525,87	9391,59
N	Na	Ni	Nu	Nous	Nan
Freq mots même ortho	5,67	377,64	21,54	4772,12	11,92
Freq même phon	5,67	401,13	62,65	4772,81	12,24
F Syll token	2900,02	323,21	1939,64	4066,88	1190,01
S	sa	Si	Su	Sous	Sans
Freq mots même ortho	1276,38	3482,79	86,12	337,04	1003,44
Freq même phon	10259,27	4514,42	92,16	393,41	1662,75
F Syll token	1176,21	662,54	54,54	2880,44	5732,7
Z	Za	Zi	Zu	Zou	Zan
Freq mots même ortho	0	0	0	1,86	0
Freq même phon	0	0	0	1,86	0
F Syll token	1296,05	2067,95	143,99	10,06	1528,36
CH	Chat	Chie	Chu	Chou	Chant
Freq mots même ortho	57,71	4,31	1,24	25,50	17,64
Freq même phon	75,64	6,42	1,68	29,7	83,9
F Syll token	1176,21	662,54	54,54	65,4	1242,13
J	ja	Ji	Jus	Joue	Gens
Freq mots même ortho	0	0	22,36	148,45	594,29
Freq même phon	1,05	71,06	22,36	207,79	594,98
F Syll token	1696,02	1767,76	498,85	348,92	864,24
F	Fa	Fit	Fut	Fou	Fend
Freq mots même ortho	5,23	21,09	88,8	213,6	1,20
Freq même phon	5,23	35,42	196,01	488,83	1,74
F Syll token	2002,57	3855,23	1246,96	279,81	823,63
V	va	Vie	Vue	Vous	Vent
Freq mots même ortho	3382,89	986,59	198	13589,70	71,50
Freq même phon	4392,11	1174,62	1214,43	13590,46	123,44
F Syll token	2499,65	4615,84	760,32	3059,46	3144,61
K	Cas	Qui	cul	cou	Quand
Freq mots même ortho	270,66	5516,52	145,85	43,71	1830,76
Freq même phon	290,81	5517,69	154,3	522,21	1953,59
F Syll token	5307,77	7858,99	2060,35	3143,45	1768,78
G	Gars	Gui	Gu	Goût	Gant
Freq mots même ortho	289,63	2,90	0	50,51	9,86
Freq même phon	289,63	2,90	0	57,94	25,02
F Syll token	1585,55	133,88	505,74	505,74	115,94
R	Rat	Rit	Rue	Roue	Rang
Freq mots même ortho	24,21	17,47	128,5	13,72	19,40
Freq même phon	56,72	73,58	159,63	26,7	195,66
F Syll token	6234,83	5468,87	826,27	665,25	2490,5
L	La	Lie	Lu	Loue	Lent
Freq mots même ortho	16028,08	4,78	166,2	6,01	9,15
Freq même phon	18171,43	246,66	347,86	36,73	10,8
F Syll token	27614,1	5998,44	1764,29	143,41	1066,5

Tableau n°5.2.2. Tableau des mots et logatomes composé à partir des 16 consonnes cibles et de 5 voyelles associées à la fréquence du mot orthographique, du mot phonologique, de la fréquence de la syllabe dite Token extraites de la banque de données « lexique.org »

La description des phonèmes de ce corpus en termes de geste et de traits a également été intégrée au fichier de données. En ce qui concerne les données phonétiques (tableau n°5.2.3), nous avons utilisé le classement proposé dans le programme OTIM (Outils de traitement d'information multimodale) avec le codage suivant pour les gestes articulatoires des phonèmes.

-Au niveau des lèvres :

- Protrusion labiale = PL : protusion, aucune
- Aperture labiale = AL : fermée, critique, étroite, moyenne, large, Non pertinent (NP)

Au niveau de la Pointe de la langue

- localisation de la constriction de la pointe de la langue = LCPL : labiale, dentale, alvéolaire, postalvéolaire, palatale, Non pertinent (NP)
- Degré de constriction de la pointe de la langue = DCPL : fermée, critique, étroite, moyenne, large, Non pertinent (NP)

-Au niveau du corps de la langue

- Localisation de constriction du corps de la langue LCCL : palatale, vélaire, uvulaire, pharyngéale, Non pertinent (NP))
- Degré de constriction du corps de la langue = DCPL : fermée, critique, étroite, moyenne, large, Non pertinent (NP)

-Au niveau du voile

- Aperture vélaire = VEL : fermé, large

Au niveau de la glotte

- Aperture glottique = GLO : fermé, critique, étroite, large.

-Au niveau du type de phonème :

- OV (Voyelles Orales),
- NV (Voyelles Nasales),
- OPLO (Plosives orales),
- NPLO (Plosives Nasales),
- FRI (Fricatives), LIQ (Liquides),
- GLI (Glissantes)

	PL	AL	LCPL	DCPL	LCCL	DCPL	VEL	GLO	TYPE
p	NP	fermé	NP	NP	NP	NP	fermé	large	OPLO
t	NP	NP	alvéolaire	fermé	NP	NP	fermé	large	OPLO
k	NP	NP	NP	NP	vélaire	fermé	fermé	large	OPLO
b	NP	fermé	NP	NP	NP	NP	fermé	fermé	OPLO
d	NP	NP	alvéolaire	fermé	NP	NP	fermé	fermé	OPLO
g	NP	NP	NP	NP	vélaire	fermé	fermé	fermé	OPLO
m	NP	fermé	NP	NP	NP	NP	large	fermé	NPLO
n	NP	NP	alvéolaire	fermé	NP	NP	large	fermé	NPLO
f	NP	critique	NP	NP	NP	NP	fermé	large	FRI
s	NP	NP	alvéolaire	critique	NP	NP	fermé	large	FRI
v	NP	critique	NP	NP	NP	NP	fermé	fermé	FRI
z	NP	NP	alvéolaire	critique	NP	NP	fermé	fermé	FRI
l	NP	NP	postalvéolaire	fermé	NP	mid	fermé	fermé	LIQ
R	NP	NP	NP	NP	uvulaire	étroit	fermé	fermé	LIQ
j	NP	NP	NP	NP	palatale	étroit	fermé	fermé	GLI
ch	protrusion	NP	NP	NP	palatale	étroit	fermé	fermé	GLI
a	NP	NP	NP	NP	NP	moyenne	fermé	fermé	OV
i	NP	NP	NP	NP	palatale	étroit	fermé	fermé	OV
u	protrusion	NP	NP	NP	palatale	étroit	fermé	fermé	OV
ou	protrusion	NP	NP	NP	vélaire	étroit	fermé	fermé	OV
an	NP	NP	NP	NP	vélaire	large	large	fermé	NV

Tableau n°5.2.3. Tableau des caractéristiques phonémiques des phonèmes utilisés comme cibles dans le corpus.

---

### 5.2.3. LA PROCEDURE DE LA TACHE DE PERCEPTION

---

#### 5.2.3.1. DESCRIPTION DE LA TACHE

---

La tache de perception correspond à une tache d'identification reposant sur la transcription orthographique de la syllabe en procédant en deux temps, d'abord la transcription orthographique avec les 16 possibilités puis une transcription pour la voyelle avec 11 possibilités (les 5 voyelles utilisés plus des leurres soit a,eu,é,i,o,ou,u,a,an,on).

Le logiciel PERCEVAL permet d'assurer une présentation sonore randomisée du corpus de syllabe. La transcription a été faite en utilisant les touches d'un clavier spécial, reprenant uniquement les consonnes constitutives du corpus et les voyelles regroupés sous la forme de macroclasses pour les « O » et les « È » et les « EU » (tableau n°5.2.3). Dans la mesure où le corpus ne comprend pas toutes les voyelles, des « leurres » seront glissés dans le corpus pour ne pas focaliser les choix des auditeurs sur un nombre limité de voyelles (réduit aux 5 voyelles choisies).

Cette forme de choix forcée élargie à la majeure partie des phonèmes du français oblige à choisir le plus proche du son perçu.

Une transcription plus fine (avec usage des diacritiques par exemple) serait source d'une augmentation ingérable de la variabilité des auditeurs

Au total, 80 syllabes par patients (10 patients), Soit 800 syllabes et 85 syllabes par témoins, 2témoins soit 170 syllabes (corpus de base incrémenté avec 5 syllabes composées avec d'autres voyelles pour servir de leurre) ont été extraites au total pour servir de stimuli dans cette manipulation.

### 5.2.3.2. LES AUDITEURS

---

En tenant compte des résultats de l'expérience de catégorisation précédant ce travail, un jury d'auditeur averti mais non expérimenté telle que les étudiants en orthophonie de première année associent les avantage suivants :

- Population ayant un niveau d'étude homogène
- Evaluation audiométrique datant d'un an
- Connaissance en phonétique théorique
- Absence d'expérience clinique et donc d'habituatation potentielle au trouble pathologique de l'articulation de la parole.

De plus des examinateurs expérimentés risquent d'avoir développer des stratégies différentes d'analyse. il n'est pas certains que cela diminue la variabilité des critères de jugement.

Un jury de 11 auditeurs en cours de deuxième année d'orthophonie a été constitué.

### 5.2.3.3. LA PROCEDURE

---

#### 5.2.3.3.1 PREPARATION DE LA PRESENTATION DES STIMULI SUR LE LOGICIEL PRECEVAL

Après des essais, une estimation de 30 minutes par locuteurs, soit une durée totale de environ 6H00 par auditeur semblait nécessaire.

Le corpus a donc été divisé en 6 blocs (170\*1 et 160\*5) pour une passation sous la forme de 3 sessions de 2H00 avec une pause entre 2 blocs pendant chaque session. L'ensemble de la liste a été



randomisé puis segmenté en 6. L'ordre des blocs a été randomisé pour chaque auditeur. Au total un tableau de 1940 trials a été créé.

### 5.2.3.3.2 PASSATION

Chaque auditeur muni d'un casque a réalisé l'expérience installé devant l'écran de l'ordinateur dans une pièce calme. Le clavier a été modifié, les touches du clavier numérique étant maquillé pour permettre l'inscription des symboles correspondant aux voyelles. Avant de démarrer, le clavier avec ces modifications était présenté à l'auditeur. Ensuite, une phase d'entraînement comprenant 5 stimuli « leurres » prononcés par les deux témoins, permettait la familiarisation avec la tâche. Puis, l'auditeur rentrait dans la phase proprement dite de l'expérience.

La consigne était la suivante :« Vous allez entendre des mots monosyllabiques. Votre tâche sera d'identifier la consonne de ce mot en appuyant sur la touche du clavier correspondant puis d'identifier la voyelle en appuyant sur la touche du clavier correspondant. Cliquez sur un bouton pour commencer... »

En fonction des auditeurs la tâche a pris entre 30 minutes et 45 minutes par bloc. Une pause de 15 minutes a été respectée entre chaque bloc et aucun auditeur n'a réalisé plus de 2 blocs par jour. Tous les auditeurs ont pu réaliser l'expérience complète sur un intervalle inférieur ou égal à une semaine.

Touche clavier modifié	Touche originale
P	p
B	b
M	m
T	t
D	d
N	n
S	s
Z	z
H	ch
J	j
F	f
V	v
K	k
G	g
R	r
L	l

Touche clavier modifié	Touche originale
A	0
E	1
EU	2
O	3
OU	4
U	5
I	6
AN	7
UN ou IN	8
ON	9

Tableau n°5.2.3 : Transcription pour le clavier dont les touches ont été spécialement modifiées pour les auditeurs

### 5.2.3. LA METHODE DE LA RECONNAISSANCE AUTOMATIQUE DE LA PAROLE.

Le traitement des données a été réalisé par Corine Fredouille au Laboratoire Informatique d'Avignon (LIA) sur la base d'une méthode de reconnaissance de mots isolés utilisé en particulier pour la reconnaissance de chiffres ou de commandes simples pour les téléphones portables (Lévy 2006, et Lévy et al 2004, Auran C 2004). Le système repose sur une boîte à outils développée par le LIA.

Les systèmes classiques de reconnaissance automatique de la parole (RAP) sont composés essentiellement de 5 modules (figure 5.2.3.1) :

- la paramétrisation qui permet de ne garder que les informations pertinentes du signal,
- la modélisation acoustique où les modèles acoustiques doivent représenter au mieux les unités acoustiques choisies (phonèmes, diphones, mots...),
- la modélisation linguistique qui doit conduire à une représentation la plus vraisemblable possible du langage,
- le dictionnaire qui contient l'ensemble des mots que l'on souhaite pouvoir reconnaître (il peut dans certains cas être spécifique à une application),
- le système de reconnaissance lui-même.

Dans le cadre de ce travail un système capable de faire du décodage de mots isolés est plus adapté qu'un système de reconnaissance de parole continue dans la mesure où nous cherchons à connaître le degré de vraisemblance par rapport à une syllabe « générique » de la syllabe présentée comme un mot dans notre corpus. Ceci permet de s'affranchir des modèles de langage et de réduire le dictionnaire à la combinaison des consonnes et voyelles offertes aux auditeurs dans la tâche de perception soit 160 syllabes (16 Cx10 V).

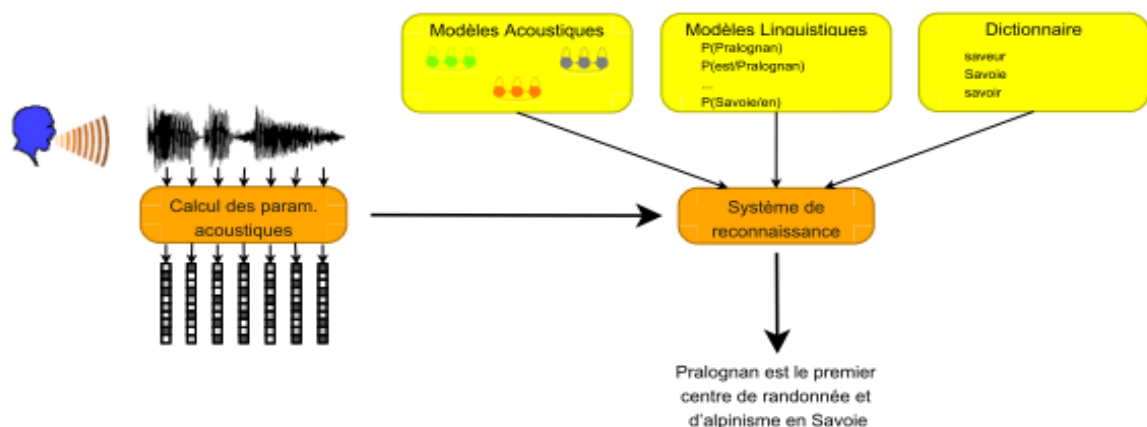


Figure 5.2.3.1 : Schéma de principe d'un système de reconnaissance automatique de la parole (source : C. Lévy)

### 5.2.3.1 LE PROCESSUS DE PARAMETRISATION

---

Le processus de paramétrisation consiste à extraire du signal les informations caractéristiques en éliminant au maximum les parties redondantes. La représentation simplifiée du signal fournie repose sur des vecteurs de paramètres acoustiques. Il est basé ici sur une analyse dans le domaine cepstral (Mel Frequency Cepstral Coefficients) avec une paramétrisation en 12 coefficients obtenus suite à une analyse en banc de filtre. Ils sont complétés par l'énergie du signal. Les dérivés de premier et second ordres sont calculées et concaténées avec le vecteur initial afin d'obtenir un vecteur de 39 coefficients pour chaque trame de signal.

Plusieurs méthodes permettent d'obtenir des coefficients cepstraux. Les coefficients Mel Frequency Cepstral (MFCC°) sont obtenus par l'utilisation d'une FFT et d'une FFT inverse avec des filtres répartis en fonction d'une échelle Mel (figure 5.2.3.2) où  $F_{mel} = 2595 \cdot \log(1 + F_{Hz}/700)$ .

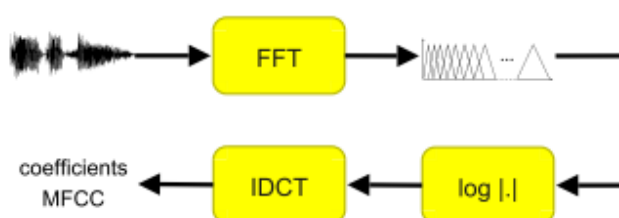


Figure 5.2.3.2. Chaîne de traitement pour obtenir les coefficients MFCC (source : C. Lévy)

L'énergie correspond à la puissance du signal. Le calcul se fait généralement sur des fenêtres glissantes de 25ms avec un décalage de 10ms soit une valeur toutes les 10ms de signal.

Pour enrichir la paramétrisation les dérivées de premier et second ordres ajoutent de l'information concernant la dynamique du signal

### 5.2.3.2 LA MODELISATION

---

Le principe de décodage est d'estimer une mesure de similarité entre la représentation d'un mot de référence et d'un mot inconnu afin d'évaluer l'écart entre ces deux mots. Le but est de réaliser un alignement temporel, le meilleur qui soit, une référence et un mot à tester.

Ici les phonèmes sont modélisés par des modèles de Markov cachés (HMM pour Hidden Markov Model) gauche-droite, à 3 états émetteurs (figure 5.2.3.3).

Un modèle de Markov est un automate fini qui change d'état à chaque unité de temps. Il repose sur une machine à états (chaîne de Markov) c'est-à-dire d'une succession d'états associés à des probabilités de transition d'un état à l'autre. Ces modèles permettent de caractériser les variations temporelles du signal de parole. Une ou plusieurs distributions de probabilité associées à chaque état caractérisent les probabilités d'émission des vecteurs acoustiques par un état. Une trame fait 10s. 3 trames (30s) sont le minimum pour la majorité des phonèmes sauf pour le burst où est attribué un seul état en plus. La fonction de densité de probabilité est composée de mixtures de gaussiennes (Gaussian mixture model : GMM)). Lors de la reconnaissance, la vraisemblance pour qu'une séquence de vecteurs de test soit issue de la chaîne de Markov est calculée. La méthode utilisée pour la ré-estimation des paramètres  $\lambda$  cherche à maximaliser la vraisemblance aussi appelée MLE pour Maximum de Likelihood Estimation.

Ici, chaque état des HMM est composé d'un mélange de 8 gaussiennes représentés par leur moyenne et la diagonale de la matrice de covariance. Ces paramètres sont estimés grâce à l'algorithme E-M (Expectation-Maximisation) en optimisant les critères du maximum de vraisemblance.

Les modèles de phonèmes ont été appris sur les données du corpus ESTER : Évaluation des Systèmes de Transcription d'Émissions Radiophoniques. Ce corpus comprend l'ensemble des données utilisées lors de la campagne d'évaluation ESTER, à l'exclusion des données textuelles. Elles reposent essentiellement sur les archives du journal Le Monde et du corpus des débats du Conseil européen, à savoir 100 heures d'émissions transcrites orthographiquement et annotées en entités nommées, soit 1,2 millions de mots pour un vocabulaire de 37 000 mots. 74 082 occurrences d'entités nommées sont annotées pour un total de 15 152 entités différentes.

Cette paramétrisation a conduit à extraire 38 phonèmes « génériques » sous la forme de vecteurs acoustiques.

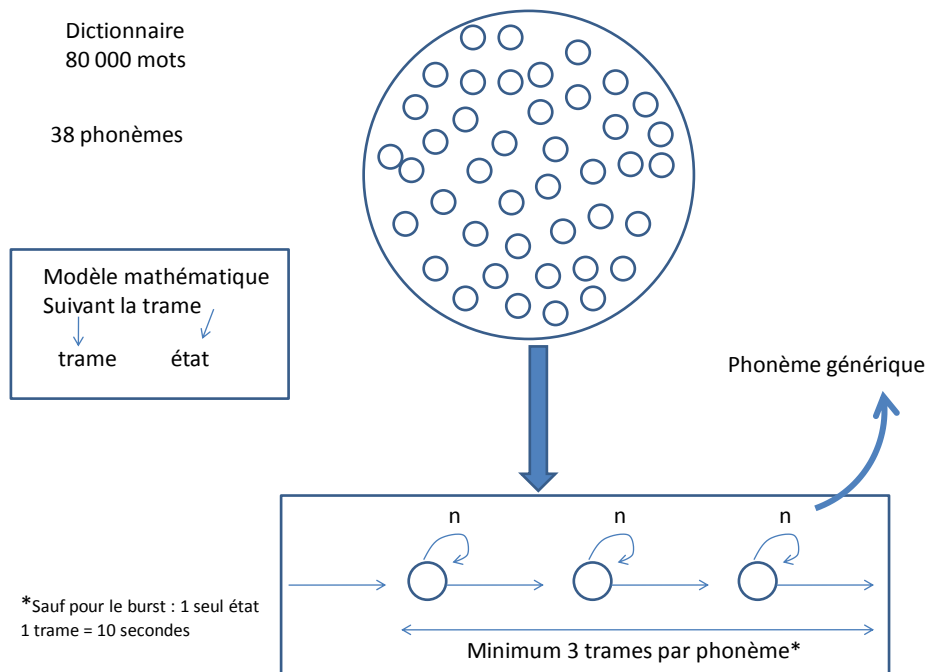


Figure 5.2.3.3. 38 phonèmes génériques par l'apprentissage du système utilisant une modélisation par HMM

### 5.2.3.3. LE SYSTEME DE RECONNAISSANCE

La reconnaissance consiste à retrouver le mot parmi un ensemble de mot possibles en fonction d'une suite d'observations (vecteurs acoustiques). Elle est fondée sur l'algorithme de décodage Viterbi. Cet algorithme calcul la séquence optimale (la plus vraisemblable) étant donné une série d'observations (maximisation de la probabilité d'émission d'un mot en fonction de la suite d'observations).

L'alignement est forcé, c'est-à-dire que la séquence de phonèmes étant pré-définie, les transitions entre les phonèmes sont fixées. Cet algorithme est donc également utilisé pour affecter les trames aux états, et à partir de la phrase porteuse c'est-à-dire d'un contexte contraint (le mot...convient), d'identifier mot « syllabe » contenu dans la phrase.

Alors que ce processus à généralement pour tâche de rechercher la mesure de similarité maximale et de désigner le candidat correspondant, cette phase n'est pas utile pour notre objectif et nous recueillerons les données fournies pour tous les candidats possibles, c'est-à-dire les mesures de vraisemblance et le classement par rang des candidats potentiels.

La méthodologie développée ici a été appliquée au corpus des 10 locuteurs témoins et des 10 locuteurs présentant un TPPP.

Les fichiers contenant les 80 cibles de chaque locuteur a été écouté pour noter les éventuelles « erreurs » de lectures : pauses, répétitions ou élisions ont été notées sur un fichier. Le nom du fichier comprend la transcription orthographique du contenu, précédé du code d'anonymat du locuteur.

#### 5.2.3.4. LES VARIABLES ISSUES DE CETTE METHODE ET INTRODUITES DANS LA BASE DE DONNEES

Les données issues de la reconnaissance par le système automatique comprennent la liste des 160 candidats possibles classés par ordre décroissant de vraisemblance. La normalisation des mesures de vraisemblance obtenus a été faite secondairement. Le score normé de  $x = \text{score candidat} / \text{score candidat } x$ , le maximum est = à 1. Ont été « injectés » dans la base de données (tableau n°5.2.4) :

- les scores normés de la syllabe cible et de la syllabe perçu
- le rang médian de la syllabe cible et de la syllabe perçue
- la syllabe candidat 1
  - o cela permettra de déterminer la consonne candidat 1 et la voyelle candidat 1
- la syllabe candidat 2
- la syllabe candidat 3
  - o cela permettra de chercher si la syllabe cible se trouve parmi les 3 premiers candidats et si la syllabe perçue se trouve parmi les 3 premiers candidats
- le rang médian de la consonne cible
- Le rang médian de la consonne perçu.

Locuteur	Auditeur	Système de reconnaissance automatique de la parole
Syll_cible (1)	syll_percu (11)	syll_reco_1
		syll_reco_2
		syll_reco_3
		Rang_syll_cible
		Rang_syll_percu
		Score_normé_syll_cible
		Score_normé_syll_percu
C_cible	C_percu	C_reco_1
		Rang_C_médian_cible
		Rang_C_médian_percu
V_cible	V_percu	Rang_médian_cible_V

Tableau n°5.2.4 : Synthèse des variables extraites du traitement par le système automatique

---

## 5.2.4. LE PLAN D'ANALYSE

---

### 5.2.4.1. GESTION DES DONNEES

---

Le logiciel PERCEVAL utilisé pour la passation du test de perception permet d'intégrer en amont les données caractérisant les locuteurs et le corpus.

Les données issues de la reconnaissance vocale ont secondairement été intégrées dans ce fichier en utilisant une clé relatif à la syllabe ou mot et au locuteur.

Les logiciels « excel » et « R » ont servi au traitement statistique des résultats. Ce traitement est basé sur l'analyse graphique.

### 5.2.4.2. PLAN DE L'ANALYSE

---

La description des résultats comprend les descriptions :

- a. des caractéristiques des locuteurs
- b. des résultats de l'analyse perceptive par les auditeurs
- c. des résultats de la reconnaissance par le système automatique.

Cette description a permis la détermination :

- a. des variables dont la fiabilité est soumise à caution pour le traitement final de l'expérience
- b. des données pertinentes pour la finalisation de la base de données, source de ce travail et la génération des variables issues de ce premier niveau de réflexion.

Les hypothèses émises dans l'introduction de ce chapitre ont été appliquées en

- a. construisant les représentations du taux de perception juste et d'autres variables apparues aussi pertinentes en fonction du rang de la consonne cible,
- b. en étudiant la distribution des consonnes en fonction du rang de la consonne cible lorsque le test de perception donné un résultat juste par consonne,

- c. en étudiant la distribution des consonnes dans les zones « extrêmes » du diagramme rang de la consonne perçue en fonction du rang de la consonne cible.

---

### 5.3. RESULTATS

---

---

#### 5.3.1. DESCRIPTION DES LOCUTEURS

---

Dix patients et 2 témoins ont été inclus dans cette étude.

##### 5.3.1.1 LES TEMOINS

---

Deux hommes âgés de 44 et 47 ans ont accepté de participer en tant que sujet contrôle à cette étude. Tous les deux originaires du Sud de la France, Perpignan pour l'un, Toulouse pour l'autre parlent avec un accent régional comme la majorité des patients. Aucun d'entre eux ne présente d'antécédents médicaux notables. Ils ont tous les deux une articulation jugée « normale ».

##### 5.3.1.2 LES PATIENTS

---

###### 5.3.1.2.1 DONNEES GENERALES ET CARCINOLOGIQUES (TABLEAU N°5.3.1.1)

Six hommes et 4 femmes âgés en moyenne de 58,7 ans (minimum 34, maximum 72 ans) au moment de l'enregistrement. Tous vivent dans la région Midi Pyrénées, 4 dans le Tarn (82), 3 en Haute Garonne (31), 2 dans l'Aude et un dans le Gers. Seule la personne vivant depuis plus de 20 ans dans le Gers et originaire d'Alsace et n'a pas l'accent régional.

En ce qui concerne les données cancérologiques, tous les patients ont présenté dans le cadre du traitement d'un carcinome épidermoïde des voies aérodigestives, un temps chirurgical d'exérèse de la tumeur comportant la résection d'une partie de langue (glossectomie partielle). Dans 3 cas la lésion était principalement localisée dans la cavité buccale intéressant principalement la langue mobile. Dans 7 cas, elle était d'origine oropharyngée partant du sillon amygdaloglosse dans 3 cas, de la base de la langue dans 2 cas, de la jonction base langue mobile dans 2 cas. Il s'agissait dans 5 cas de tumeurs classées T2, dans 4 cas T3 et un cas T4 en rapport avec l'estimation en surface de la lésion (cf Classification TNM en annexe 5.3.1.1). Quatre patients présentaient des adénopathies cervicales.



Aucun ne présentait de métastases au moment du diagnostic. Le délai par rapport à la fin du traitement est en moyenne de 5 ans et demi avec un minimum de 1 an et un maximum de 11 ans.

	Origine	Sexe	Age au moment diagnostic (années)	Délai entre la fin du traitement et l'enregistrement (années)	Localisation région	Localisation site	T	N	M
ALJ	Aude	H	64	8	Oropharynx	Langue	2	0	0
BAC	Aude	H	59	3	Oropharynx	Base de langue	2	0	0
BOF	Tarn	H	25	9	Oropharynx	Base de langue	3	0	0
BOJ	Haute Garrone	F	68	9	Oropharynx	Langue	4	1	0
COJ	Tarn	F	60	1	Cavité buccale	Langue	3	0	0
DEC	Haute Garrone	F	35	11	Cavité buccale	Langue	2	0	0
EUP	Haute Garrone	H	63	2	oropharynx	Amygdalogue	3	1	0
GAJ	Tarn	H	64	8	Cavité buccale	Langue	2	0	0
RAC	Tarn	F	40	1	oropharynx	Amygdalogue	3	1	0
WIF	Alsace/gers	H	54	3	Oropharynx	Amygdalogue	2	2a	0

Tableau n°5.3.1.1 Données générales

#### 5.3.1.2.2 DESCRIPTION DU TRAITEMENT

Le traitement carcinologique de ces tumeurs (tableau n°5.3.1.2) repose sur la résection chirurgicale, la radiothérapie et ou la chimiothérapie. Pour deux patients, l'intervention chirurgicale a été réalisée en raison de l'évolution de la maladie malgré un traitement par radiothérapie et chimiothérapie avant.

Dans 6 cas, le geste chirurgical était suivi d'une radiothérapie en raison soit des limites de résection jugées insuffisantes, soit en raison du statut ganglionnaire. Dans 2 cas, le geste chirurgical était suivi d'une radiochimiothérapie pour les mêmes raisons. Le temps chirurgical associé pour tous les patients un temps d'évidement ou curage cervical pour les ganglions et un temps de résection de la tumeur.

La résection chirurgicale a consisté en une oropharyngectomie pour 7 patients avec des résections étendues principalement à la base de la langue dans 4 cas, à la partie postérieure de la langue mobile et à la base de la langue dans 2 cas et à la partie postérieure de la langue mobile dans 1 cas. La voie d'abord est alors une buccopharyngectomie transmandibulaire dont le principe est exposé en annexe (annexe 3.3.1.2). Une résection de la moitié de la mandibule était nécessaire dans 3 cas et une résection était possible dans 2 cas. Dans 3 cas, la résection intéressait la langue principalement la langue mobile (partie postérieure et antérieure). Une voie d'abord naturelle, c'est-à-dire par la bouche a pu être réalisée, évitant ainsi les séquelles propres à la voie d'abord. Lorsque la surface de

la résection le justifie et que la situation vasculaire du patient l'a permis, le comblement de la zone de résection a été effectué par un lambeau. L'ensemble de ces informations est colligée dans le tableau n°5.3.1.2.

	Année Traitement	Voie D'abord	Type résection	Traitement	Résection linguale	Résection mandibulaire	Lambeau	Complications éventuelles
ALJ	2002	BPTM	oropharyngectomie	Chirurgie + RTE	Postérieure mobile et base de langue	Oui sans reconstruction	Aucun	Risque vasculaires
BAC	2007	BPTM	oropharyngectomie	Chirurgie + RTE	Base de langue	non	Grand Pectoral	Mélanome Lèvre supérieure
BOF	2001	BPTM	oropharyngectomie	Chirurgie + RTE+CT	Base de langue	Non	Antébrachial	
BOJ	2001	BPTM	oropharyngectomie	Chirurgie + RTE	Postérieure mobile et base de langue	non	Grand Pectoral	Ostéoradionécrose mandibulaire
COJ	2009	Naturelle	Glossectomie partielle	RTE+CT puis Chirurgie	Postérieure et antérieure mobile	non	lambeau de buccinateur	Ostéoradionécrose mandibulaire
DEC	1999	Naturelle	Glossectomie partielle	Chirurgie + RTE	Postérieure et antérieure mobile	non	Aucun	
EUP	2008	BPTM	oropharyngectomie	RTE+CT puis Chirurgie	Base de langue	oui avec reconstruction par plaque	Grand Pectoral	
GAJ	2002	Naturelle	Glossectomie partielle	Chirurgie + RTE	Postérieure et antérieure mobile	non	Aucun	
RAC	2009	BPTM	oropharyngectomie	Chirurgie + RTE+CT	Postérieure mobile	oui avec reconstruction par plaque	Antébrachial	
WIF	2007	BPTM	oropharyngectomie	Chirurgie + RTE	Base de langue	Non	Grand Pectoral	

Tableau n°5.3.1.2 : description des modalités thérapeutiques

La zone de la résection est schématisé, sur un même figure (figure n°5.3.1.1), sans tenir compte du coté de la tumeur pour chaque patient.



Figure n°5.3.1.1: zone de la résection schématisée sur une vue de  $\frac{3}{4}$  de carrefour aérodigestif (sans tenir compte du côté de la tumeur) pour chaque patient.

### 5.3.1.2.3 DESCRIPTION DU TRAITEMENT

En ce qui concerne la situation fonctionnelle des patients à la date de l'enregistrement, le délai après le traitement étant supérieur à 1 an, la phase de récupération est considérée comme terminée et les séquelles sont considérées comme stables ou lentement évolutives (tableau n°5.3.1.3). En effet, l'association chirurgie et radiothérapie avec le temps conduit à des modifications de la qualité des tissus avec notamment un processus de fibrose qui induit une limitation de l'ouverture de la bouche, des problèmes de dentition... De plus, le statut fonctionnelle est la résultante de l'ensemble du traitement et des éventuelles complications ou pathologies intercurrentes survenues à distance. Ainsi, deux patients ont présentés une ostéoradionécrose mandibulaire et un, un mélanome de la lèvre supérieure.

Ainsi ; bien que tous les patients aient bénéficiés d'une prise en charge rééducative comprenant au moins une rééducation orthophonique pour la déglutition et/ou la phonation, leur situation avec le délai par rapport à la fin du traitement continue à s'individualiser. Pour cette raison les données cliniques que nous avons recueillies sont limitées aux possibilités dynamiques des structures anatomiques suivantes : sangle labio-jugale, langue mobile, isthme oropharyngé et mouvement mandibulaire. L'échelle utilisée est ordonnée à 4 niveaux, 4 étant le niveau d'atteinte le plus sévère et 1 le niveau le moins sévère. Enfin a été recueilli un indice d'altération globale subjectif de l'intelligibilité noté de 0 (pas d'altération) à 4 (altération la plus sévère).

	Degré d'altération du TPPP (grade)	sangle Labio-jugale	Langue	Isthme oropharyngée	Mandibule
ALJ	2	2	2	2	2
BAC	3	3	2	3	3
BOF	3	3	2	3	3
BOJ	4	2	2	3	2
COJ	3	3	1	2	1
DEC	2	4	3	3	2
EUP	2	3	2	3	3
GAJ	2	3	2	3	3
RAC	3	3	2	2	3
WIF	2	3	3	2	3

Tableau n°5.3.1.3 : description des données fonctionnelles recueillies.

---

## 5.3.2. RESULTATS DU POINT DE VUE DES AUDITEURS

---

Ces résultats sont issus du traitement des données obtenues par la tâche de perception réalisée par les 11 auditeurs. Nous traiterons d'abord le taux de bonne perception à partir du nombre de perception correcte, puis le taux de confusion à partir des erreurs et enfin la pondération du taux de perception par le taux de confusion. Les tableaux ayant permis la réalisation des représentations graphiques figurent dans l'annexe à la section 7.3.1.2.

### 5.3.2.1. ETUDE DU TAUX DE BONNE PERCEPTION

---

La consigne dans cette expérience préconisait un choix forcé. Cependant, après un délai de 8 secondes, si l'auditeur n'appuyait pas sur une touche du clavier, l'expérience se poursuivait en inscrivant dans le fichier de résultat « xxx », permettant ainsi de différencier la bonne perception notée « ok », de la confusion avec un autre phonème, codée « err », de l'absence de réponse dans un délai largement suffisant (codé « xxx »). Nous faisons l'hypothèse que de non appuyer signifie que la reconnaissance n'a pas été possible et que cela ne modifie donc pas le pourcentage de bonne perception dans nos résultats.

Avant de décrire les résultats, la recherche des facteurs pouvant influencer leur interprétation est réalisée.

#### 5.3.2.1.1 QUELQUES VERIFICATIONS

Avant de tirer des conclusions, une vérification préalable en termes de comportement des auditeurs et des résultats obtenus auprès des locuteurs témoins est réalisée.

##### 5.3.2.1.1.1 LE TAUX DE PERCEPTION JUSTE PAR AUDITEURS

La moyenne du taux de perception juste est de 61% avec une faible dispersion (minimum 58%, maximum 63%, écart type de 1,64). A noter que le score de l'absence de choix (par le fait de non appuyer) est faible avec une moyenne de 0,9% (minimum 0,1%, maximum 1,7%, écart type de 0,59 %). Globalement, comme en atteste le graphique n°5.3.2.1, il n'y a pas de déviance de comportement perceptif décelable au sein de ce groupe d'auditeur.

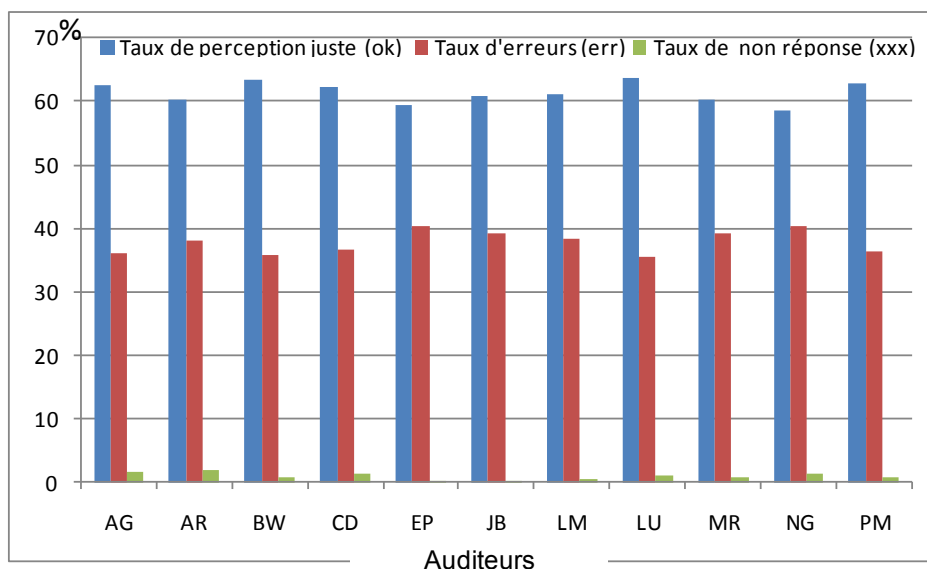


Figure n°5.3.2.1 : histogramme des réponses au test de perception par les auditeurs en pourcentage  
 Les compétences des auditeurs sont équivalentes et paraissent satisfaisantes pour la suite de l'expérience.

#### 5.3.2.1.1.2 LES LOCUTEURS TEMOINS

Le taux de bonne perception pour les locuteurs témoins est de 90% avec un taux de non réponse de 0,51% alors que le taux moyen de bonne réponse pour les patients est de 55% avec un taux de non réponse de 0,98% (figure n°5.3.2.2 ).

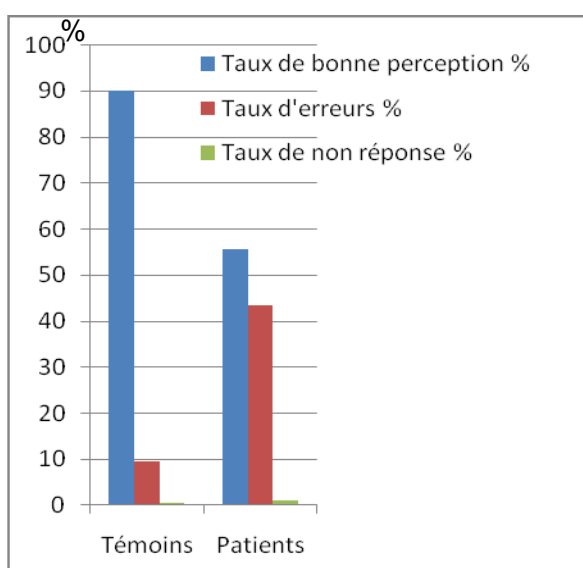


Figure n°5.3.2.2 : Description des réponses au test de perception pour les deux groupes de locuteurs (témoins et patients.

Cependant, il existe également une différence entre les deux témoins que nous allons explorer (figure n°5.3.2.3). Les consonnes produites par le locuteur PRM sont moins bien perçues que celles produites par le locuteur BAD, sauf pour les sons /k/,/b/,/ch/,/j/,/f/,/l/.

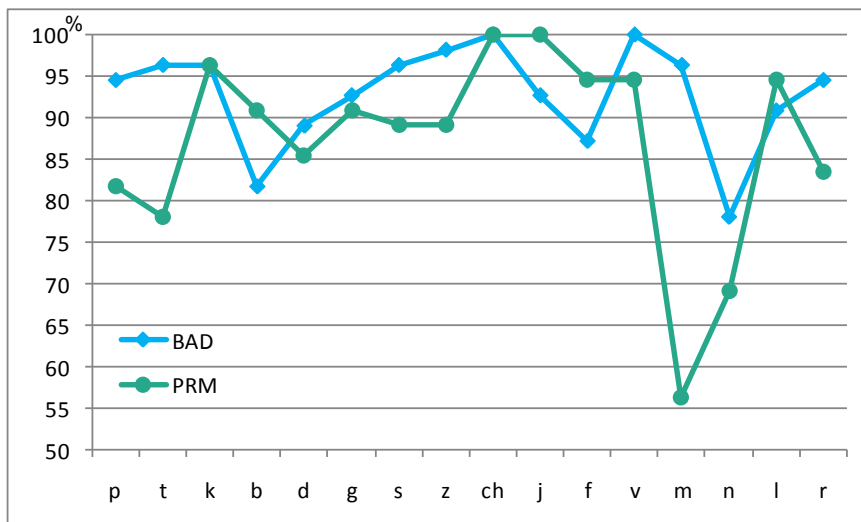


Figure n°5.3.2.3 : Pourcentage de perception juste par consonne pour les deux locuteurs témoins

Globalement, les nasales sont les moins correctement perçues, notamment le /n/ qui contrairement au /m/, a un taux inférieur à 80% pour les deux locuteurs (figure n°5.3.2.3).

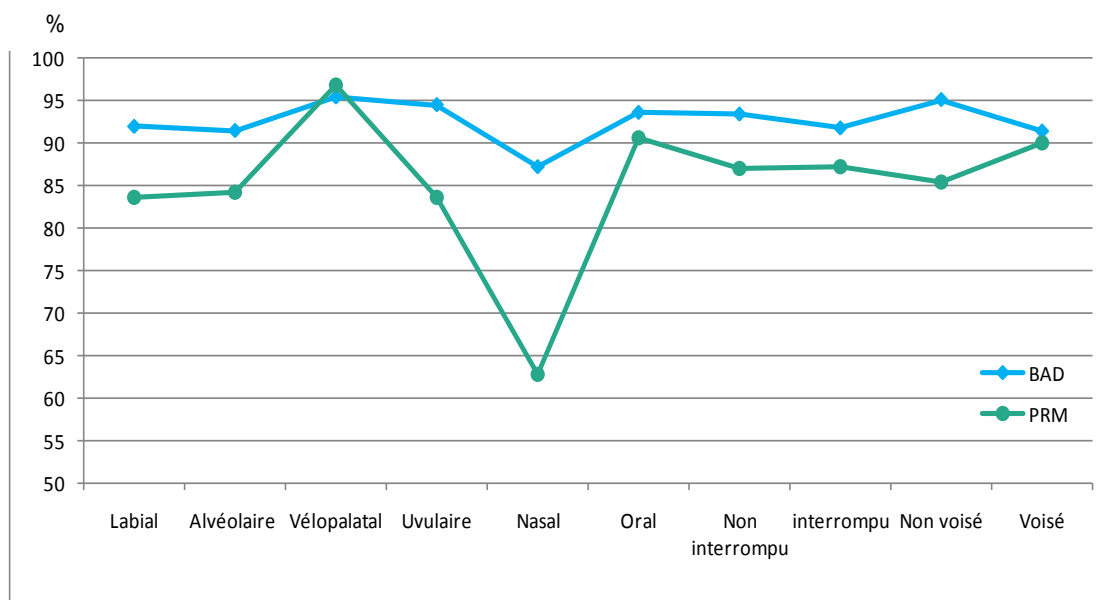


Figure n°5.3.2.4 : Taux de perception juste par traits pour les locuteurs témoins

Ceci est confirmé par l'analyse par traits (figure n°5.3.2.4). La différence est particulièrement nette sur le trait de nasalité. A noter que le trait vélopalatal est le lieu mieux perçu et que les non voisés sont mieux perçus que les voisés.

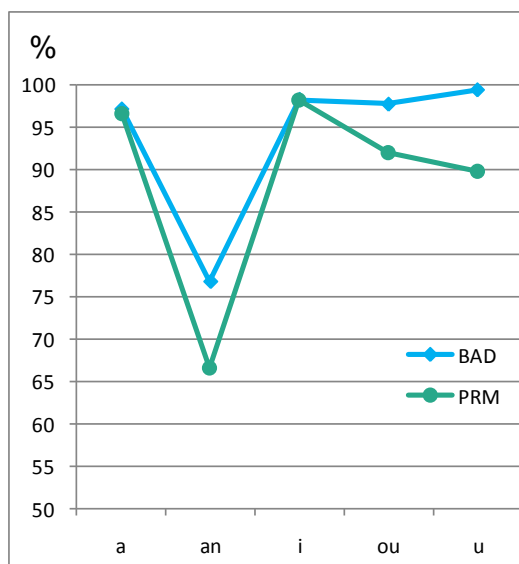


Figure n°5.3.2.5 : Taux de bonne perception des voyelles pour les deux locuteurs témoins

A noter que la mauvaise perception du trait de nasalité est également confirmée par l'analyse du taux de perception des voyelles (Figure n°5.3.2.5). La nasale /an/ est nettement moins bien reconnue que les autres.

Des études similaires décrivant des sujets sains révèlent, qu'en général, les taux de reconnaissance sans exclusion des auditeurs ou élimination d'une partie du corpus sont supérieurs à 85%. (Phatak et al 2008). Cette analyse confirme que le taux de reconnaissance est inférieur à 80% pour les deux locuteurs pour les phonèmes /an/ et /n/. En revanche apparaît une différence qui semble être spécifique au locuteur PRM sur le /m/ puisque seulement 56% des /m/ sont correctement reconnus alors que 96% de ceux produit par le locuteur BAD le sont. Une autre discordance apparaît pour le phonème /t/ dont le taux de reconnaissance est à 78% pour le locuteur PRM (figure n°5.3.2.3). Il s'agit avant tout d'erreurs de lieu qui peuvent relever des particularités des locuteurs dont l'accent régional est nettement plus marqué chez PRM (tableau n°5.3.2.1) .

Phonèmes cible	BAD		PRM	
	Phonème perçu 1er	Phonème perçu 2nd	Phonème perçu 1er	Phonème perçu 2nd
an	on		on	o
n	m	l	l	d
m			b	
t			p	

Tableau n°5.3.2.1 : Analyse des confusions issues de l'observation de la matrice de confusion des locuteurs témoins annexée en 7.3.1.2 sous le numéro 5.3.2.1 bis.



Nous n'avons pas retrouvé d'expérience similaire pour le français. Pour l'anglais les résultats révèlent que les taux de perception des inter dentales plus spécifiques à cette langue sont faibles (Miller & Nicely 1955, Allen 2005). La situation est donc similaire en ce qui concerne la bonne perception du trait de nasalité des voyelles, spécificité du français. Ces résultats, dans la limite des études disponibles, sont en accord avec la littérature. Fonagy (1989) signale que les erreurs prédominent sur le lieu. Les nasales sont des voyelles moins stables. Elles présentent une tendance à la neutralisation des contrastes. La stabilité articulaire du trait aigu décrit parmi les indices invariants par Stevens & Blumstein (1978) peut expliquer la perception du /p/ à la place du /t/. Ceci ne met pas en jeu la validité de l'expérience.

#### 5.3.2.1.1.3 INFLUENCE DE LA FREQUENCE DU SEGMENT SONORE

La fréquence du segment sonore produit dans la langue française influe t-elle sur le taux de bonne reconnaissance des phonèmes ? Pour répondre à cette question, nous avons recueilli et intégré dans la base de donnée, pour chaque mot du corpus la fréquence de la syllabe correspondant au mot, la fréquence phonétique et orthographique du mot (section 5.2.2.4). L'étude de la réponse au test de perception en fonction de ces différentes fréquences est en faveur d'une influence sur le taux de perception plus marqué avec la fréquence de la syllabe. Ce résultat est représenté sur les graphes en boite de la figure n°5.3.2.6.

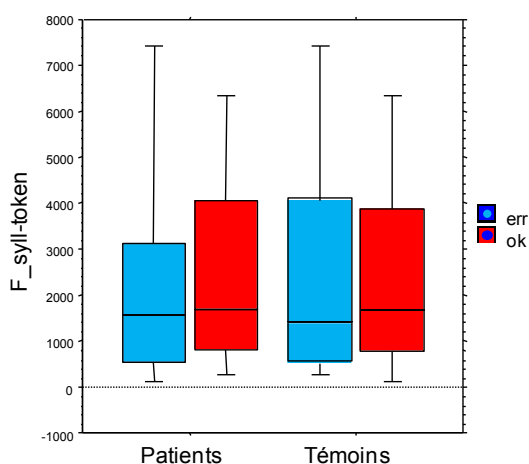


Figure n°5.3.2.6 : Fréquence des syllabes correspondant aux mots du corpus en fonction de la réponse au test de perception pour les patients et l témoins.

Cet effet n'est cependant pas suffisamment marqué pour justifier sa prise en compte systématique dans la suite de ce travail.

#### 5.3.2.1.1.4. A RETENIR AVANT D'ALLER PLUS LOIN...

Les résultats des auditeurs sont en faveur d'une bonne concordance /

Les locuteurs témoins dans ce corpus donnent lieu à une faiblesse pour la perception des nasales notamment le /n/

La fréquence des unités linguistiques influence le taux de réussite à l'épreuve perceptive. Mais cet impact est trop faible pour être prise en compte de manière systématique.

### 5.3.2.1.2 TAUX DE PERCEPTION POUR LE GROUPE PATIENTS

#### 5.3.2.1.2.1. RESULTAT PAR LOCUTEUR

Le Taux de perception juste des consonnes pour les patients est compris entre 29 et 85% (figure n°5.3.2.7).

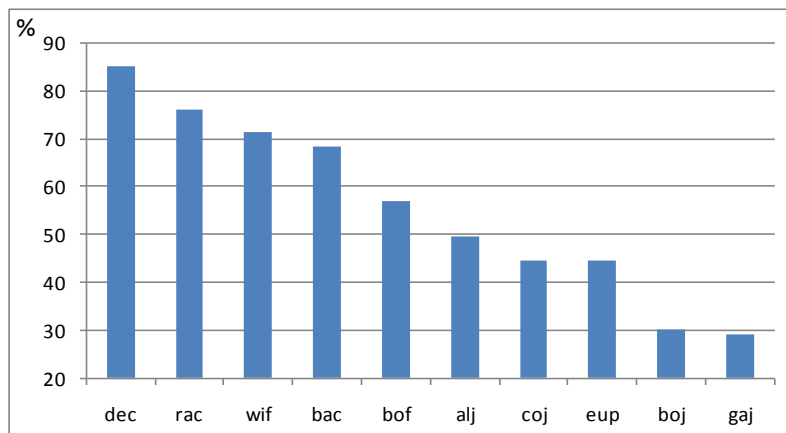


Figure n°5.3.2.7: Taux de perception juste des consonnes par ordre décroissant du corpus par patient  
Cette étendue concerne des patients de la même pathologie ce qui permet d'étudier le même type de distorsion à des degrés de sévérité différents.

#### 5.3.2.1.2.2. LES CARACTERISTIQUES DES PATIENTS INFLUENÇANT LE TAUX DE PERCEPTION DES CONSONNES

Interviennent les facteurs en rapport avec la tumeur : sa taille (plus elle est grande, plus le pourcentage de perception est faible), sa localisation (si le site initial est la langue en avant du V lingual, le pourcentage de perception plus faible). En revanche, le type de voie d'abord (incision et

dissection permettant l'accès à la tumeur) ou la résection ou non de la mandibule induit peu de différence (figure n°5.3.2.8).

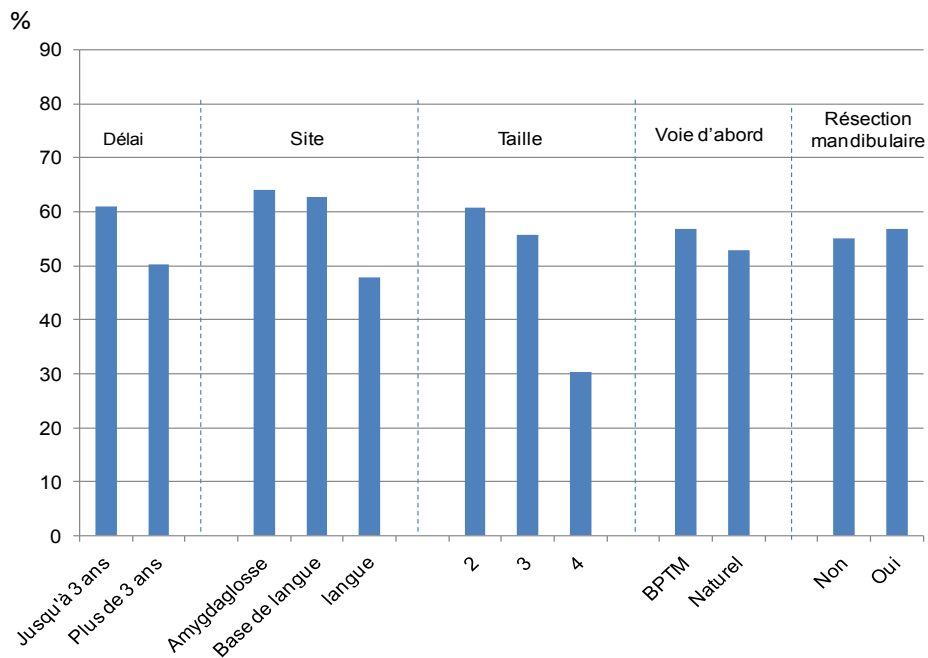


Figure n°5.3.2.8 : Impact des facteurs en rapport avec la tumeur sur le taux de perception juste.

Le délai par rapport à la fin du traitement Intervient également : plus le délai est grand, plus le taux de perception est faible avec une différence de l'ordre de 10% entre avant et après 3 ans. Ceci implique une dégradation avec le temps.

Les facteurs en rapport avec l'évaluation au moment de la réalisation de l'enregistrement sont rapportés sur la figure n°5.3.2.9. Il s'agit des résultats de l'analyse par rapport au degré subjectif d'altération de l'intelligibilité et à la cotation des possibilités dynamiques. Pour une lecture plus aisée des résultats, le degré subjectif d'altération de l'intelligibilité est rendu cohérent avec la cotation des possibilités dynamiques coté 1 normal à 4 très altéré.

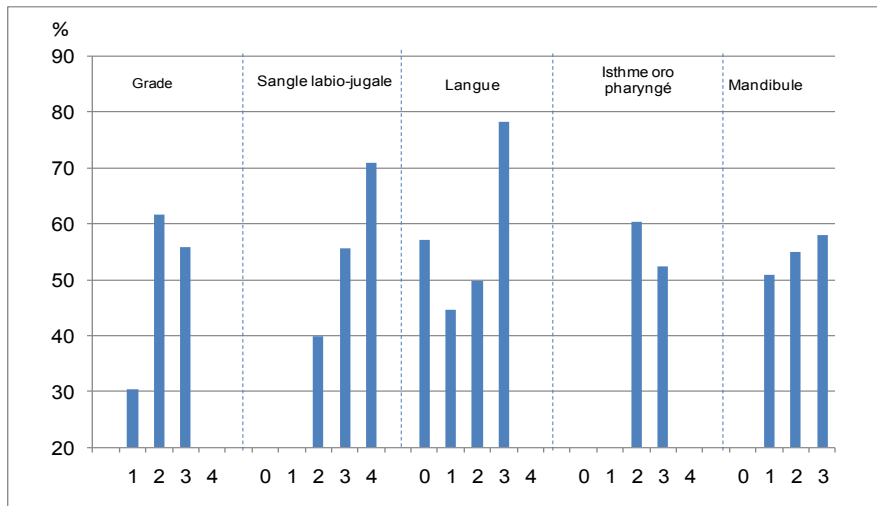


Figure n°5.3.2.9 : Taux de perception juste en fonction du degré de conservation des possibilités dynamiques

Une diminution du taux de perception est observée avec l'altération subjective de l'intelligibilité particulièrement marqué pour le grade 3 en sachant qu'aucun sujet n'avait été jugé comme ayant une intelligibilité subjectivement normale.

Il en est de même avec les altérations de la dynamique des structures anatomiques excepté l'isthme oropharyngé pour lequel il y a une tendance inverse mais seuls 2 niveaux d'altération sont présents ne permettant probablement pas une discrimination suffisante (figure n°5.3.2.10). A noter que le taux de perception le plus bas est associé à l'atteinte de la sangle labio-jugale la plus sévère.

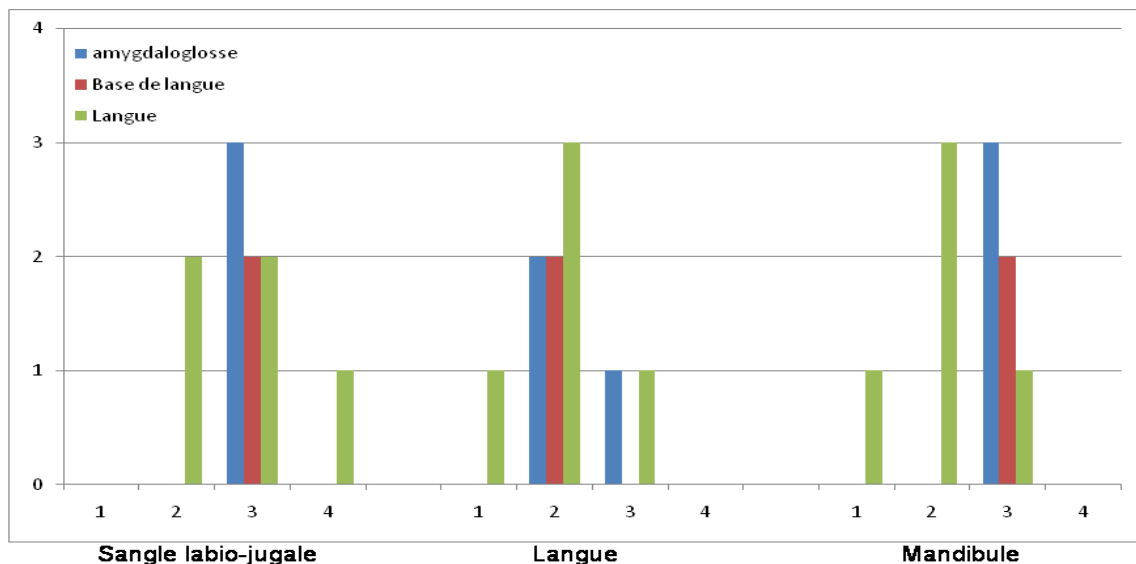


Figure n°5.3.2.10 : Degré d'altération de la dynamique des structures labio-jugale, linguale et mandibulaire en fonction du site tumoral.

Un déficit de la sangle labio-jugale est présent pour tous les sites tumoraux alors que le déficit de la dynamique de la langue mobile est relativement plus marqué pour le site basilingual et que le déficit d'ouverture mandibulaire est relativement moins marqué pour le site lingual. Au vu des ces résultats,

nous avons choisi de retenir le site tumoral initial comme un des paramètres à contrôler éventuellement dans les analyses futures.

#### 5.3.2.1.2.3. A RETENIR

1. L'étendue du taux de perception est compatible avec l'étude des distorsions présentes à différents degrés de sévérité.
2. Le site tumoral initial est le facteur lié au locuteur qui pourra être pris en compte dans l'analyse.

#### 5.3.2.1.3 TAUX DE PERCEPTION PAR CONSONNE ET PAR TRAIT POUR LES PATIENTS

Cette section est consacrée à l'analyse des résultats du taux de perception chez les patients.

##### 5.3.2.1.3.1. TAUX DE PERCEPTION PAR CONSONNE

Les taux de perception juste pour les patients sont comparés aux taux de perception pour les témoins sur le tableau n°5.3.2.4.

	% perception juste patients	% perception juste témoins	Différence de %
k	27,27	96,36	69,09
g	25,27	91,82	66,55
j	39,82	96,36	56,55
ch	46	100	54
d	47,82	87,27	39,45
t	50	87,27	37,27
b	52,18	86,36	34,18
z	61,82	93,64	31,82
l	61,82	92,73	30,91
s	62	92,73	30,73
p	58,18	88,18	30
v	72,73	97,27	24,55
f	71,27	90,91	19,64
m	67,27	76,36	9,09
r	80,91	89,09	8,18
n	65,64	73,64	8

Tableau n°5.3.2.4. Taux de perception par consonne pour les patients et les témoins classés par ordre décroissant de la différence de pourcentage.

Pour les patients, les occlusives dorsovélares /g/ et /k/ sont les moins bien perçues suivies des fricatives palatales post alvéolaires /j/ et /ch/. Les nasales /m/, /n/ et l'uvulaire /r/ sont perçues presque au même niveau que les témoins (< à10% de différence) (figure n° n°5.3.2.11).

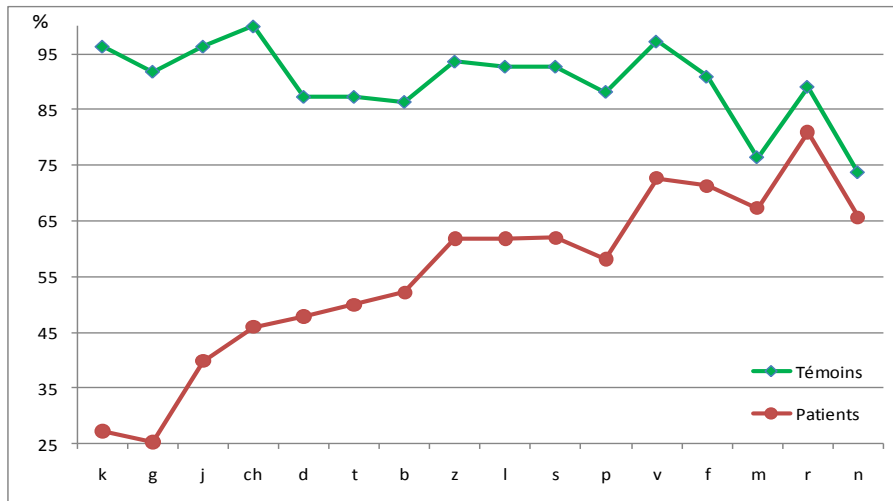


Figure n°5.3.2.11 : Représentation graphique des taux de bonne perception pour les patients et les témoins classés par ordre décroissant de la différence entre ces deux pourcentages.

L'analyse par locuteur (figure n°5.3.2.12) révèle quelques particularités. Par exemple les sujets BOJ et GAJ qui présentent un taux de perception particulièrement faible pour le // ou COJ pour le /r/ ou DEC qui présente dans un contexte globalement moins sévère, un taux bien conservé pour le /k/.

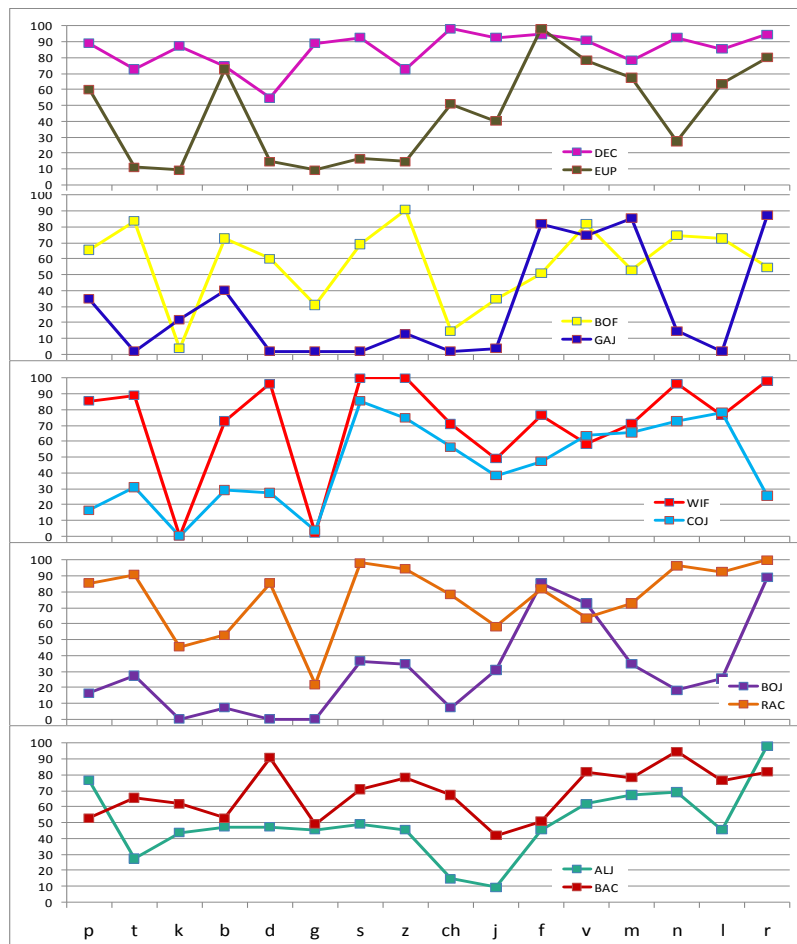


Figure n°5.3.2.12 : Taux de perception juste par consonne et locuteur

Les facteurs déterminants cette diversité sont liés aux profils des patients et seront étudiés à travers le site tumoral initial au vu des résultats décrits dans la section 5.3.2.1.

#### 5.3.2.1.3.2. TAUX DE PERCEPTION PAR TRAITS

Le trait le moins bien perçu est le lieu vélopalatal (34,5%) suivi du trait interrompu (43,4%). Le mieux perçu est le trait de lieu uvulaire (80,9%) (figure n°5.3.2.13)

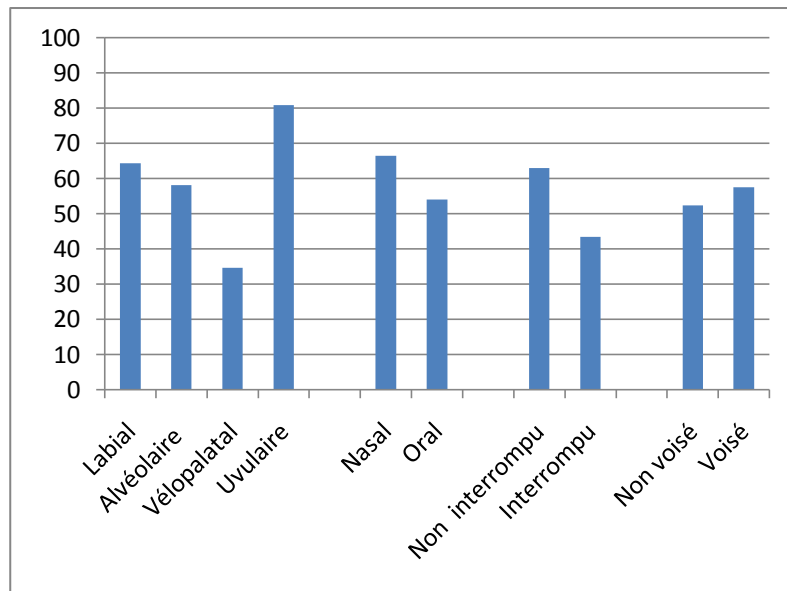


Figure n°5.3.2.13 : Taux de perception juste par trait

L'analyse en fonction du site tumoral montre que le faible taux de perception du lieu vélopalatal n'est pas lié au site tumoral (figure n° 5.3.2.14). Mais le site tumoral initial lingual est associé à un taux de perception plus faible du trait alvéolaire, du trait oral, du trait interrompu (figure n°5.3.2.15 à n°5.3.2.17). Le trait de voisement donne lieu à un taux de perception plus faible pour le site basi lingual traduisant l'impact possible d'une résection de cette région anatomique sur les performances laryngées.

Les différences observées en fonction des traits s'expliquent par la difficulté production propre au traitement réalisé en fonction du site et de taille de la tumeur. Ainsi l'absence de résection labiale même si la lèvre inférieure a été sectionnée chez au moins 6 patients explique la relative préservation de ce lieu. En ce qui concerne le voile du palais, une résection importante de l'hémi voile a été réalisée une seule fois (BOJ) ce qui peut expliquer le meilleur score du lieu uvulaire (figure n°5.3.2.23).

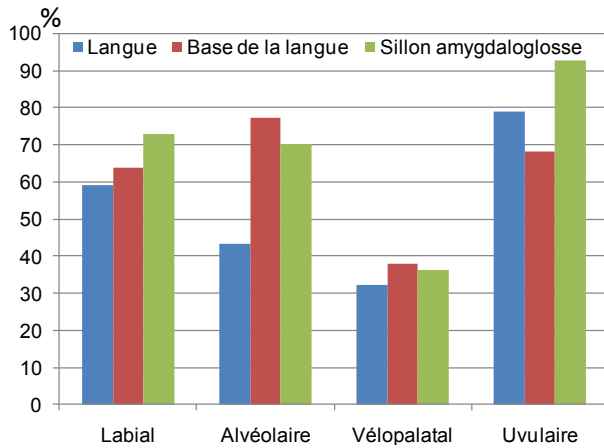


Figure n° 5.3.2.14 : Pourcentage de perception juste du lieu en fonction du site tumoral

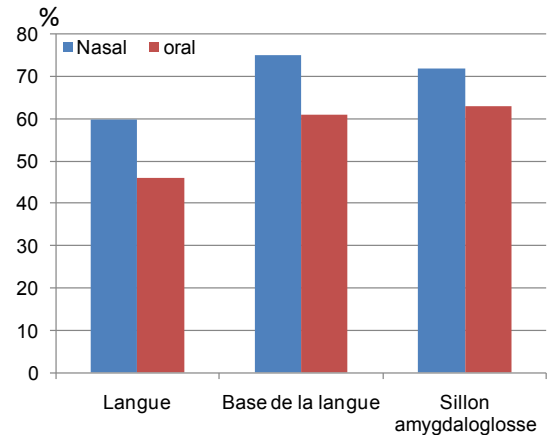


Figure n° 5.3.2.15 : Pourcentage de perception juste de la nasalité en fonction du site

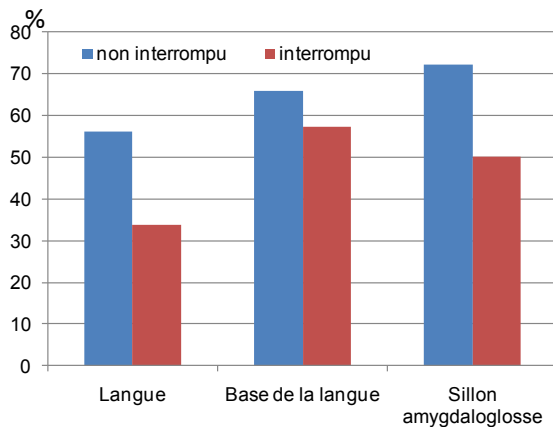


Figure n° 5.3.2.16 : Pourcentage de perception juste du trait interrompu en fonction du site

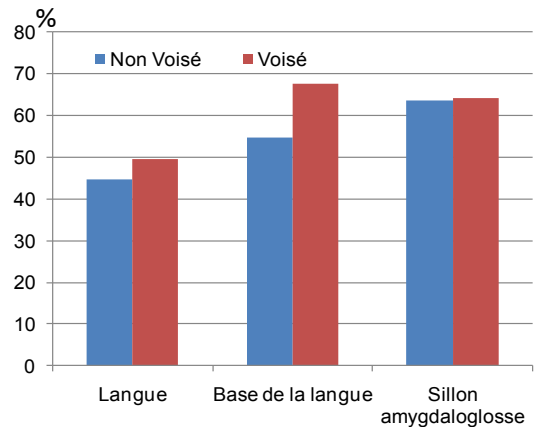


Figure n° 5.3.2.17 : Pourcentage de perception juste du caractère voisé en fonction du site

### 5.3.2.1.3.3. A RETENIR

1. Le taux de perception est particulièrement faible pour les consonnes /k/, /g/, /j/, /ch/ et le trait vélopalatal. Ce phénomène est probablement en rapport avec la faible fréquence des syllabes comprenant ces consonnes dans la langue française notamment pour /g/, /ch/ et /j/ (voir section 5.2.4), mais est probablement lié aussi au profil des distorsions liés à la pathologie pouvant expliquer le faible taux de perception du [k] avec un effet du site tumoral lingual net sur le lieu alvéolaire.
2. Le taux de perception reste satisfaisant pour consonnes /r/, /m/, /n/ et traits uvulaire et nasal mais également labial puis non interrompu. L'effet fréquence de la syllabe n'est pas évident ici.



## 5.3.2.2 ETUDE DES CONFUSIONS

### 5.3.2.2.1. MATRICE DE CONFUSION

Les confusions sont présentées tout d'abord sur une matrice avec les consonnes cibles en colonne et les consonnes perçues en ligne. Sont colorées en orange les confusions dont l'occurrence est supérieure à 11 (c'est-à-dire supérieur au nombre d'auditeurs), dans la limite des 3 premières par consonne (tableau n°5.3.2.6). Est remarquable l'incidence que les confusions ont sur l'occurrence de certaines consonnes en terme de perception, notamment les labio-dentales /v/ et /f/.

Cibles	perçues																	Total	
	b	ch	d	f	g	j	k	l	m	n	p	r	s	t	v	z	xxx		
b	382		37	1	45	5		26	9	11	10	1		1	122	8	2	660	
ch		363		172		9							110	3				3	660
d	13		359		35	13	1	45	9	5	2	1	1	9	108	53	6	660	
f	9	36		494		2	4				11		70	4	19	4	7	660	
g	16	1	63	5	240	43	1	33	10	15	1	81	1	1	120	21	8	660	
j		5	4	3	37	326		10			2	15	3	2	138	109	6	660	
k	1	38	3	123	7	10	256	6	2	1	8	72	26	64	21	8	14	660	
l	21	1	8		15	11		442	28	18	17	18			53	14	14	660	
m	25							31	454	70		5			68	3	4	660	
n	4	1	6		5	5		66	88	444					19	11	11	660	
p	18	6	4	92	3		50	4	2	1	417	2	13	26	13	5	4	660	
r	18			6	2	13	3	10	2		5	543			44	7	7	660	
s		46	1	143			1				1		444		12	11	1	660	
t		29	17	60	4	12	69	1			19	2	31	373	29	13	1	660	
v	9		11	2	6	41		14	1		5	19			507	42	3	660	
z	2		3	1	9	42		5				3	4		144	443	4	660	
<b>Total</b>	518	526	516	1102	408	532	385	693	605	565	498	762	703	483	1417	752	95	10560	

Tableau n°5.3.2.6 : Matrice de confusion des consonnes cibles en colonne, perçues en ligne

### 5.3.2.2.2. CALCUL D'UN TAUX DE CONFUSION

Le calcul du taux de confusion est réalisé à partir de la matrice (Tableau n°5.3.2.6). Le nombre de confusion rapportée au nombre total d'occurrence de la consonne perçue permet le calcul du taux de confusion (tableau n°5.3.2.10). soit :

$$\text{Taux de confusion} = \frac{nbX - nbXok}{nbrepX}$$

Xok = le nombre de réponse juste pour X

nbX = le nombre de X présentés comme cible

nbRepX = le nombre de réponses X fourni par les auditeurs.

#### 5.3.2.2.2.1. PAR CONSONNES

Pour les consonnes, le taux de confusion est supérieur à 50% pour /v/ 69%, /f/ 60%, /g/ 52%. Il est par contre faible pour le /p/ (17%) puis /n/ 23%, /b/ 28%, /t/ et /m/ 27% (figure n°5.3.2.18).

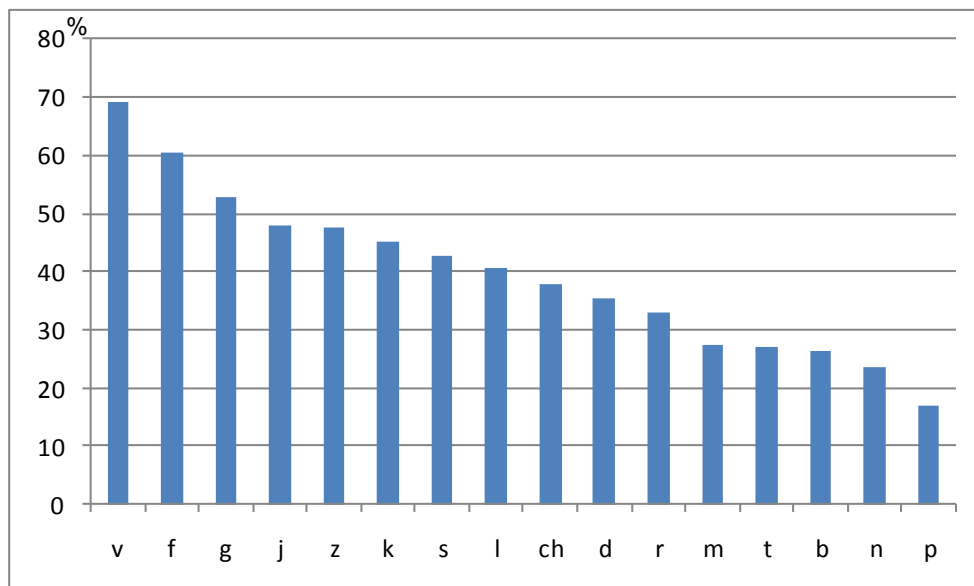


Figure n°5.3.2.18 : Taux de confusion des consonnes

#### 5.3.2.2.2. PAR TRAITS

Le taux de confusion prédomine sur le trait de lieu puis le trait non interrompu. Il est particulièrement élevé pour le trait labial. Les traits conservés sont le trait de voisement puis le trait de nasalité (figure n°5.3.2.19).

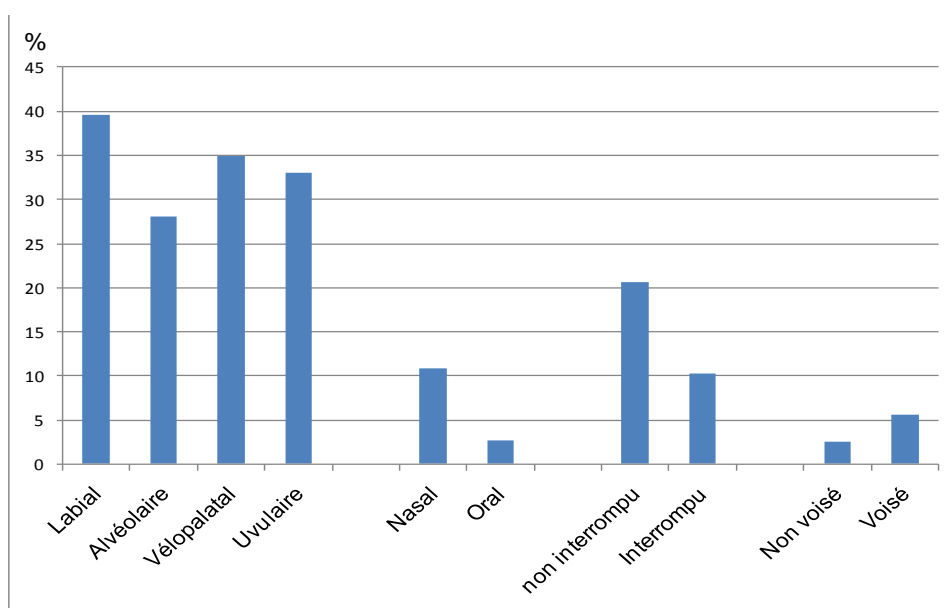


Figure n°5.3.2.19 : Taux de confusion par traits

Le taux de confusion a ensuite été étudié en fonction du site. Les résultats sont présentés sur les figures n° 5.3.2.20 au n° 5.3.2.23

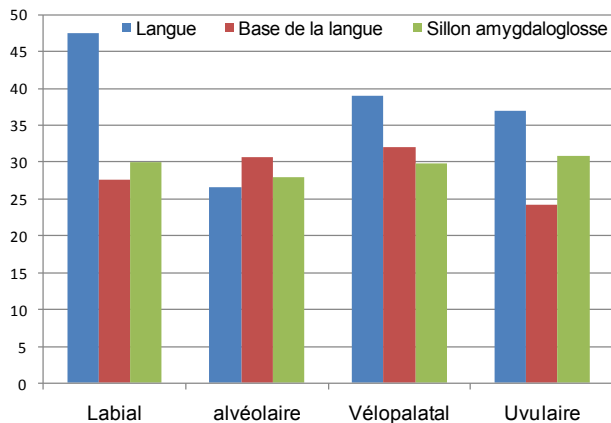


Figure n° 5.3.2.20 : Taux de confusion pour le trait interrompu

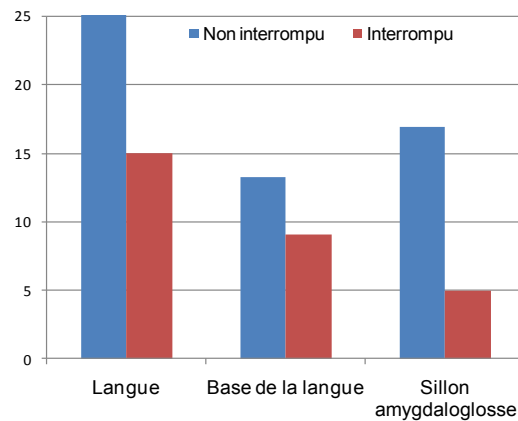


Figure n° 5.3.2.21 : Pourcentage de confusion le trait de lieu

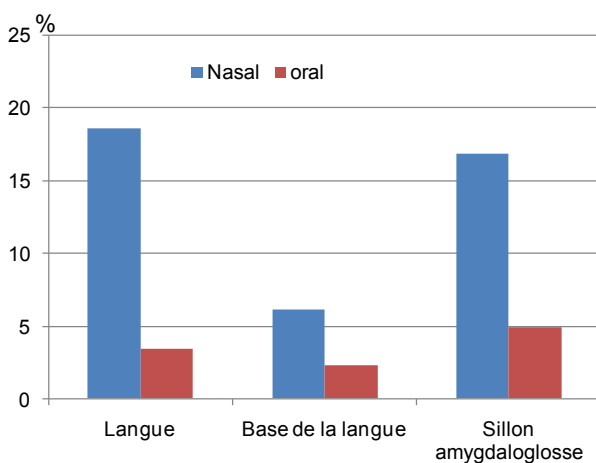


Figure n° 5.3.2.22 : Pourcentage de confusion pour le trait de voisement

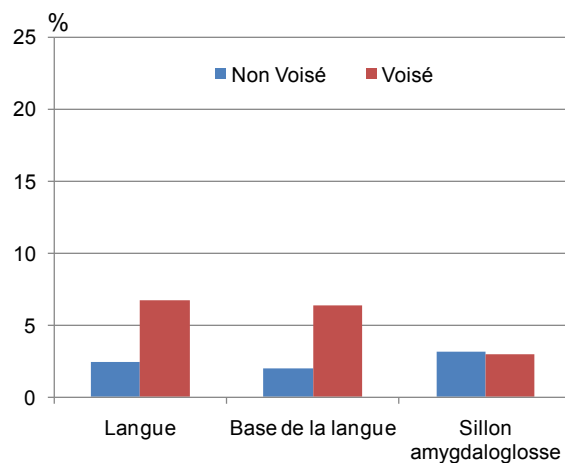


Figure n° 5.3.2.23 : Pourcentage de confusion pour le trait de nasalité

Le site tumoral initial lingual est associé à un plus grand taux de confusion avec une nette prédominance sur le lieu labial. Il n'y a pas de différence pour le trait alvéolaire mais ceci est le trait de moins bonne perception pour ce site initial. A noter également, une tendance avec des confusions plus importantes sur le trait velopalatal et moins importante sur le trait uvulaire pour la base de la langue. Rien n'apparaît nettement pour le site du sillon amygdalogue. Les confusions sur le trait interrompu sont plus fréquente également pour le site de la langue. Il y a peu d'impact sur le trait nasal (un peu moins de confusion pour le site base de la langue) et aucun de décelable pour le trait de voisement.

### 5.3.2.2.3. DESCRIPTION DES CONFUSIONS

Une approche qualitative des confusions explicite comment ces confusions peuvent avoir un impact sur le processus d'accès au sens dès le processus de perception de la parole.

	bilabial		labiodental		Alveolaire		Palatal		Vélaire		Pharyngeal	Glottal
Plosive	P 58% 52% 50% 11%	b			T 50% 48% 11%	D 48% 12%			K 27% 25% 12%	G		
Nasale	M 67% 14%					N 66% 12%						
Fricative			F 31% 22% 16% 25%	V 73% 17% 22% 15%	S 62% 20% 12% 10% 12% 25%	Z 62% 20%	J 46% 40%					
Latéral fricative								R 80%				
Approximante								14% 13%				
Latéral approximante						L 61%						

Figure n°5.3.2.24 : Taux de perception juste (en vert) et pourcentage de confusion (en rose) des consonnes et leur relation.

La figure n°5.3.2.24 illustre la convergence de la perception des productions vers les fricatives labio-dentales avec respect en général du trait de voisement. A noter une tendance pour un pourcentage relativement plus faible à une confusion du lieu vélopalatal et uvulaire. Cette représentation fait évoquer le rôle important des confusions sur les difficultés de compréhension dans la mesure où il apparaît évident que le taux de bonne perception du [v] par exemple perd de sa pertinence du fait de la source majeur de confusion que ce son représente. De même, le trait nasal est bien perçu mais avec un important taux de confusion. Cette observation justifie le calcul d'un taux de confusion et une méthode de pondération du taux de bonne perception. Ces résultats sont confirmés par l'étude des confusions par traits illustrés ici par les tableaux n°5.3.2.8 et 5.3.2.9 concernant le lieu et sur le caractère interrompu ou non.

Lieu prévu	Lieu perçu				NA	Total
	L	A	VP	U		
L	2158	363	186	27	16	2750
A	742	2214	284	24	36	3300
VP	611	483	910	168	28	2200
U	65	17	17	445	6	550
<b>Total</b>	<b>3576</b>	<b>3077</b>	<b>1397</b>	<b>664</b>	<b>86</b>	<b>8800</b>

Figure n°5.3.2.8 : Matrice de confusion des traits de lieu (en vert, le nombre de correctement perçu, en orange du plus foncé au plus clair le nombre décroissant de confusions) NA = Non attribué

Interrompu prévu	Interrompu perçu		NA	Total
	Non	Oui		
Non	5226	218	56	5500
Oui	1360	1910	30	3300
<b>Total</b>	<b>6586</b>	<b>2128</b>	<b>86</b>	<b>8800</b>

Tableau n°5.3.2.9 : Matrice de confusion du trait interrompu (NA = Non attribué)

#### 5.3.2.2.4. A RETENIR

1. Comme pour la bonne perception, les confusions prédominent sur les traits de lieu et le trait interrompu
2. On observe un profil plus marqué pour le site lingual avec des confusions plus importantes pour le trait labial

#### 5.3.2.3 PONDERATION DU TAUX DE PERCEPTION PAR LE TAUX DE CONFUSION

La pondération du taux de bonne perception par la prise en compte des confusions paraît intéressante pour une mesure de l'intelligibilité discriminante à un niveau supérieur comme celui de la reconnaissance des mots ou de la compréhension d'énoncé.

L'idéal serait de calculer la probabilité de reconnaître un phonème à juste titre parmi l'ensemble des sons perçus de la même catégorie en y intégrant les paramètres participant à cette probabilité. Le problème peut être posé comme suit : On peut entendre  $n$  occurrences de la cible, il y en a  $x$  de juste et  $y$  de fausses, quelle est la probabilité de tomber sur  $x$  en sachant que la fréquence de la cible dans la langue française est  $f$  ?

##### 5.3.2.3.1. LA F MESURE

Nous n'avons pas trouvé de méthode spécifique pour répondre à cette question dans la littérature. Mais ce problème s'apparente à celui de l'efficacité d'une technique de recherche d'informations. Pour cela, deux mesures distinctes sont utilisées :

- Le rappel donne le pourcentage des réponses correctes parmi l'ensemble des propositions disponibles dans une base de donnée. Elle correspond au taux de perception juste.
- La précision donne le pourcentage de réponses correctes parmi les propositions retrouvées dans la base par la technique évaluée. Cette notion est complémentaire de la confusion telle que nous l'avons définie. Elle donnerait le pourcentage de réponses incorrectes parmi les propositions retrouvées.

Afin d'évaluer un système, l'interprétation entre le rappel et la précision se fait souvent avec un graphique du rappel par rapport à la précision (ou vice versa) ou par le calcul de la F mesure en maximisant la moyenne harmonique de la précision et du rappel :

$$F=2*(rappel*précision)/(rappel+précision).$$

Appliqué à nos données, le développement de la formule générale de la F mesure est la suivante :

$$F=Xok/[(nbX+nbRepX)/2] \text{ où}$$

Xok = le nombre de réponse juste pour X

nbX = le nombre de X présentés comme cible

nbRepX = le nombre de réponses X fourni par les auditeurs.

Présenté par ordre décroissant avec le coefficient de perception pondérée (figure n°5.3.2.2.25), la F mesure est plus faible (inférieure à 0,5) pour /g/ 0,32, /k/ 0,36 /v/ 0,43. Elle est supérieure à 0,7 pour /r/ 0,73, /n/ 0,70, /m/ 0,69, /p/ 0,68.

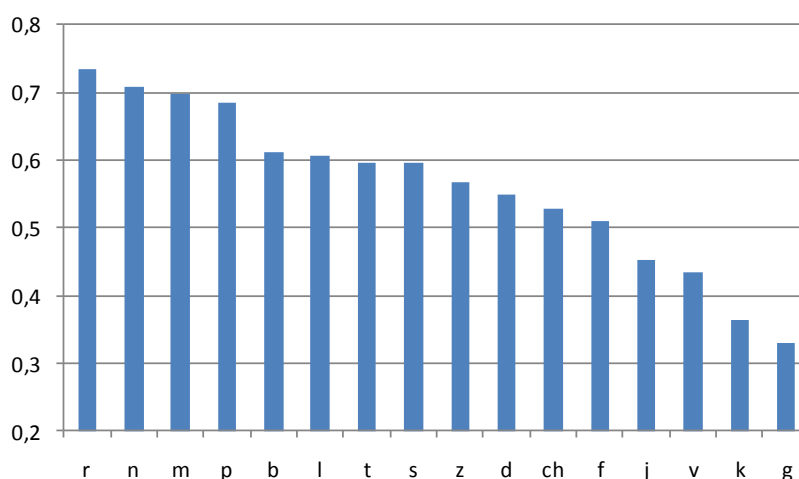


Figure n°5.3.2.2.25 : F mesure pour les consonnes, listées par ordre décroissant.

Pour les traits, le lieu vélopalatal (0,5) présente le score le plus faible. Les autres traits ont un score supérieur à 0,7 (figure n°5.3.2.2.26). Le score le plus élevé concerne le trait oral (0,97).

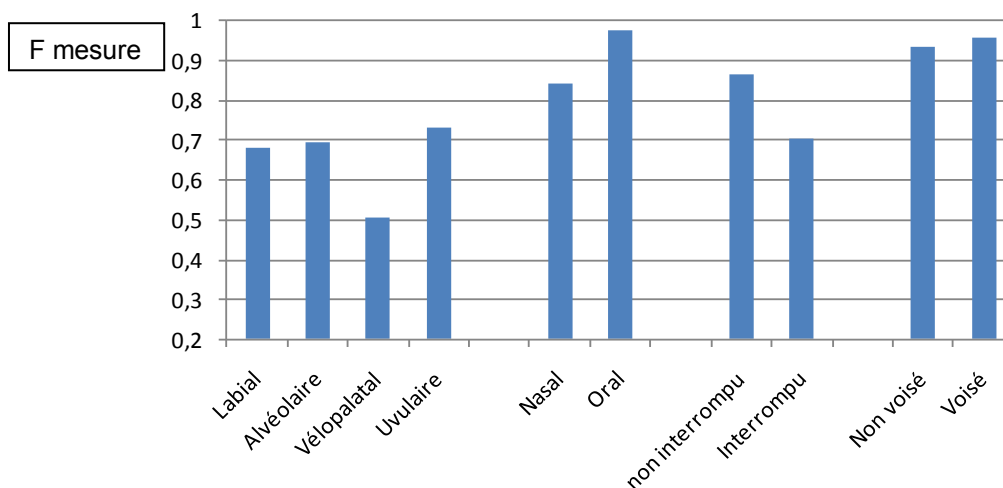


Figure n°5.3.2.2.36 : F mesure calculée pour les traits

### 5.3.2.3.2. A RETENIR PAR RAPPORT AUX CONFUSIONS

1. Les confusions sont plus fréquentes sur le trait de lieu et le trait interrompu. Elles induisent une surreprésentation perceptive des fricatives notamment labio-dentales
2. De pondérer les résultats du taux de perception par les confusions ne change rien pour les taux de perception bas.
3. En revanche la pondération pour les taux élevés de perception juste fait glisser les labiodentales dans les mauvaises performances. Le son /p/ glisse dans le groupe le mieux perçu juste avec les nasales et le /r/.
4. Les nasales et le /r/ étaient déjà bien positionnés notamment dans le classement issu du calcul de la différence par consonne du taux de bonne perception entre patients et témoins (tableau n°5.3.2.2.14)

F mesure	% de perception juste	Différence de % patient/témoins
g	g	k
v	k	g
k	j	j
j	ch	ch
f	d	d
d	t	t
ch	b	b
z	p	z
b	z	l
s	l	s
t	s	p
l	n	v
m	m	f
p	f	m
n	v	r
r	r	n

Tableau n°5.3.2.2.14 : Les listes de consonnes obtenues par la F mesure, le taux de perception juste et la différence entre le taux de perception des témoins et des patients

Quelque soit la méthode utilisée, la consonne uvulaire /r/ et les nasales /n/,/m/ sont très bien perçues. La pondération fait basculer la position des labiodentales /f/,/v/ et celle du /p/.

#### 5.3.2.4. APPROCHE DE LA ROBUSTESSE PAR LES AUDITEURS

Afin de pouvoir comparer les différents aspects de la robustesse par les différentes méthodes utilisées, nous allons d'abord utiliser les résultats du test de perception par une approche classique : la part de la variance du taux d'intelligibilité, ici le taux de perception juste pour chacune des unités et une approche un peu différente : l'impact de la sévérité donné soit par le score d'impression globale lors du bilan soit par taux de perception juste par locuteur

##### 5.3.2.4.1. ETUDE PAR RAPPORT A L'IMPRESSON GLOBALE DE SEVERITE

Pour cela le seul patient ayant un grade très sévère a été ajouté au grade 2, permettant de comparer la courbe du taux de perception juste en fonction de ces 2 niveaux de sévérité par ordre croissant de différence entre les 2 grades (figure n°5.3.2.2.27).

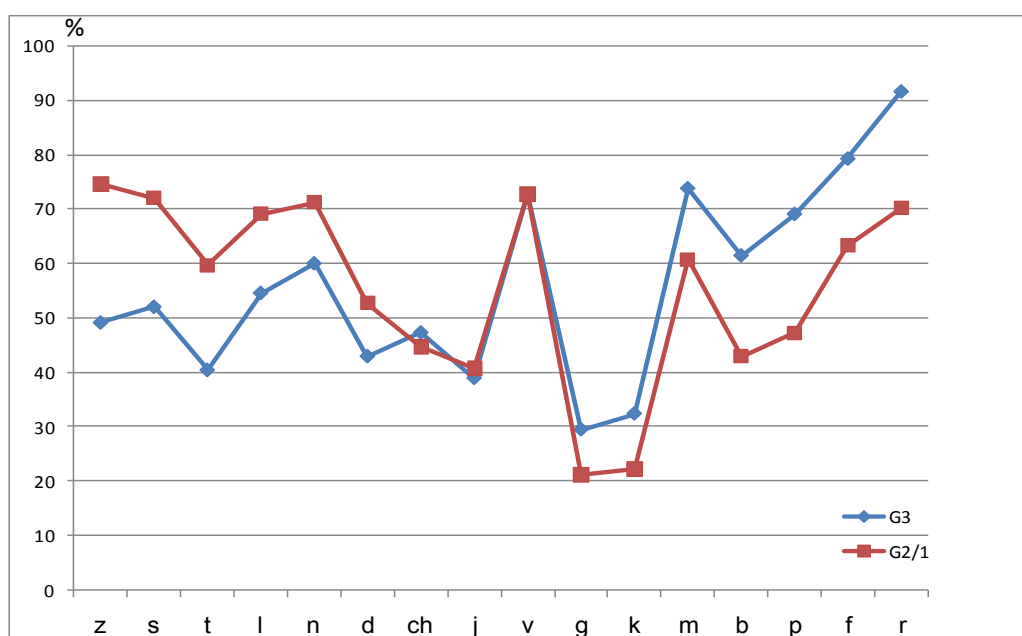


Figure n°5.3.2.2.27. Comparaison des courbes de perception juste en fonction de ces 2 niveaux de sévérité par ordre croissant de différence entre les 2 grades.

Cette représentation montre qu'en considérant le jugement de sévérité des experts sous la forme de l'impression globale de la parole, les consonnes qui ne changent pas malgré une impression d'altération sont le /v/, /j/ et le /ch/. Certaines Consonnes sont mieux perçues malgré l'altération globale : z, s,t,d,n,l (robuste à la distorsion dans ce contexte pathologique ou compensation par une



modalité de production qui permet une identification ?) et d'autres sont moins bien perçues : k,g, m,p,b,f,r (dégradation proportionnelle ou fragilité ?).

#### 5.3.2.4.1. ETUDE PAR RAPPORT AUX TAUX DE PERCEPTION PAR LOCUTEUR

La figure n° n°5.3.2.2.28 représente la distribution du nombre de réponse juste en fonction du taux de perception juste par locuteur donnant une vision de l'impact d'un indice de sévérité globale

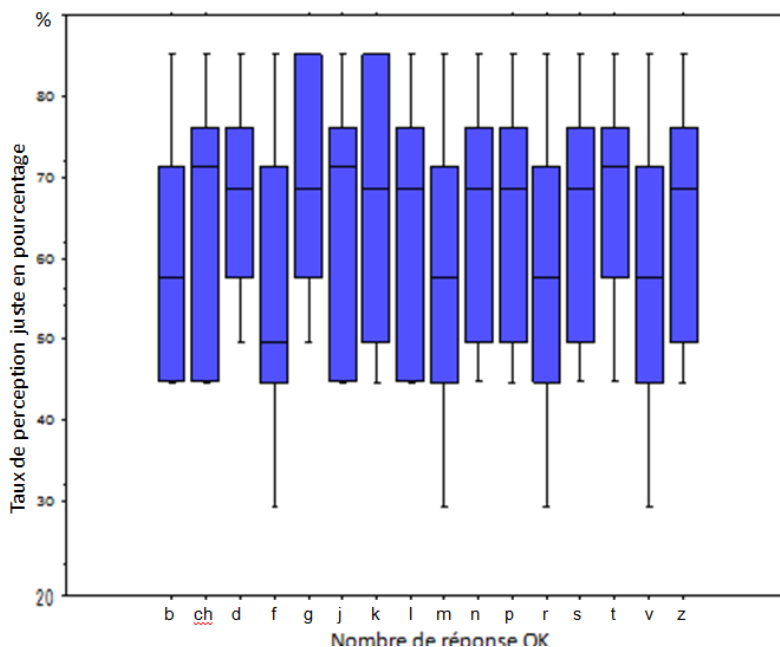


Figure n°5.3.2.2.28 : distribution du nombre de réponse juste en fonction du taux de perception juste par locuteur

Avec cette représentation, le /f/ est la consonne parait la plus robuste (la médiane du nombre de Ok correspond à un taux de perception juste de 50%, c'est-à-dire que ce taux, il reste la moitié de ok en dessous) puis ce sont les consonnes v, r,m,b. En revanche, il est plus difficile de déterminer les plus fragiles : /t/, /ch/, /j/,/d/.

#### 5.3.2.5. A RETENIR

1. Quelle que soit l'approche utilisée, les sons /g/,/k/,/j/ et le trait vélopalatal restent les moins bien perçus. Cette observation résulte probablement du profil de distorsions produit par la sélection de notre population de patients mais également de la fréquence des syllabes contenant ces consonnes dans la langue française pour l'échantillon de notre corpus.

2. Si le taux de perception est toujours bon pour /r/, /n/,/m/, la place de certains sons changes si l'on prend en considération une pondération par le taux de confusion. Ainsi, la perception juste du son /p/ serait sous-estimée alors que celle des sons /v/ et /f/ serait surestimée.

3. Le site tumoral initial lingual est associé à une diminution de la perception du trait alvéolaire et une augmentation de la confusion pour le trait labial. Cela est moins net ou moins systématique pour les autres sites.

4. Un calcul de probabilité à partir du taux de perception juste, du taux de confusion, de la fréquence de la syllabe et pour notre population éventuellement du site tumoral initial pourrait avoir un intérêt pour prévoir l'impact sur la compréhensibilité (forme de prédiction)

---

### 5.3.3. RESULTATS DU SYSTEME AUTOMATIQUE DE RECONNAISSANCE VOCALE

---

Nous allons maintenant analyser les résultats du système automatique de reconnaissance vocale pour estimer le degré de dégradation de la production.

#### 5.3.3.1. QUELQUES VERIFICATIONS PREALABLES

---

##### 5.3.3.1.1. ETUDE DES RESULTATS DES LOCUTEURS CONTROLES (AU TOTAL 10 LOCUTEURS, 5 FEMMES ET 5 HOMMES)

##### 5.3.3.1.1.2. QUE NOUS APPREND L'ANALYSE DES RESULTATS OBTENUS AVEC LES SUJETS CONTROLES SUR LA RECONNAISSANCE DES SYLLABES DU CORPUS ?

Le taux de reconnaissance du mot cible utilisé ici est déterminé par l'occurrence de la cible parmi les trois premiers candidats proposés par le système. Il est de 77% en moyenne et varie de 62,5 à 88,75% par locuteur.

Sont particulièrement mal reconnus : gi (1/10), vou (2/10) puis gou, gu, kou (3/10), bu, vu , zou, zu, mou (4/10). L'association à la voyelle « u » et à moindre degré « ou » réduit le taux de reconnaissance des syllabes (figure n°5.3.3.1).

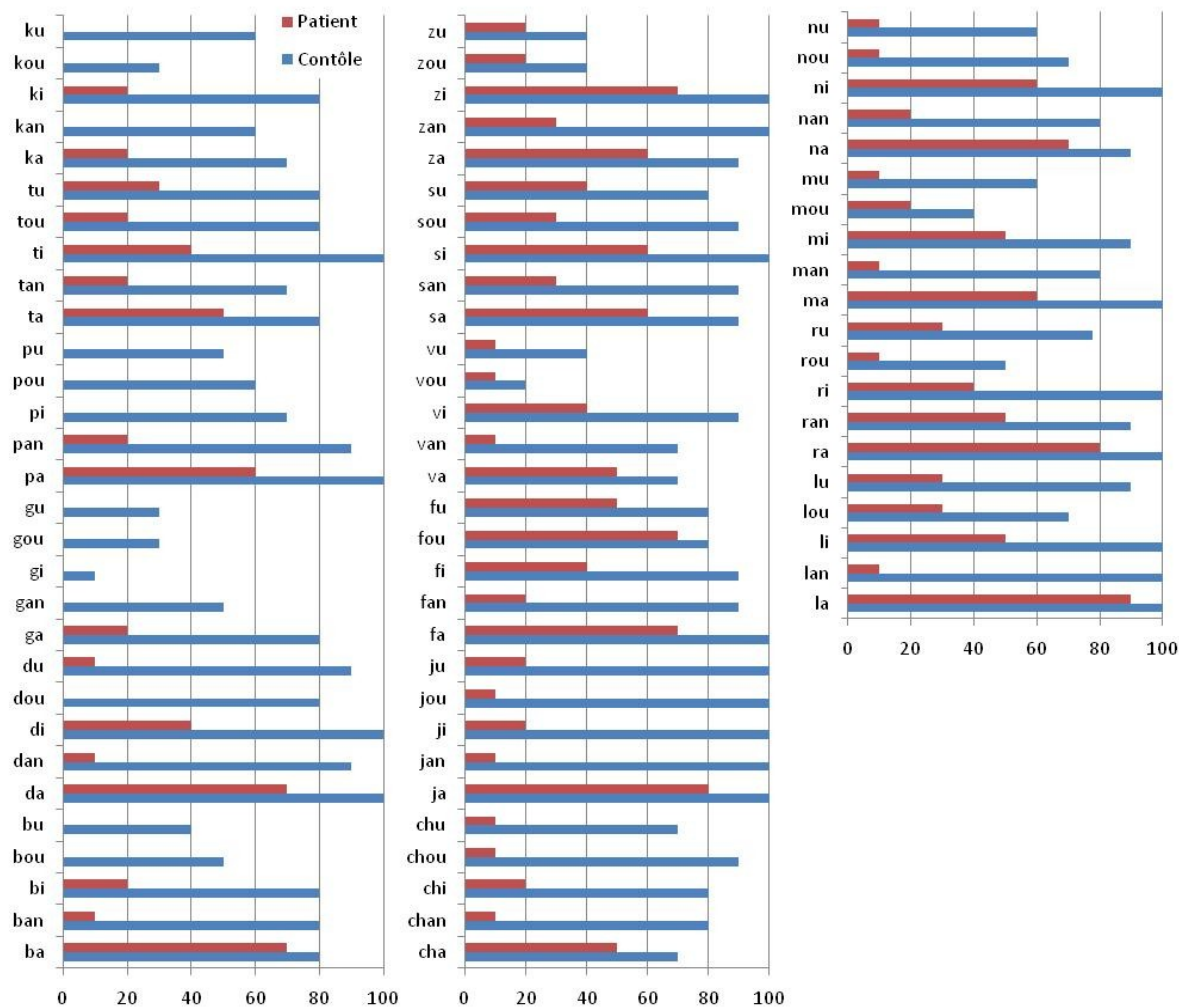


Figure n°5.3.3.1. Taux de reconnaissance du mot cible (déterminé par l'occurrence de la cible parmi les trois premiers candidats proposés par le système) pour les contrôles et les patients.

Pour expliquer le taux relativement faible de reconnaissance peut être évoquée la nature de l'expérience elle-même avec l'absence d'utilisation d'une modélisation linguistique venant compléter la modélisation acoustique et l'approche statistique.

Cependant ce taux de reconnaissance paraît suffisant pour que ce système de reconnaissance soit utilisé pour mener à bien cette expérience.

#### 5.3.3.1.1.2. VERIFICATION DE L'ALIGNEMENT

Cette vérification s'est limitée à la question suivante : Est-ce que les syllabes précédant ou succédant à la cible ont une fréquence anormalement élevée en tant que 1<sup>er</sup> candidat (figure n°5.3.3.2) ?

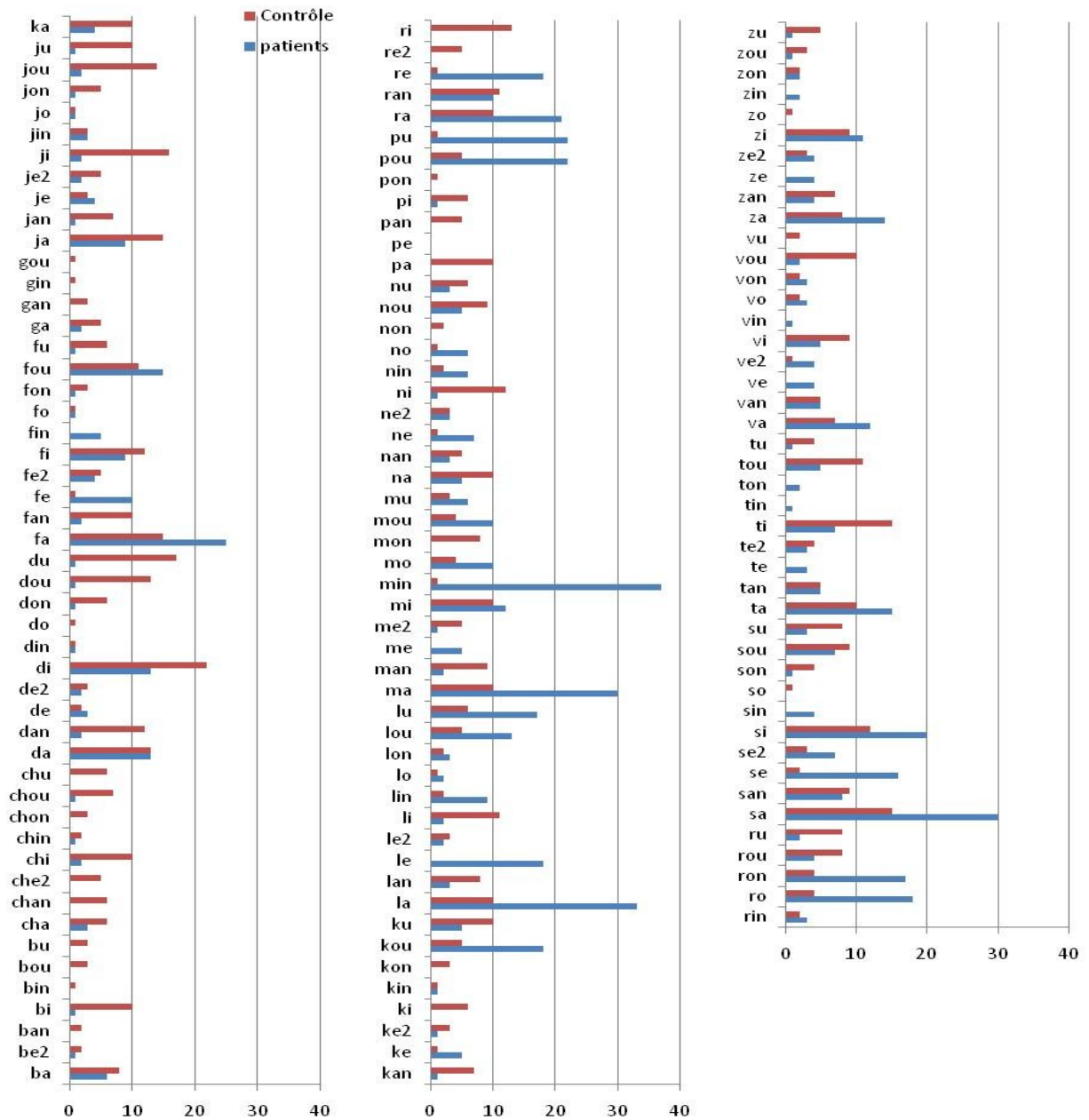


Figure n°5.3.3.2. Nombre d'occurrence des syllabes proposées comme 1<sup>er</sup> candidat par le système.

Pour les sujets contrôle, la syllabe correspondant au mot « mot » n'a été proposée que 4 fois et est en 6<sup>ème</sup> position des 1ers candidats qui ne correspondent pas à une cible du corpus. La syllabe « kon » présente dans le mot « convient » est encore moins fréquent (3 fois). Pour les patients, « mo » est sorti 12 fois comme premier candidat et est en 7<sup>ème</sup> position des 1ers candidats qui ne correspondent pas à une cible. « kon » n'a jamais été proposé par le système. Ceci est en faveur de l'absence de problème d'alignement spécifique à ce corpus.

### 5.3.3.1.2. RESULTATS DES PATIENTS ET DES DEUX TEMOINS DONT LE CORPUS A ETE UTILISE LORS DU TEST DE PERCEPTION

#### 5.3.3.2.2.1. POUR LES SYLLABES

Le taux de bonne reconnaissance par le système des syllabes est en moyenne de 15,5% pour les patients si l'on considère que le premier candidat et de 28,75% si l'on considère les 3 premiers candidats (3,75 à 56,25%).

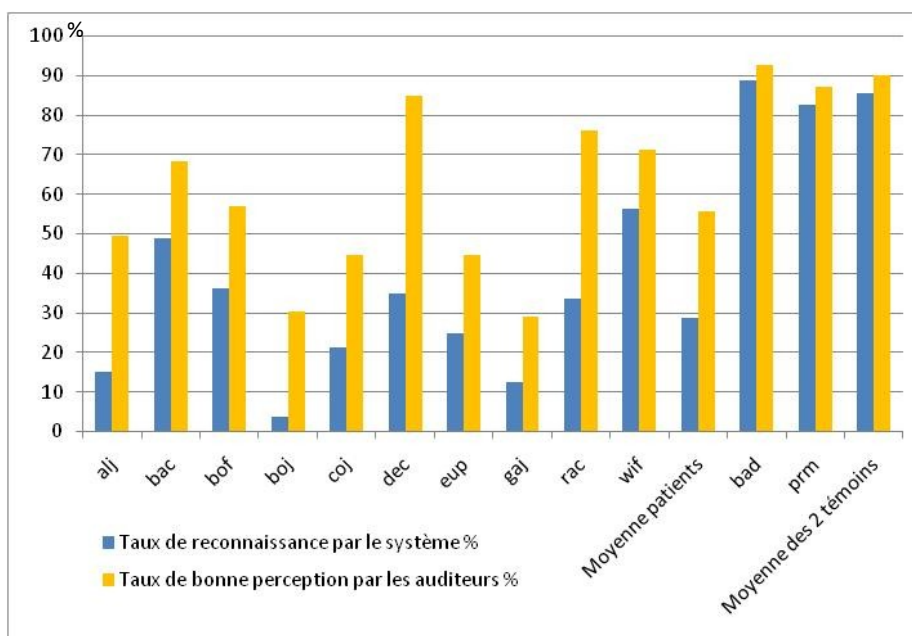


Figure n°5.3.3.3 : Histogramme des taux de reconnaissance du système et de perception juste pour les auditeurs par rapport à la cible pour les syllabes

La différence entre le taux de reconnaissance par le système et le taux de perception juste par les auditeurs se situe entre 20 à 50% en fonction des patients. Cette différence est de moins importante (environ 10%) pour les deux témoins (figure n°5.3.3.3).

#### 5.3.3.2.2.2. DESCRIPTION DES RESULTATS POUR LES CONSONNES

Les résultats en terme de taux de reconnaissance des consonnes cible en se basant sur la consonne du candidat 1 ont été analysés. Le taux de reconnaissance par le système des consonnes cible est de 28,63% pour les patients (figure n°5.3.3.4). Les phonèmes les mieux reconnus sont: /s/, /f /, /r/, /m/, /z/, /l/ ; les moins reconnus sont : /g/, /p/, /k/.

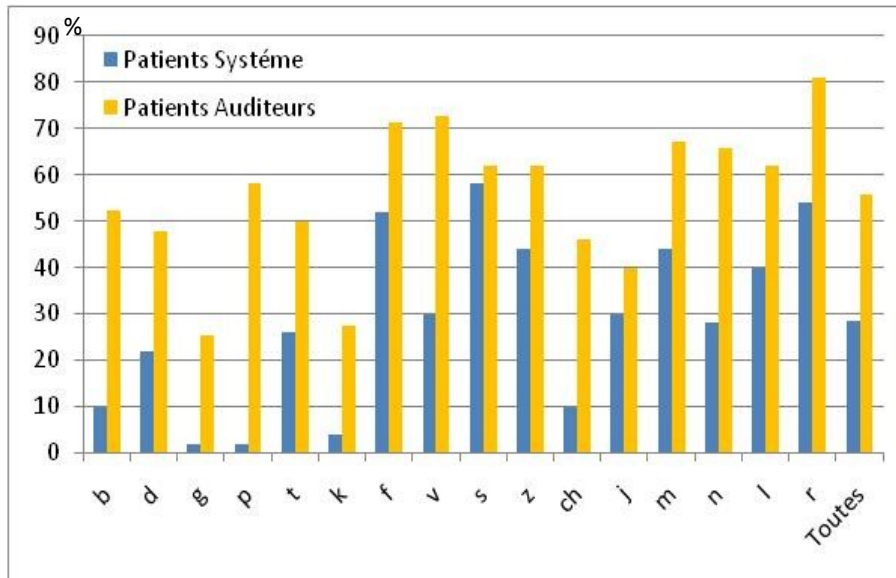


Figure n°5.3.3.4 : Histogramme des taux de reconnaissance du système et de perception juste par les auditeurs par rapport à la cible pour la consonne du candidat 1 pour les patients.

On note que les taux de reconnaissance sont bien meilleurs pour les fricatives non voisées /f/,/s/ et qu'il est particulièrement moins bon pour les occlusives notamment le /g/ comme cela était noté chez les sujets contrôlés mais également le /p/ puis sur le /k/, /ch/ et le /b/.

Ces taux de reconnaissance présentent des variations importantes en fonction de la voyelle et sont meilleurs avec la voyelle [a] (tableaux n°5.3.3.1).

Voyelle_cible	Accord cible_reco_cand1			Taux (%) de reconnaissance des consonnes
	Erreur	Correct	Total patients	
a	97	63	160	39,38
an	117	43	160	26,88
i	116	44	160	27,50
ou	122	38	160	23,75
u	119	41	160	25,63
<b>Total patients</b>	<b>571</b>	<b>229</b>	<b>800</b>	

Tableau n°5.3.3.1 : Taux de reconnaissance des consonnes (candidat 1) en fonction de la voyelle cible pour les patients

Enfin, ont été comparés les résultats de la reconnaissance par le système et par les auditeurs. Ces résultats sont présentés sur la figure n°5.3.3.4 pour les patients et la figure n°5.3.3.5 pour les témoins.

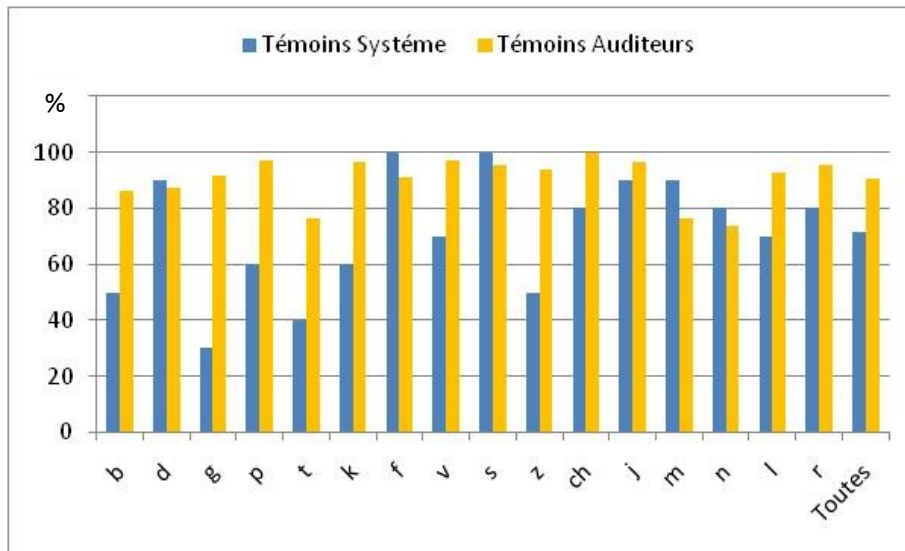


Figure n°5.3.3.5 : Histogramme des taux de bonne reconnaissance du système et de bonne perception par les auditeurs par rapport à la cible pour la consonne du candidat 1 pour les témoins

Pour les témoins, quelques consonnes principalement occlusives sont relativement moins bien reconnues par le système : /g/,/t/ puis /z/,/b/ et /p/,/k/. Pour les patients, ce phénomène est encore plus prégnant sur les occlusives /g/,/p/,/k/,/b/ alors que cela est moins important sur les fricatives /ch/, /v/ et le /n/. Ce phénomène traduit la difficulté que rencontrent en général les systèmes de reconnaissance automatique pour les consonnes occlusives.

Cependant, le but de ce travail n'est pas d'évaluer la performance d'un système de reconnaissance mais d'utiliser sa paramétrisation pour obtenir une mesure d'un indice de vraisemblance qui dans notre problématique répondra à un indice de degré de distorsion de la production. Cette analyse a été faite à titre de vérification préalable à l'utilisation de la reconnaissance pour calculer cet indice par rapport à la cible. Les résultats sont suffisants pour poursuivre avec une limite d'interprétation concernant la consonne [g] et une vigilance quant à l'exploitation du trait dorsovélaire.

#### 5.3.3.1.3. EN RESUME

1. Un taux de reconnaissance suffisant (77% en moyenne) et l'absence de problème d'alignement décelé sur le groupe contrôle sont des arguments qui autorisent la poursuite de ce travail.
2. A la lumière de l'analyse des taux de reconnaissance des syllabes du groupe contrôle (10 locuteurs) et des résultats des 2 témoins permettant de comparer le taux de reconnaissance et le taux de perception, la consonne /g/ pose pour ce corpus et dans la condition particulière d'utilisation du

système des difficultés qui risque rendre ce phonème inutilisable pour le reste du travail. Pour les mêmes raisons mais à un degré moins important, une vigilance s'impose pour le son /k/ et donc le trait dorsovélaire. Les sons /t/, /p/, /b/ sont moins bien identifiés pour les 2 témoins ayant participé au test seulement. Une attention particulière sera portée aux occlusives notamment labiales.

3. En comparaison avec le test de perception,

- pour les témoins, les résultats sont différents avec des taux les plus bas pour /g/, /k/ et /v/ dans un environnement vocalique /ou/, /u/, /i/ alors que pour les tests perceptifs les taux les plus bas sont observés pour le /n/ et /m/ dans un environnement /an/ ;

- en revanche pour les patients, le même phénomène est observé en ce qui concerne le rôle de la voyelle (meilleur score avec le /a/ dans les 2 cas), assez proche pour les bons scores (/f/, /r/ dans les 2 cas mais une divergence pour le /v/ par exemple). Les différences sont plus nettes dans les mauvais scores notamment pour le p.

### 5.3.3.2. DETERMINATION DE L'INDICE DE VRAISEMBLANCE DE LA CONSONNE CIBLE A PARTIR DU SCORE DE VRAISEMBLANCE SYLLABIQUE.

Plusieurs méthodes sont possibles pour obtenir une mesure du degré de distorsion pour la consonne cible. Ont été calculés la moyenne des rangs, la médiane des rangs et le score normalisé moyen de la consonne cible. Les statistiques descriptives concernant ces indices sont présentées sur la figure n°5.3.3.6. La médiane et la moyenne des rangs sont en moyenne respectivement de 48,59 et 49,2 avec une déviation standard de 39,74 et 36,62 pour les patients. Elles sont de 22,58 et 23,61 pour la moyenne et de 22,43 et 21,31 pour les témoins. Le score normalisé est de 0,98 avec une déviation standard de 0,14. Elle est de 0,99 en moyenne avec une déviation standard de 0,06 pour les témoins.

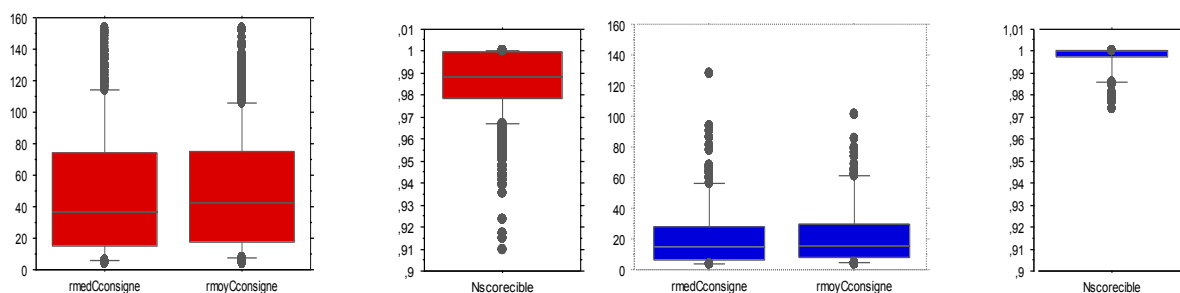


Figure n°5.3.3.6 : Représentation graphique en boîte à moustache des indices du degré de distorsion de la consonne cible. En rouge pour les patients, en bleu pour les témoins – de gauche à droite : la médiane du rang, la moyenne du rang et les score normé moyen.



Ces résultats donnent une information sur la dispersion, la symétrie, l'étendue des mesures laissant apparaître une exploitation plus aisée des mesures utilisant le rang avec une dispersion un peu moins grande pour la moyenne mais une étendue un peu plus grande pour la médiane et une distribution moins symétrique. Ces différentes méthodes de mesures ont été évaluées quant à leur pertinence dans les différentes situations d'analyse : par consonnes et par trait, par type de locuteur (patients, témoins) et par site tumoral (pour les patients). Les graphiques représentant les résultats figurent en annexe, section 7.3.1.3.

Les 3 modalités de mesure sont concordantes. Les résultats sont moins visibles du fait de l'asymétrie de la distribution propre au score normé. Les différences au niveau des traits sont plus nettes avec le rang médian par rapport au rang moyen.

Au vu de ces différents résultats, nous avons retenu de traiter les résultats avec la moyenne des rangs médians de la consonne cible.

### 5.3.3.3. ANALYSE DES RESULTATS AVEC LE RANG MEDIAN

#### 5.3.3.3.1. POUR LES TEMOINS

Les répartitions du rang médian en fonction des consonnes et des traits est représentée pour les témoins sur les figures n°5.3.3.10 et n°5.3.3.11.

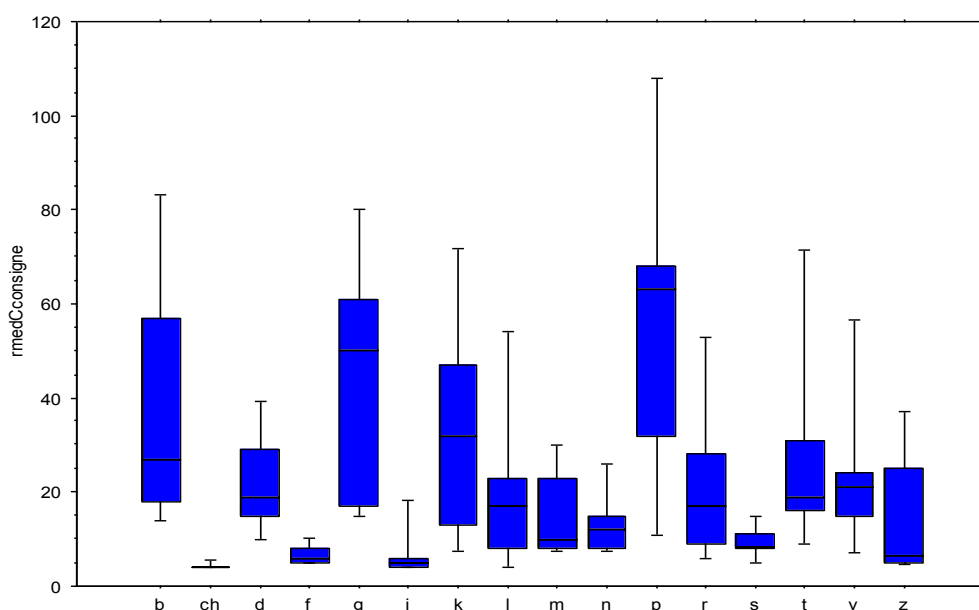


Figure n°5.3.3.10 . Graphe en boîte de la Médiane du rang des consonnes cibles pour les témoins.

L'analyse de la moyenne du rang médian pour la consonne cible apporte des précisions sur les difficultés du système pour reconnaître les consonnes /g/ et /k/, /p/ et /b/.

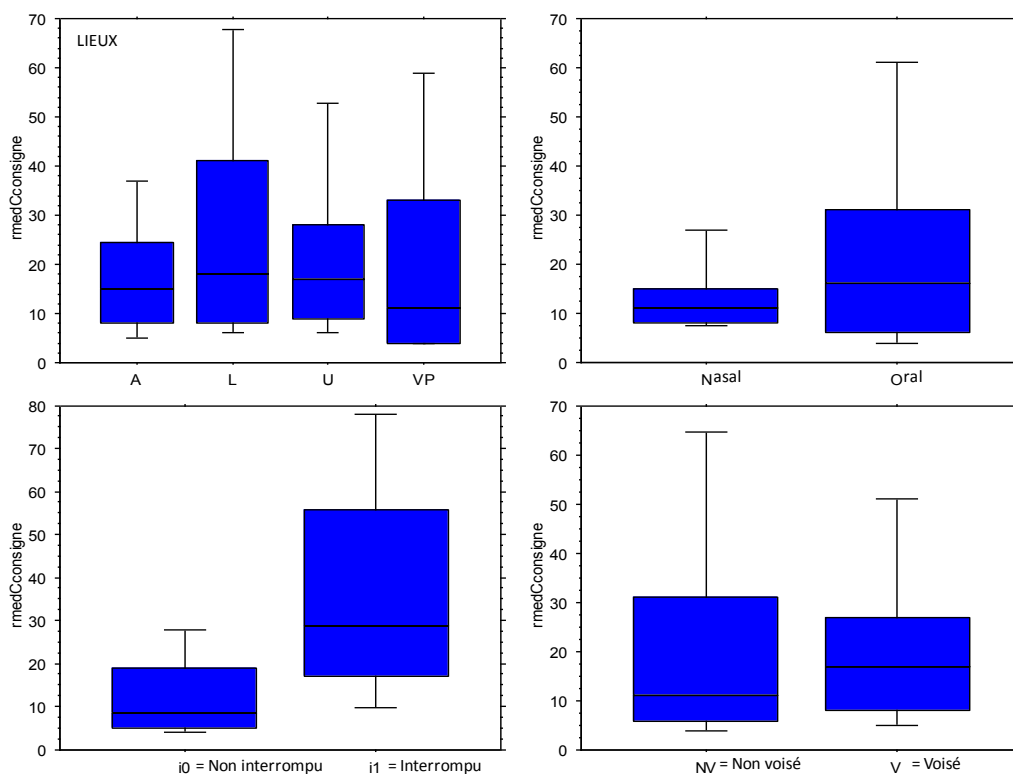


Figure n°5.3.3.11 : Graphe en boîte de la Médiane du rang de la consigne ou cible en fonction des traits pour les témoins.

Les performances sont très bonnes pour les fricatives non voisées : /ch/, /j/, /f/.

### 5.3.3.3.2. POUR LES PATIENTS

Pour les patients les résultats sont représentés sur les figures n°5.3.3.12 et n°5.3.3.13.

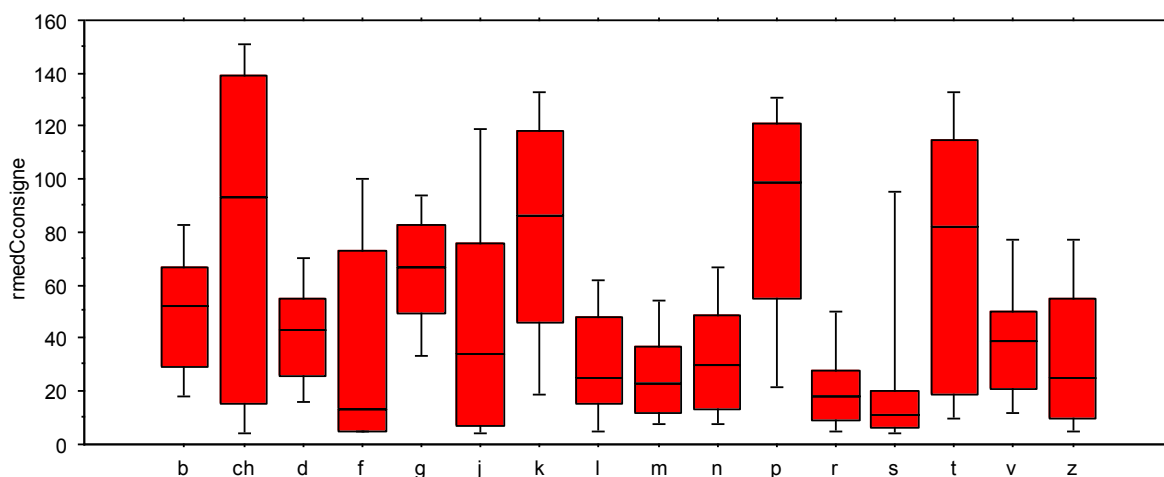


Figure n°5.3.3.12 : Graphe en boîte la Médiane du rang des consonnes pour les patients

Pour les patients /ch/,/p/,/k/,/t/ présentent une médiane de rang très éloignée. A l'opposé, /s/ et /r/ puis /f/,/z/,/l/,/m/,/n/ présentent un rang proche (figure n°5.3.3.12).

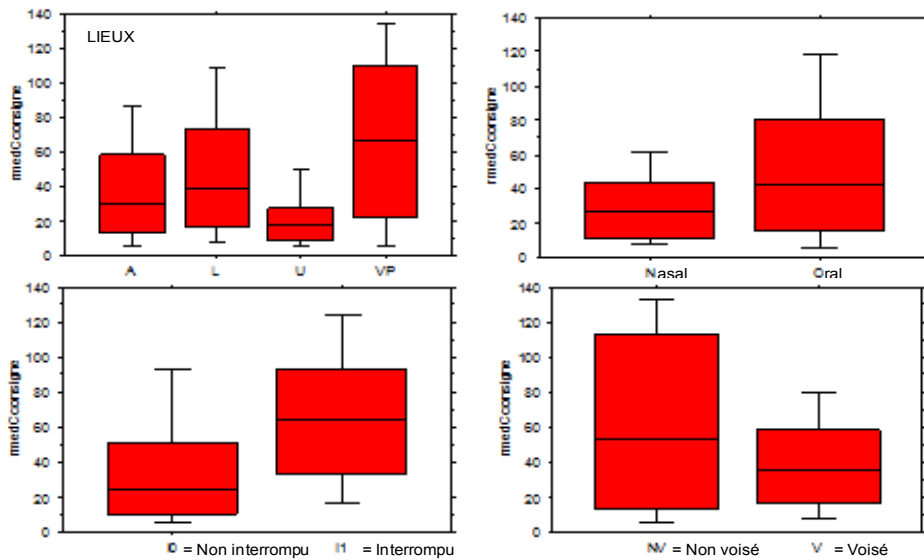


Figure n°5.3.3.13 : Graphe en boîte de la Médiane du rang en fonction des traits (A : alvéolaire, L : labial, U : uvulaire, VP : vélopalatal).

Le trait uvulaire présente une meilleure médiane de rang suivi du trait non interrompu, nasal, alvéolaire, voisé puis labial. Le trait vélopalatal a le rang le plus éloigné. (Figure n°5.3.3.13).

### 5.3.3.3.3. COMPARAISON DES RESULTATS POUR LES DEUX GROUPES

En prenant en considération la différence de moyenne entre les témoins et les patients, on note une différence particulièrement importante pour la consonne /ch/, puis /k/, /t/, et /j/ relativisant la position pour le /p/ par exemple. Aucune différence n'est observée pour le /r/. Cette différence reste inférieure à 20 pour /l/,/b/,/m/,/v/ (figure n°5.3.3.3.14).

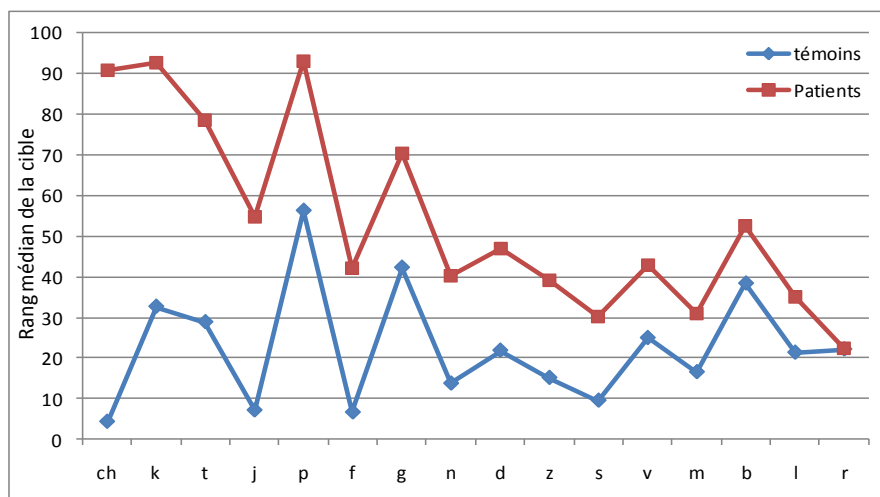


Figure n°5.3.3.3.14. Histogramme représentant la moyenne de la médiane des rangs par consonne pour les témoins et les patients par ordre décroissant de la différence entre témoins et patients

La conservation des différences relatives entre les consonnes est remarquable, comme pour le taux de perception juste. Elle n'est pas respectée pour les consonnes /ch/ et à moindre degré pour /l/ et /n/.

#### 5.3.3.3.4. A RETENIR

1. L'analyse sera faite avec la moyenne du rang médian pour la consonne cible.
2. L'étude des témoins met en évidence une difficulté du système pour les consonnes occlusives bilabiales et dorso vélares.
3. L'étude des patients, l'éloignement du rang médian des consonnes /ch/, /p/, /k/, /t/ reste net même si il est relativisé par la comparaison aux scores des témoins pour le P. Pour les rangs proches : /s/ et /r/ puis /f/, /z/, /l/, /m/, /n/, la comparaison avec l'observation des témoins relativise beaucoup plus les résultats avec notamment une possible surestimation du /s/ et de manière plus générale, des fricatives.

---

### 5.3.4. TEST DES HYPOTHESES

---

#### 5.3.4.1. RAPPEL DES HYPOTHESES

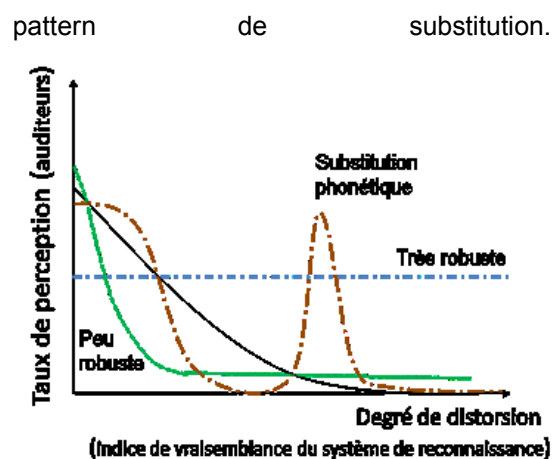
Dans la partie 5.1.2 Hypothèses, nous avons proposé que l'utilisation d'un système automatique de reconnaissance vocale pouvait permettre d'estimer un degré de dégradation du signal et de comparer l'impact de ce degré de dégradation sur l'intelligibilité en termes de taux de perception en fonction des phonèmes ou autres unités linguistiques (des traits à la syllabe) et ainsi de déterminer la robustesse du unités linguistiques (figure n°5.3.4.1). En utilisant une représentation graphique du taux de reconnaissance juste et de la précision en fonction du rang médian de la consonne cible et une représentation de la F mesure en fonction rang médian de la consonne cible nous repérerons les consonnes ou les traits qui malgré la distance par rapport au générique de la cible maintiennent leur taux de perception. Nous espérons démontrer :

- a. qu'il existe des unités linguistiques dont la distorsion donne lieu à des compensations, soit par ce qu'elle résiste à la distorsion (pente faible, caricaturée par la droite horizontale en pointillés bleus sur la figure) soit parce qu'il existe des substitutions phonétiques donnant lieu à des restaurations phonologiques provoquant une

réascension en cloche de la courbe (remontée du taux de perception pour des cibles distantes de la production « canonique » du système).

- b. qu'il existe des unités peu robustes se traduisant par la chute très rapide du taux de perception figuré par la courbe en pointillés verts.

Figure n°5.3.4.1 : La comparaison de l'impact du degré de dégradation sur l'intelligibilité en termes de taux de perception. La courbe en trait plein correspond à la situation supposée normale, c'est-à-dire une diminution progressive du taux de perception avec l'éloignement par rapport à la production canonique du système. Les courbes en pointillées illustrent nos hypothèses : en vert, la situation de faible robustesse, en bleu, la situation de grand robustesse et en marron, la présence d'une compensation avec un



Nous avons également proposé la recherche du même type d'information en étudiant la distribution des différentes unités dans la représentation graphique du rang de l'unité perçue en fonction du rang de la cible (figure 5.3.4.2). Nous espérons ainsi repérer :

- c. Les unités linguistiques particulièrement robustes dans l'extrémité de la bissectrice formée par l'adéquation du rang cible et du rang perçu correspondant à la perception juste des auditeurs alors que la production des cibles est très peu « vraisemblante ».
- d. les unités linguistiques particulièrement fragile correspondant à des groupes situés sur le diagramme de la figure à un rang Percu >> au rang Cible soit une localisation dans la partie supérieure gauche
- e. les unités linguistiques particulièrement exposés aux confusions situées en bas à droite sur le diagramme de la figure.

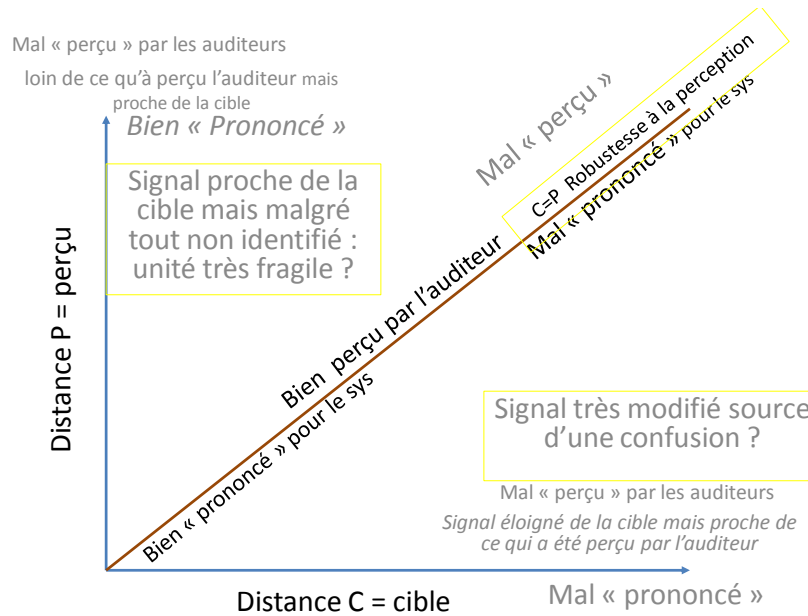


Figure n°5.3.4.2 : Hypothèses d'interprétation de la distance par rapport à la cible ou à la perception

### 5.3.4.2. LA COMPARAISON DE L'IMPACT DU DEGRE DE DEGRADATION (OBTENU PAR LE SYSTEME) SUR L'INTELLIGIBILITE (OBTENU PAR LE TEST DE PERCEPTION PAR LES AUDITEURS).POUR LES CONSONNES.

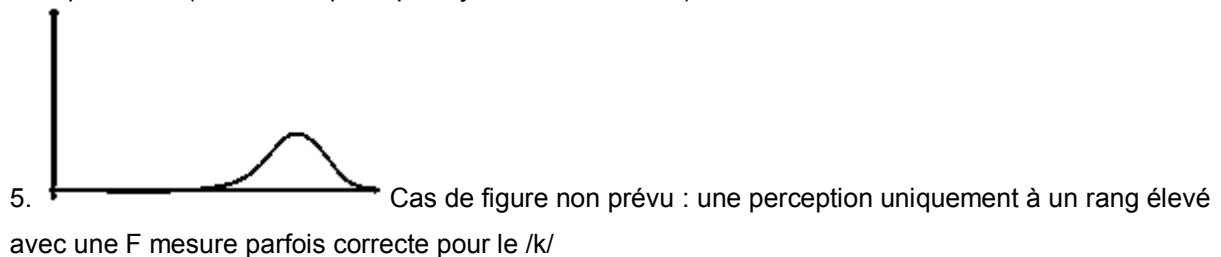
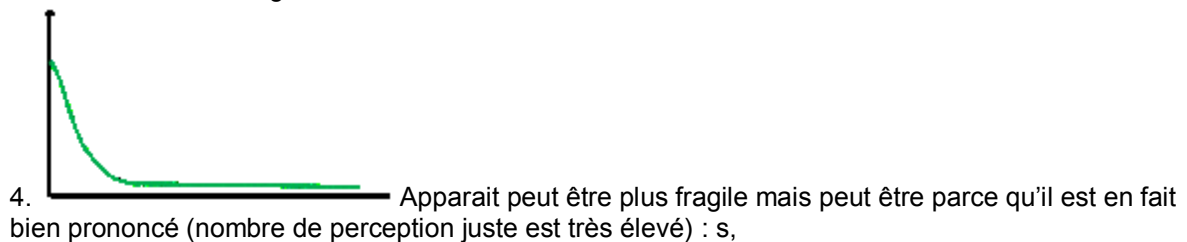
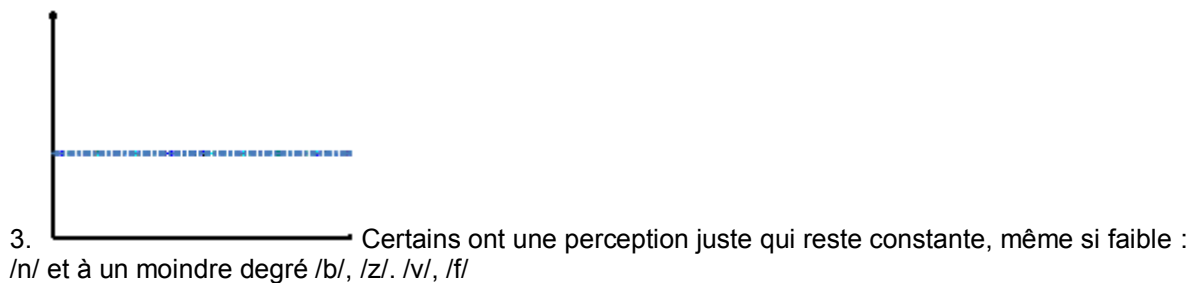
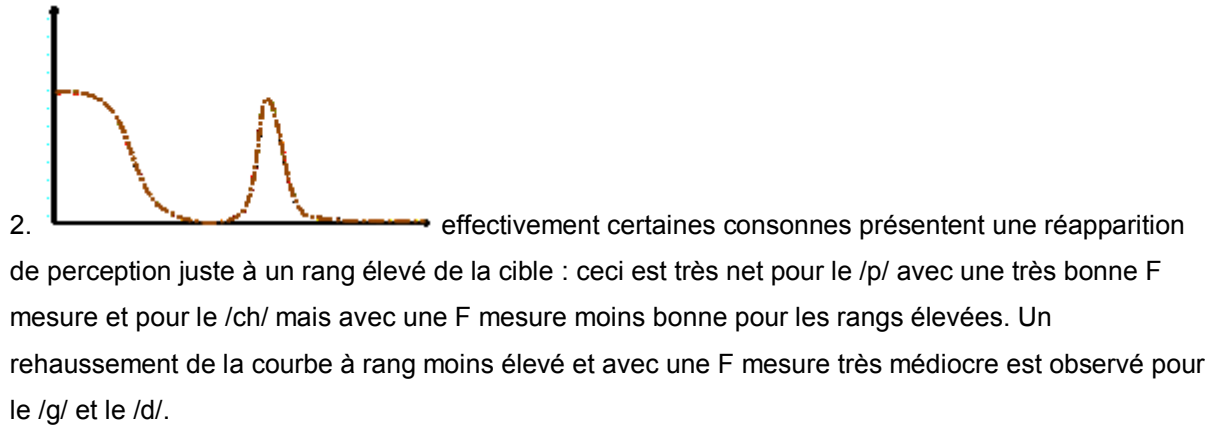
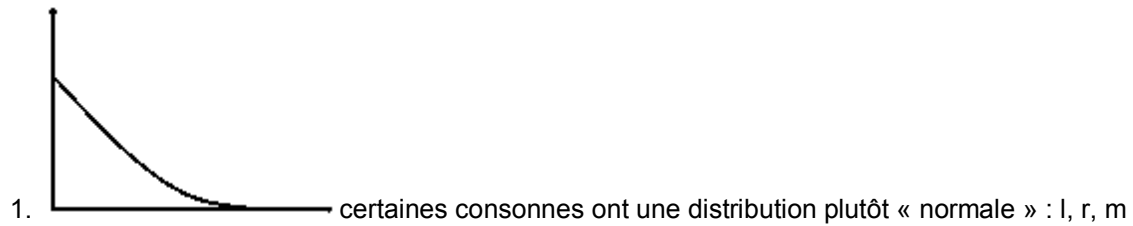
Compte tenu des résultats du test de perception, nous allons utiliser dans un premier temps les résultats des courbes de la F mesure puis interpréterons si besoin les graphiques associant le taux de perception juste ou rappel et le taux de précision (en se rappelant que le taux de confusion = 1 - précision).

#### 5.3.4.2.1. LA F MESURE EN FONCTION DU RANG MEDIAN POUR LES CONSONNES.

Sur la figure n°5.4.3.3, les points représentent le nombre de cibles présentes en fonction de la valeur du taux et du rang. Plus les points sont gros, plus il y a de cibles. La courbe correspond à un lissage, le halo gris qui l'encadre correspond à la probabilité de la courbe, plus il est étroit, plus la courbe bleue est vraisemblable, plus le halo est large, plus elle perd de sa signification. Pour bien interpréter ces graphiques, il faut relativiser la tendance de la courbe en fonction du nombre de points présents dans la zone de rang représentés par les points eux-mêmes et la taille du halo gris. Ceci est important dans la mesure où même en l'absence de points, une courbe est tracée. Cette représentation, du fait du

lissage, donne des courbes de tendance qui « collent » moins aux modèles proposés lors de l'élaboration des hypothèses mais ce sont les plus justes que nous puissions proposer.

On repère cependant que :



Le /j/ et le /t/ sont particulièrement difficiles à interpréter. Il faut de plus garder en mémoire que les performances du système automatique sont mauvaises pour /g/ et à un moindre degré /z/, /b/ puis /p/ et /k/.

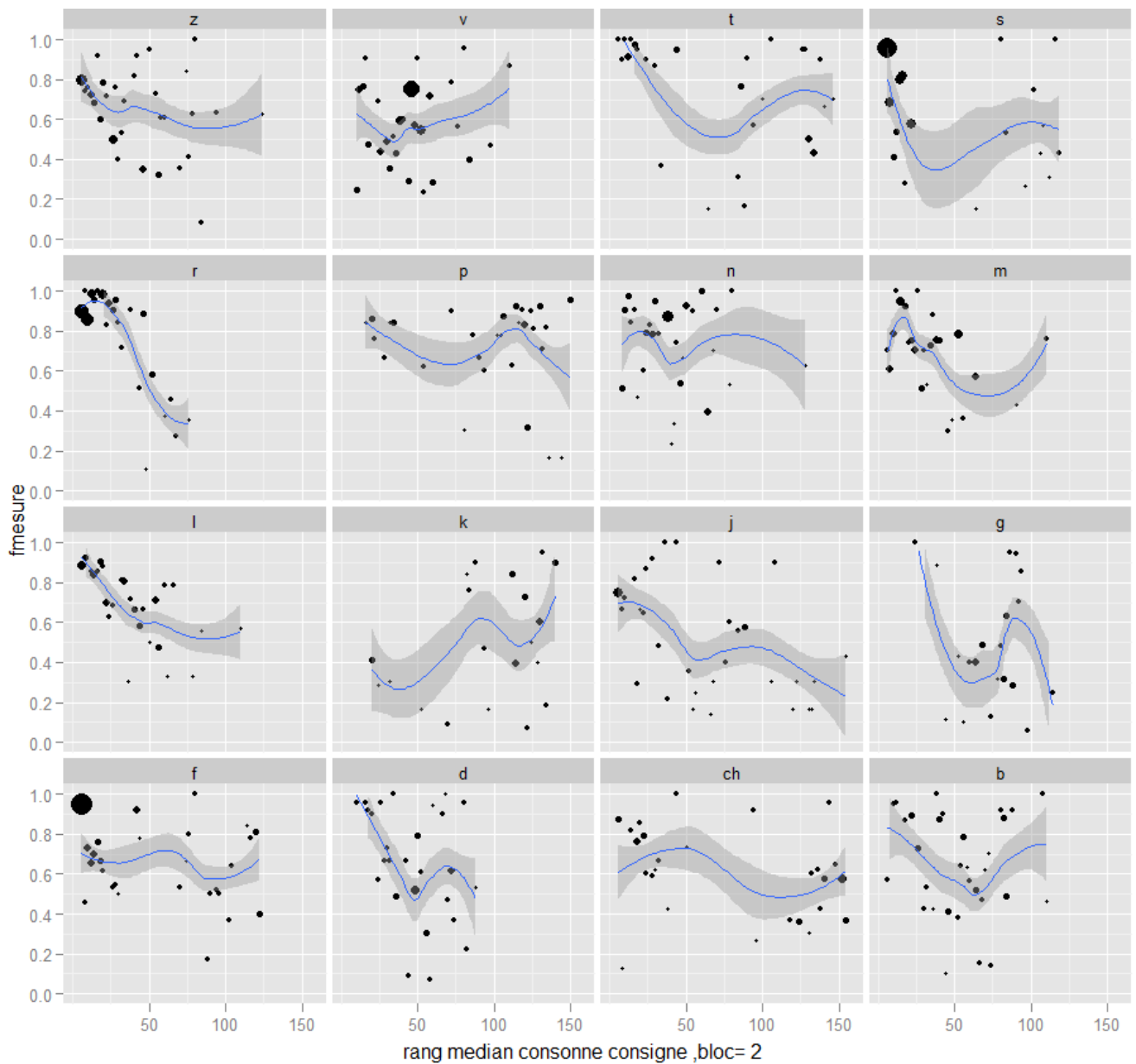
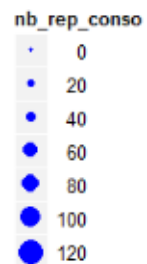


Figure n°5.4.3.3 : Représentation graphique de la f mesure en fonction du rang médian cible.

Sur ces figures, les points représentent le nombre de cibles présentes en fonction de la valeur de la f mesure et du rang. Plus les points sont gros, plus il y a de cibles. La courbe correspond à un lissage, le halo gris qui l'encadre correspond à la probabilité, plus il est étroit, plus la courbe bleue est vraisemblable, plus le halo est large, plus elle perd de sa signification





#### 5.3.4.2.2. LE RAPPEL OU TAUX DE PERCEPTION JUSTE ET LA PRECISION EN FONCTION DU RANG MEDIAN POUR LES CONSONNES.

Figure n°5.3.4.4, le trait bleu représente le rappel, le trait rouge la précision. Comme sur la figure précédente, les points représentent le nombre de cibles présentent en fonction de la valeur du taux et du rang. Plus les points sont gros, plus il y a de cibles. La courbe correspond à un lissage, le halo gris qui l'encadre correspond à la probabilité, plus il est étroit, plus la courbe bleu est vraisemblante, plus le halo est large, plus elle perd de sa signification.

Ainsi, l'aspect de courbe « normale » attendue n'est pas retrouvée. Le r s'en rapproche avec un taux de perception plus élevé sur les 50 premiers rangs, puis une chute assez rapide jusqu'au rang 75. Le z, le l et le m également, moins bien perçu, le taux chute plus lentement jusqu'au rang 100. Après le faible nombre de points ne permet de retenir la tendance. Le /j/, qui n'était pas classé, peut être mis dans cette catégorie.

L'aspect de la courbe robuste est retrouvé pour le /v/, /n/. En revanche, cette représentation conduit à reclasser le /b/ dans une robustesse normale en tenant compte du fait qu'il est mal produit et le d dans le groupe des unités produites peut être de deux manières différentes avec donc un partie de substitution.

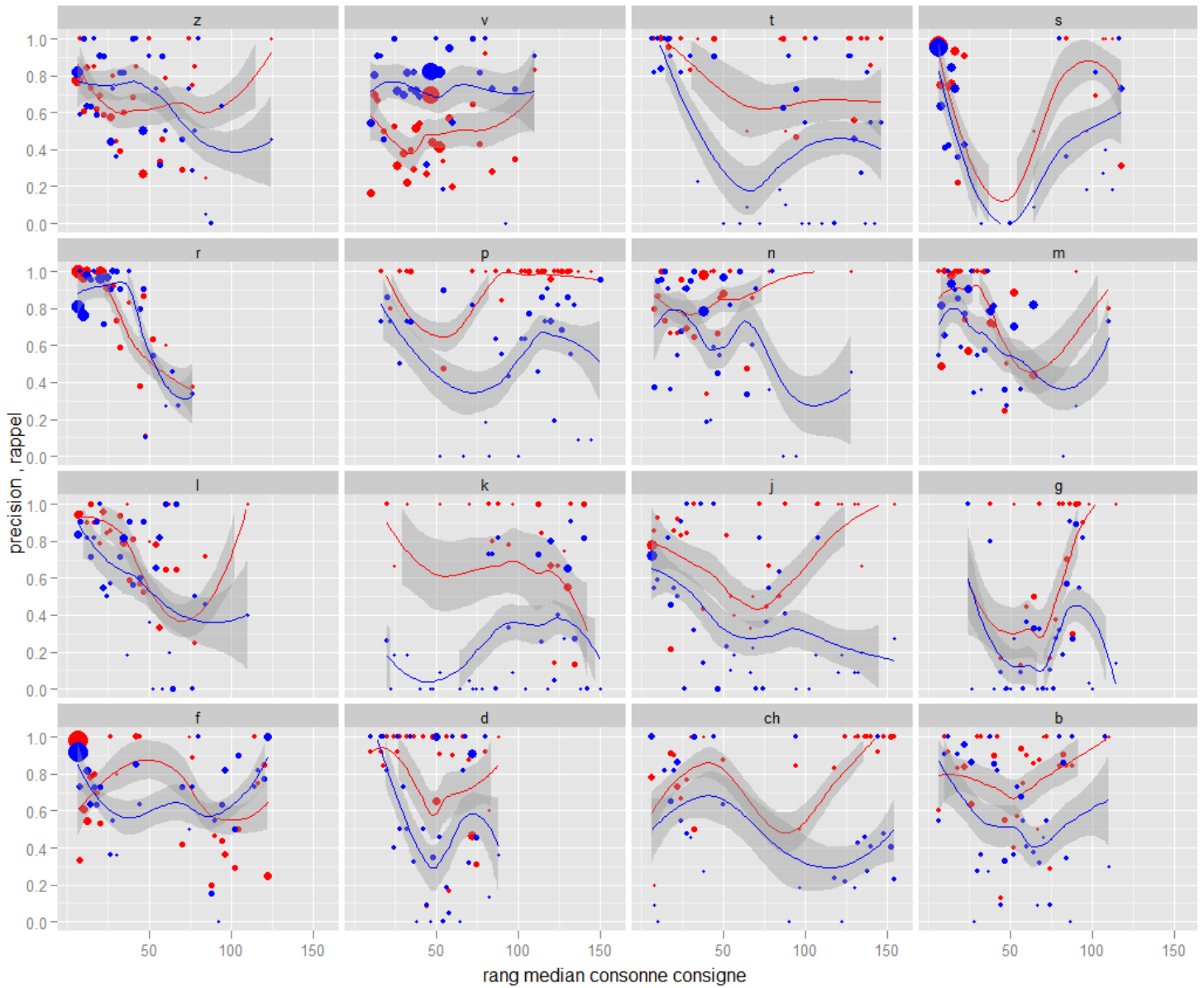
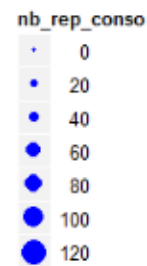


Figure n°5.3.4.4 : Représentation graphique du rappel (en bleu) et de la précision (en rouge) en fonction du rang médian cible.

Sur ces figures, les points représentent le nombre de cibles présentes en fonction de la valeur de la f mesure et du rang. Plus les points sont gros, plus il y a de cibles. La courbe correspond à un lissage, le halo gris qui l'encadre correspond à la probabilité, plus il est étroit, plus la courbe bleu est vraisemblable, plus le halo est large, plus elle perd de sa signification.



### 5.3.4.2.3. EN SYNTHÈSE

Le tableau n°5.3.4.1 illustre la synthèse de l'interprétation des résultats du test de perception en fonction du rang de la consonne cible.

	F mesure	Rapp el	Précision	Synthèse
Normale	l,r,m	l, z, m,	r, l,m	<b>l, m</b>
Normal mal produit		j, b	j	j?
Normale mais mieux produit ou moins robuste ?		r		<b>r ?</b>
Robuste	n,b,v,z,f	v, n	b,n,v,z	<b>n,v z, b ?</b>
Robuste mal produit			t	
Compensation pour une partie	p,ch	p,d	p	P, ch ?
Compensation sur fragilité	d, g			<b>d ,g</b>
Fragile		f	d	
Fragile ? bien produit	s	s	s	<b>s</b>
Fragile mal produit		t		
Substitution possible sur une grande distorsion	k	k, g	g	<b>g , k</b>
Non défini	t,j	ch	ch, f,k	t

Tableau n°5.3.4.1 synthèse de l'interprétation des résultats du test de perception en fonction du rang de la consonne cible

En considérant la fiabilité des résultats obtenus par le système automatique de reconnaissance, les consonnes /n/ et /v/ semblent particulièrement robuste perceptivement et nous disposons d'arguments forts pour conclure qu'ils sont plus contributifs pour l'intelligibilité en français que d'autres consonnes, notamment les consonnes /l/ et /m/.

Des phénomènes de compensation phonologiques par attribution d'un sens à des unités dont la distorsion est compatible avec une « ré-allocation » catégorielle sont fortement suspectés mais le degré de certitude concernant leur identification est limité par leur résultat lors de la reconnaissance automatique. Il s'agit de /p/, /d/, /g/, /k/.

### 5.3.4.2 INTERPRÉTATION DES DIAGRAMMES R\_CIBLE EN FONCTION DE R\_PERCU

### 5.3.4.2.1. RECHERCHE DES UNITES LINGUISTIQUES PARTICULIEREMENT ROBUSTE A LA PERCEPTION

En reprenant l'hypothèse concernant la bissectrice du diagramme Distance cible/Distance perçue, l'objectif est de connaître la distribution des consonnes le long de cette droite (figure n°5.3.4.5).

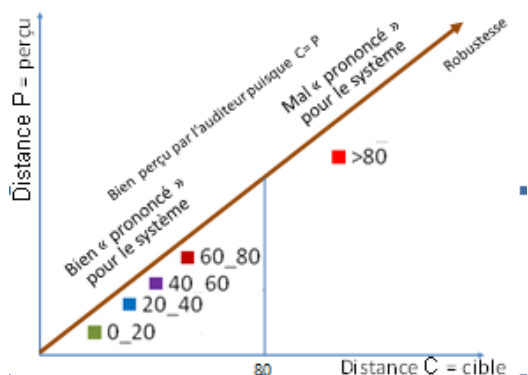


Figure n°5.3.4.5 : Etude de la bissectrice du diagramme Distance C, cible, en fonction de la distance P, perçue

Pour cela, nous avons recherché le nombre de chaque consonne cible en « découpant » la droite en 5 portions : du rang 0 à 19, du rang 20 à 39, du rang 40 à 59, du rang 60 à 79 et du rang supérieur ou égal à 80 comme l'illustre la figure n°5.3.4.6.

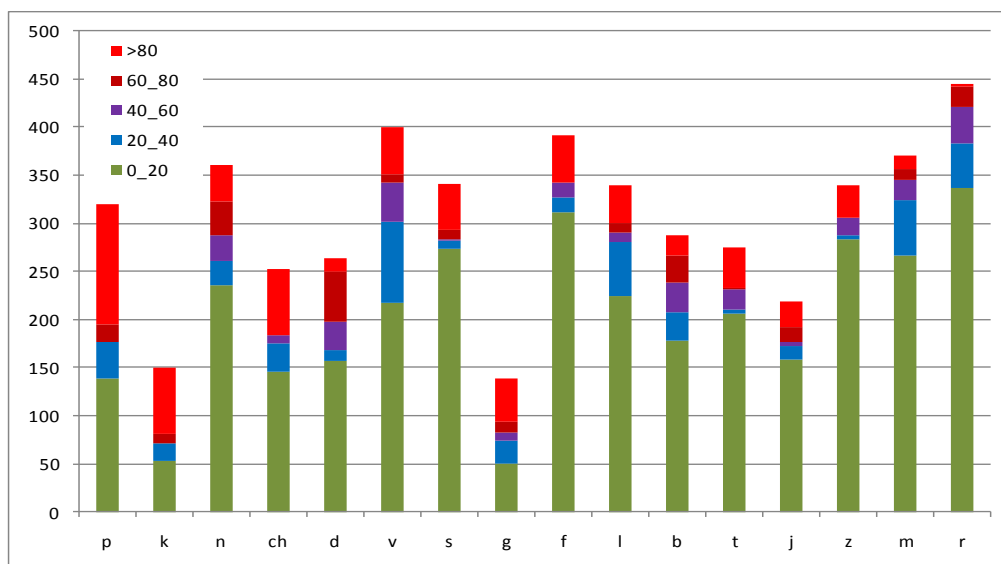


Figure n°5.3.4.6 : Répartition des différentes consonnes en répondant à la définition de la bissectrice (soit C=P) en fonction du rang médian de la cible classé par ordre décroissant de consonnes à partir du rang 60.

En nombre d'occurrences, dans la portion de la bissectrice supérieure au rang médian 60, les 5 consonnes les plus fréquentes sont : /p/, /k/, /n/, /ch/, /d/.

### 5.3.4.2.2. RECHERCHE DES UNITES LINGUISTIQUES PARTICULIEREMENT FRAGILES

La figure 5.3.4.8 illustre l'objectif de cette recherche. L'objectif est ici de décrire le contenu du rectangle défini par un rang médian cible inférieur à 50 et un rang médian perçu supérieur à 50.

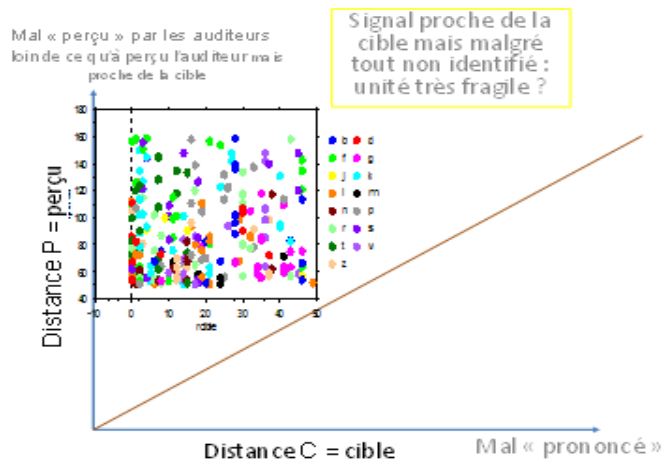


Figure n°5.3.4.8 : Principe de la recherche des unités fragiles

Le contenu du « rectangle » est décrit en comparant le pourcentage des consonnes restantes à un seuil de rang perçu > à 50 et > à 100, pourcentage par rapport au nombre de consonnes figurant en dehors de la bissectrice du graphe (figure n°5.3.4.9).

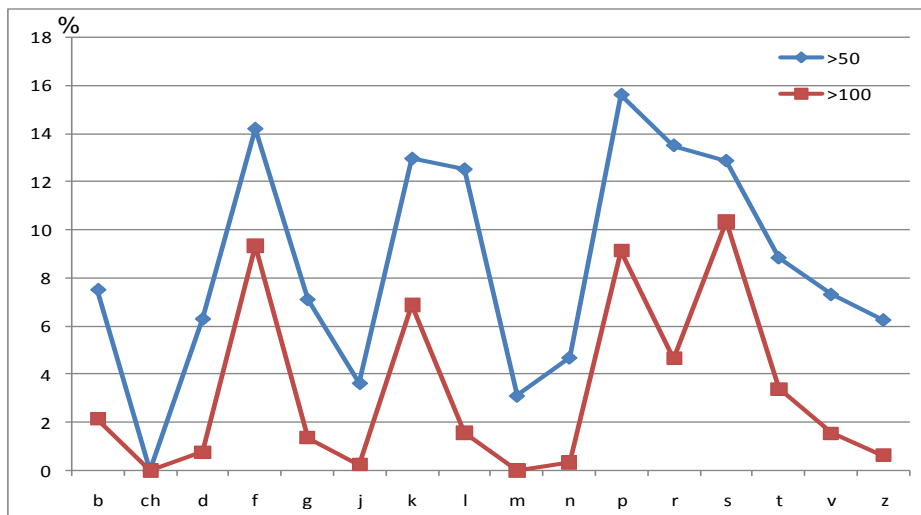


Figure n°5.3.4.9 : Pourcentage de consonnes restantes dans la zone définie par un rang cible < à 50 avec un rang perçu > à 50 en bleu et à 100 en rouge.

Dans la partie extrême (>100), sont donc « bien prononcés » pour le système mais loin de ce qui a été par l'auditeur : /s/,/f/, /p/,/k/ correspondants respectivement à 33, 25, 31,35 occurrences.

### 5.3.4.2.3. ANALYSE DE L'AUTRE ZONE « EXTREME » A LA RECHERCHE DE CONSONNES SOURCE DE CONFUSION

La figure n°5.4.3.10 illustre l'objectif de cette recherche. Il s'agit ici de décrire le contenu du rectangle défini par un rang médian perçu inférieur à 50 et un rang médian cible supérieur à 50.

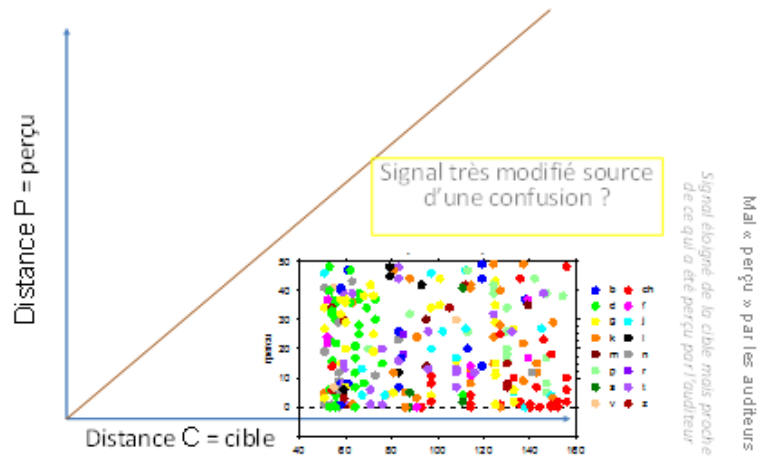


Figure n°5.4.3.10 : Principe de la recherche des unités sources de confusion

Le contenu du « rectangle » est décrit en comparant le pourcentage des consonnes restantes à un seuil de rang cible > à 50 et > à 100, pourcentage par rapport au nombre de consonnes figurant en dehors de la bissectrice du graphe (figure n°5.4.3.11).

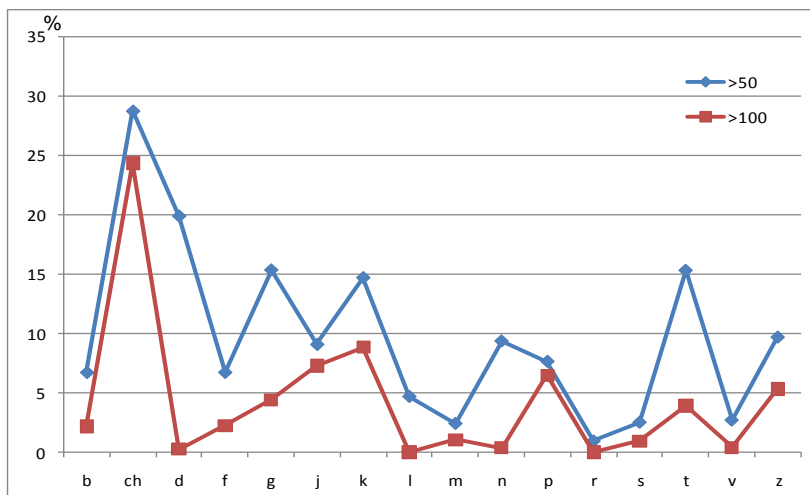


Figure n°5.4.3.11 : Pourcentage de consonnes restantes dans la zone définie par un rang perçu < à 50 avec un rang cible > à 50 en bleu et à 100 en rouge.

Dans la partie extrême (>100), sont donc « bien perçu mais mal prononcé» pour le système alors qu'ils ont été mal perçu par l'auditeur : /ch/ et à un moindre degré /k/, /p/, /z/ et /t/ correspondants

respectivement à 99, 45, 22,17,15 occurrences. Ce résultat pourrait sembler contradictoire avec les résultats de la matrice de confusion qui met en évidence un fort taux de confusion pour le f et v. L'explication réside dans le fait que la consonne /ch/ donne lieu à des transformations alors que le /f/ et le /v/ sont eux plutôt bien reconnus et sont la conséquence des transformations. Les arguments confirmant cette hypothèse sont que /ch/ est le moins bien perçu juste par les auditeurs des trois et que les confusions à partir de la cible /ch/ sont à l'origine de la perception d'un /f/ dans 60% des cas.

#### 5.3.4.2.4. EN SYNTHÈSE

La figure n°5.4.3.12 illustre la synthèse de l'interprétation des résultats issus du système automatique de reconnaissance

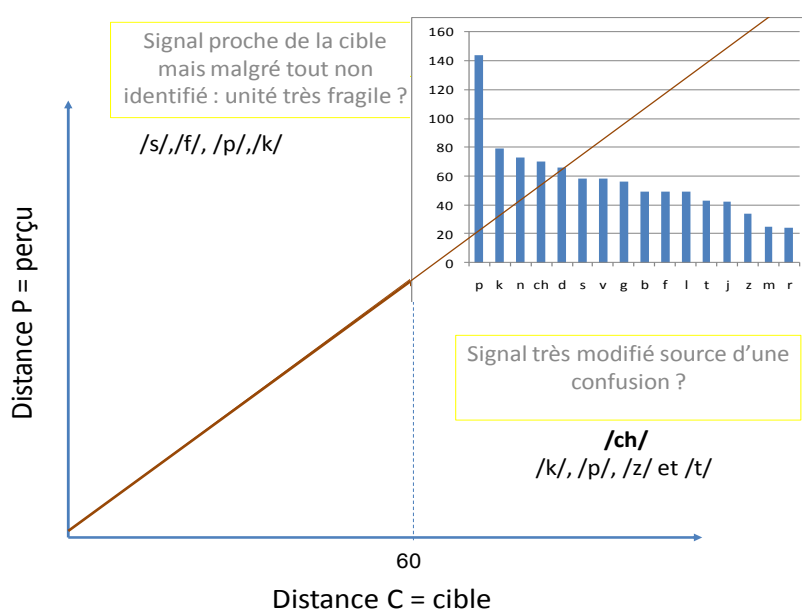


Figure n°5.4.3.12 : synthèse de l'interprétation des résultats issus du système automatique de reconnaissance

Encore un fois, ces résultats doivent être interprétés en fonction du degré de certitude quand au rang calculé par le système automatique pour certaines consonnes. Ainsi, parmi les consonnes « fragiles », le /s/ contribue plus à l'intelligibilité en français que les autres consonnes. Le /ch/ peut être retenu comme une unité qui est particulièrement source de confusion. Parmi les unités robustes repérées sur l'extrémité de la bissectrice, le /n/ et le /ch/ ne pose pas de problème par rapport à la fiabilité du rang donné par le système automatique.

## 5.4. DISCUSSION

### 5.4.1. SYNTHÈSE DES RESULTATS

Les différentes approches contribuent à identifier la robustesse perceptive des consonnes et à en déduire leur contribution à l'intelligibilité. Ces résultats sont à pondérer en raison des mauvaises performances de la reconnaissance vocale pour les consonnes suivantes citées par degré décroissant : /g/ et /t/ puis /z/ et /b / et enfin /p/ et /k/.

Le tableau n°5.4.1 regroupe toutes les informations recueillies et autorise les conclusions illustrées par la figure n°5.4.1.

	% de perception /sévérité	F mesure / rang médian	% de perception ou Rappel / rang médian	% de perception Précision	Bissectrice du diagramme	R perçu >100 dans le diagramme	R cible >100 dans le diagramme	Conclusion
Normal		l,r,m	l, z, m,	r, l,m				<b>l, m</b>
Normal mal produit			j, b	j				j ?
Normale mais mieux produit ou moins robuste ?			r					<b>r</b>
Robuste	f v, r,m,b	n,b,v,z,f	v, n	b,n,v,z	p,			<b>n,v</b> <b>z, b ?</b>
Robuste mal produit				t	k			
Compensation pour une partie		p,ch	p,d	p	n	s		<b>P,</b> <b>Ch ?</b>
Fragilité		s	s	s	ch	f		<b>s</b>
Compensation sur fragilité		d, g			d	p		<b>d,g</b>
Substitution possible sur une grande distorsion		k	K, g	g		k	ch	<b>ch</b> <b>g ? k ?</b>
Fragile	/t/, /ch/, /j/,/d/. ?		f	d				
Fragile ? bien produit		s	s	s				<b>s</b>
Fragile mal produit			t					
Non défini		t,j	ch	Ch, f,k				t,f

Tableau n°5.4.1 informations recueillies par les différentes méthodes utilisées pour déterminer la contribution de chaque consonne pour l'intelligibilité. En gras, les consonnes pour lesquelles les résultats des différentes approches convergent. Suivies d'un point d'interrogation, les consonnes pour lesquelles la conclusion est discutable, soit par manque d'arguments, soit par doute sur le rang. Les consonnes robustes dans le sens d'une résistance à la distorsion (robustesse à la production) sont /r/ et /m/ et à moindre degré le //.

Les consonnes robustes dans le sens d'une résistance avec une bonne perception malgré une altération de la production (robustesse à la perception) sont le /n/ et /v/.



Certaines consonnes sont bien identifiées alors qu'elles sont produites d'une manière différente : /ch/ et /d/. Les consonnes /p/ et /g/ ont un profil similaire mais la mauvaise performance du système à leur égard discrédite ce résultat. Le rang médian obtenu par le système pour ces deux consonnes peut être en rapport avec la difficulté du système pour certaines occlusives.

La consonne repérée comme fragile est le /s/. L'association d'une fragilité associée à une amélioration de la perception à partir d'un seuil de distorsion paraît possible notamment sur les consonnes : /p/, /g/.

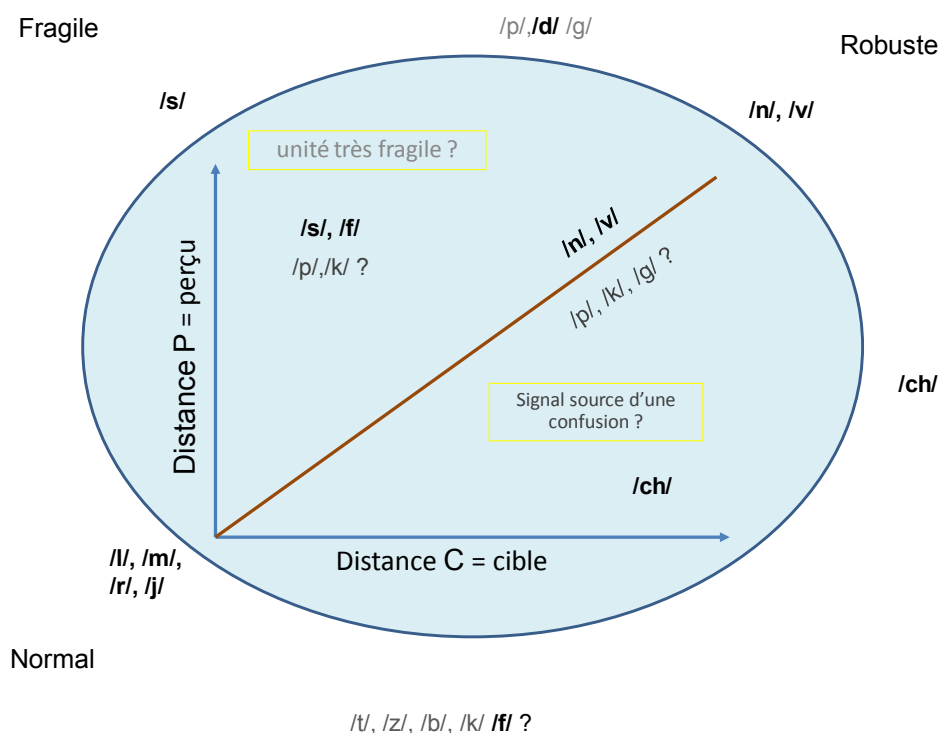


Figure n°5.4.1. Schéma synthétique des résultats de la recherche de la robustesse à la perception des consonnes. En gras, les consonnes pour lesquelles les résultats des différentes approches convergent.

## 5.4.2. LA ROBUSTESSE AUX DISTORSIONS

### 5.4.2.1. LES TROUBLES DE LA PRODUCTION DE LA PAROLE DANS LES CANCERS DE LA CAVITE BUCCALE ET DE L'OROPHARYNX

Dans notre travail, le déficit mesuré en termes de taux de perception juste varie entre 29 et 85%. Les consonnes atteintes sont dans un ordre décroissant: /k/, /g/, /ch/, /d/, /t/. Les mieux conservées et donc robustes dans ce contexte pathologique sont le /r/, /v/, /f/, /n/.

Si l'on compare ces résultats avec ceux disponibles dans la littérature dans le même cadre pathologique, les difficultés sont souvent décrites par rapport à la topographie linguale. La pointe est plutôt associée à des erreurs sur / t /, / d / et / s /, la partie arrière de la langue avec des sons tels que / k /, / g / (Terai & Shimahara 2004).

On remarquera cependant, qu'une étude finlandaise (Knuuttila et al 1999) retrouve un taux pour une tâche de perception juste pour les consonnes de 50% à 92%. Les consonnes concernées sont systématiquement le /r/ et le /l/ et dans plus de la moitié des cas le /j/, /t/, /d/,/k/. Les auteurs soulignent que le /s/ est toujours dégradé mais comme le finois ne possède qu'une seule sifflante, cela induit peu de confusions avec ce phonème. En comparaison aux résultats de notre travail en français, le /r/ est au contraire mieux conservé. On notera ici l'importance de la différence entre les langues. En français (Delattre 1946), le /r/ peut être produit de nombreuses manières en référence à une seule référence phonologique alors qu'en finlandais seul le /r/ vibré est reconnu. Autre différence, la présence d'autres sifflantes en français.

Bressmann et al (2004) dans une étude dont le but est d'analyser le rapport entre l'intelligibilité de la parole et le déficit de mobilité linguale après glossectomie partielle, retrouve un taux de perception juste des consonnes plus élevé, à 71,25%. Les erreurs sont plus fréquentes sur le trait de lieu (31,25%), mais également sur le voisement (30,2%). Une erreur a été identifiée comme une articulation compensatoire impliquant une consonne qui ne fait pas partie du système phonologique allemand (/ d / a été perçue comme /ð/mais interprété comme /d/). Cette observation concorde avec la mise en évidence dans notre travail d'une compensation par l'attribution du phonème /d/ dans une zone à fort degré de déviance de production.

Ces deux observations sont en faveur :

1. De particularités liées à la structure des différentes langues qui, alors que le trouble de production est probablement similaire en terme de déficit structurel, n'aura pas la même conséquence en terme de déficit fonctionnel, dès lors que la perception est dépendante du système phonologique de la langue native.
2. De la possibilité de réallouer une identité phonétique à une production déviante. Cette possibilité peut être interprétée comme un réajustement dans l'espace articuloire en fonction des possibilités structurelles restantes.

Ce résultat justifie une réflexion sur la capacité des analyses perceptives à décrire les déficits structurels dans la mesure où le niveau phonologique permet des compensations dans le domaine de la perception de la parole.

#### 5.4.2.2. COMPARAISON AVEC LES AUTRES SOURCES D'ALTERATION DU SIGNAL

---

Pour mener la discussion suivante, nous avons essayé de nous concentrer sur les études produites à partir de corpus de langue française.

##### 5.4.2.2.1. LA COMPARAISON PAR RAPPORT AUX AUTRES SOURCES D'ALTERATION D'ORIGINE PATHOLOGIQUE

La comparaison par rapport aux autres sources d'altération d'origine pathologique confirme la spécificité de l'impact des déficits anatomiques sur la production de la parole.

Dans le même cadre pathologique, nos résultats sont concordants avec ceux de la littérature quand à la prédominance d'un déficit sur le lieu alvéolaire si la tumeur est à une extension plus antérieure dans la cavité buccale et sur le trait de voisement quand elle a une extension postérieure dans la pharynx (Wakumoto et al 1996, Terai & Shimahara. 2004, Stelzle F et al 2011).

Le profil des erreurs est très différent quand on s'intéresse aux résections laryngées (Crevier-Buchman et al 2002). En revanche, elle se rapproche du profil des paralysies linguales isolées (Gatignol et al 2003) ou d'atteintes plus diffuses comme dans les scléroses latérales amyotrophiques (Robert et al 2006).

Hors contexte pathologique, dans le chant, certaines consonnes résistent mieux que d'autres à l'élévation de la fréquence fondamentale. La constrictive palatale « j », arrive en tête, puis ce sont les consonnes nasales suivies de la consonne liquide /l/ et des constrictives palatales /ch/ et /z/. Viennent ensuite les occlusives palatales puis labiales puis /r/ et enfin les occlusives et les constrictives dentales /d/ /v/ /t/ /f/ /s/ (Scotto Di Carlo 2005).

Les traits de lieu sont les plus fragiles dans le cadre des « troubles » de la production de la parole suivi par le trait interrompu. Le trait de voisement déjà asymétrique en français, est plus résistant.

#### 5.4.2.2. LA COMPARAISON AVEC DES CONDITIONS DE TRANSMISSION ALTERANT LA QUALITE DU SIGNAL DE PAROLE

Dans le bruit les traits de lieux sont les plus altérés, les traits de voisement sont les mieux préservés (Hazan & Simpson 2008).

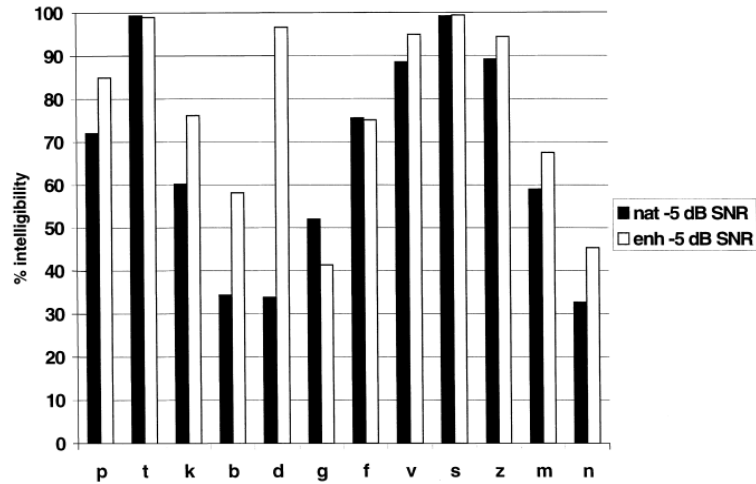


Figure n°5.4.2 : Identification correcte dans des conditions de bruits à -5dB SNR pour un stimulus naturel et un stimulus renforcé sur le plan acoustique d'après Hazan & Simpson (2008)

Meyer (2010) a étudié le taux de réponse juste en fonction de la distance entre la source de production et la réception du signal de parole. Les consonnes les plus résistantes sont les constrictives palatales non voisées /ch/, /s/ / j / puis les sonantes /l/ et /r/, les nasales et enfin les occlusives. La moins résistante est le /f/ suivi par l' autre constrictive labiodentale /v/ et le /z/.

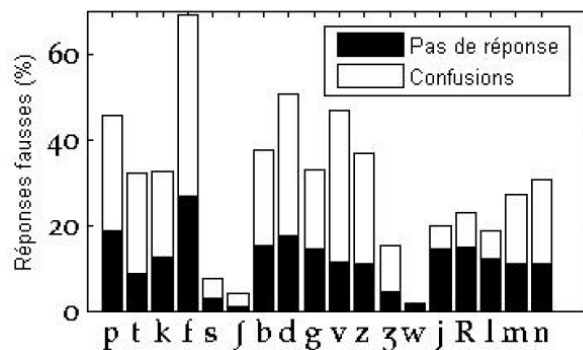


Figure n°5.4.3. : Distribution des erreurs pour chacune des consonnes

La diversité des altérations que subit le signal de parole dans des conditions naturelles (naturelles en opposition à la manipulation artificielle des signaux) est associée à des profils différents d'impact sur la perception de la parole.

	Glossectomies Etude actuelle	Chant Scotto di Carlo 2005	Bruit (-5 dB NSR) Hazan et simpson 1998	Distance Meyer et al 2010
b				
p				
d				
t				
k				
g				
v				
f				
z				
s				
j				
ch				
m				
n				
r				
l				

Tableau n°5.4 2 : Résultats de quelques travaux sur les troubles de la production de langue française. En rouge rouge foncé, les consonnes les moins bien perçues, en vert foncé les phonèmes les mieux perçus

Que cela soit dans un cadre pathologique ou non, ce recueil d'information ne met pas en évidence de résistance « universelle » de certaines catégories phonétiques ou de certains traits pour les troubles de la production (tableau n°5.4 2). Elle montre seulement que certains traits sont plus fragiles et ou résistants dans certaines conditions de production comme le voisement dans le bruit (Miller & Nicely 1955, Benki 2003). Ceci ne permet pas de dégager des stratégies correctives issues directement de l'observation de l'impact sur la perception des troubles et justifie la réflexion sur la recherche d'une méthode permettant de connaître, pour un système linguistique donné, la contribution à l'intelligibilité de la parole de ses unités.

---

#### 5.4.2. LES METHODES POUR EVALUER L'IMPACT DU DEFICIT DE PRODUCTION SUR LA PERCEPTION DE LA PAROLE

---

Les méthodes utilisées pour quantifier le déficit de la parole ont été exposées au chapitre 2 (section 2.4). Elles reposent couramment sur des tests d'intelligibilité à partir de mots isolés ou de mots dans des phrases ou de phonèmes

##### 5.4.2.1. LES LIMITES DES TAUX DE PERCEPTION JUSTE

---

Au cours de ce travail nous avons, lors du test de perception, classé les consonnes en fonction des résultats des différentes méthodes à partir de la matrice de confusion.

Nous ne parlerons pas ici, de l'utilisation du taux de perception juste pour évaluer le déficit de production, mais de son impact sur la fonction de communication à son plus bas niveau, celui de la perception du code du langage parlé.

La capacité de l'interlocuteur à percevoir le signal en y repérant les unités linguistiques permettant le processus de décodage devient naturellement l'objet de l'étude.

Trois écueils se dressent dans cette démarche.

1. La production de la parole est toujours associée à des « imperfections » de production impactant la perception.
2. Les auditeurs ont des aptitudes différentes à décoder des signaux altérés.
3. En fonction des types de distorsion, la survenue de confusions s'ajoute aux unités non identifiées pour complexifier le processus d'identification.

En ce qui concerne le premier point, les taux de perception juste inférieurs à 85% pour certains phonèmes produits par les locuteurs témoins nous ont conduit à rechercher les matrices de confusion de locuteurs « normaux » dans des conditions « normales ». Elles sont rares et nous n'en avons pas retrouvé en français de matrice comparable . De plus, celles que nous avons pu consulter en langue anglaise, révèlent que les taux de perception des inter dentales, plus spécifiques à cette langue sont faibles (Miller & Nicely 1955, Allen 2005). Ces études sont souvent faites après avoir « épuré » le corpus en supprimant les locuteurs ou les stimuli qui donnent de « mauvais » scores. Pourtant la perte sur l'intelligibilité doit être relativisée par rapport à la situation naturelle et non par rapport à une situation virtuelle où tous les locuteurs parleraient idéalement bien. D'ailleurs dans notre travail, il semble exister une relation entre la distribution des taux d'intelligibilité des témoins et celui des sujets pathologiques qui méritera des investigations supplémentaires.

Le second point justifie le nombre d'auditeurs nécessaires pour « neutraliser » l'effet de la grande variabilité de la perception dans la parole. L'utilisation des systèmes automatiques pourrait remédier à cet écueil. Cependant l'utilisation de tel système est récente dans les troubles de la production de la parole et le choix des algorithmes aux différentes étapes du traitement n'est pas sans conséquence (Van Nuffelen 2009). De plus, les situations de communication en interaction restent difficiles à modéliser pour les systèmes automatiques (Oudeyer 2003).

Le troisième, encore plus complexe, est la traduction par des variables « simples » du contenu de la matrice de confusion. Le calcul d'un indice de transfert d'information initié par Miller et Nicely (1955) a été développé (Sagi & Svirsky 2008). Il consiste à quantifier en nombre de traits respectés dans la matrice de confusion, l'information correctement transférée. Cependant, cet indice n'inclut pas la difficulté pour identifier une unité représentée par la présence de nombreuses confusions. Il est certain que l'utilisation d'une méthode reposant sur un choix forcé est un biais puisqu'elle oblige à donner un sens à tout signal quel que soit le degré d'altération. Cependant la prépondérance de certaines confusions (comme ici le /v/ et le /f/) constitue un véritable facteur aggravant.

#### 5.4.2.2. LES PERSPECTIVES POUR DES EVALUATIONS PLUS PERTINENTES

---

Le calcul des taux de confusions est une source d'information complémentaire. Mais, en pratique, ils ne peuvent être utilisés qu'en comparaison aux taux de perception juste.

Le calcul d'un indice de confusion (Kayasith & Theeramunkong 2009) a été proposé dans la littérature, dans le but de mesurer la sévérité du trouble en réponse à la question : comment le signal peut être interprété comme un mot erroné? Il repose sur la détermination pour toutes les confusions réalisées lors de l'écoute d'une cible, de la distance entre la cible et le perçu en utilisant un système automatique. Contrairement à notre approche, le système n'utilise pas les algorithmes classiques des systèmes de reconnaissance. Les étapes du traitement comprennent une extraction de traits, puis une comparaison de trait, une représentation en cluster et enfin le calcul de l'indice. Les résultats en terme de prédiction de sévérité sont meilleurs que ceux obtenus par un test d'intelligibilité ou d'articulation d'après les auteurs. Cette approche paraît prometteuse mais n'exclut pas l'intérêt d'une méthode ne recourant pas à des systèmes automatiques et donc applicable à partir d'un test de perception.

Une solution simple de pondération du taux de perception juste a été proposée sous la forme de la F mesure, méthode utilisée pour évaluer des systèmes de reconnaissance automatique de la parole. Cette F mesure est issue des travaux sur la pertinence des recherches de documentation. Elles sont plus simple à appréhender que les notions de sensibilité et spécificité. Dans notre travail, cet indice s'est avéré cohérent avec les autres indices. Il mérite que nous poursuivions les travaux pour l'évaluer en tant qu'indice prédictif de l'impact sur la communication à ces différents niveaux (de l'intelligibilité

en terme d'identification de mots à la compréhensibilité en termes d'interprétation de message voire de l'analyse de conversation).

---

### 5.4.3. LA ROBUSTESSE A LA PERCEPTION

---

Notre travail avait pour objectif de chercher une méthode pour déterminer si les unités minimales de la langue contribuaient de manière différente à l'intelligibilité. Pour nous affranchir du problème de la fréquence relative de l'unité linguistique, nous avons utilisé le degré d'altération de la production de cette unité et recherché si certaines unités continuaient ou non à être identifiées malgré cette altération. Si nous avons pu mener à bien la validation de cette approche, l'interprétation des résultats est contrainte par les choix méthodologiques que cette expérience a imposés.

#### 5.4.3.1. LES LIMITES D'INTERPRETATION DE NOS RESULTATS

---

Nos résultats ne sont interprétables que dans la mesure où les unités linguistiques présentent des degrés d'altération suffisamment étendus dans le corpus. Or, pour certaines consonnes comme le /f/, leur bonne prononciation ne permet pas de conclure. Ce problème pourra être résolu par l'incrémentation du corpus avec des productions pathologiques provenant de locuteurs altérant ces consonnes.

Dans notre travail, lors de la segmentation du signal pour le test de perception par le jury, les frontières ont été déterminées à la fin de la voyelle précédent la syllabe cible et à la fin de la voyelle de cette syllabe. Ainsi les informations comprises dans la transition précédant la consonne n'ont pas été incluses dans les stimuli. C'est une des raisons que nous pouvons évoquer pour expliquer le faible taux de reconnaissance par le système automatique. Cela explique peut être, aussi, la faiblesse du taux de perception juste pour certaines consonnes chez les locuteurs témoins, notamment le locuteur PRM.

L'influence de la fréquence de l'unité linguistique, qu'il s'agisse de la consonne, de la syllabe ou du mot et de l'environnement vocalique n'a pas été spécifiquement étudiée dans ce travail. Pour en discuter, nous avons calculé, à partir des données disponibles dans lexic.org, la fréquence dans la langue française des consonnes de notre corpus dans leur contexte syllabique.



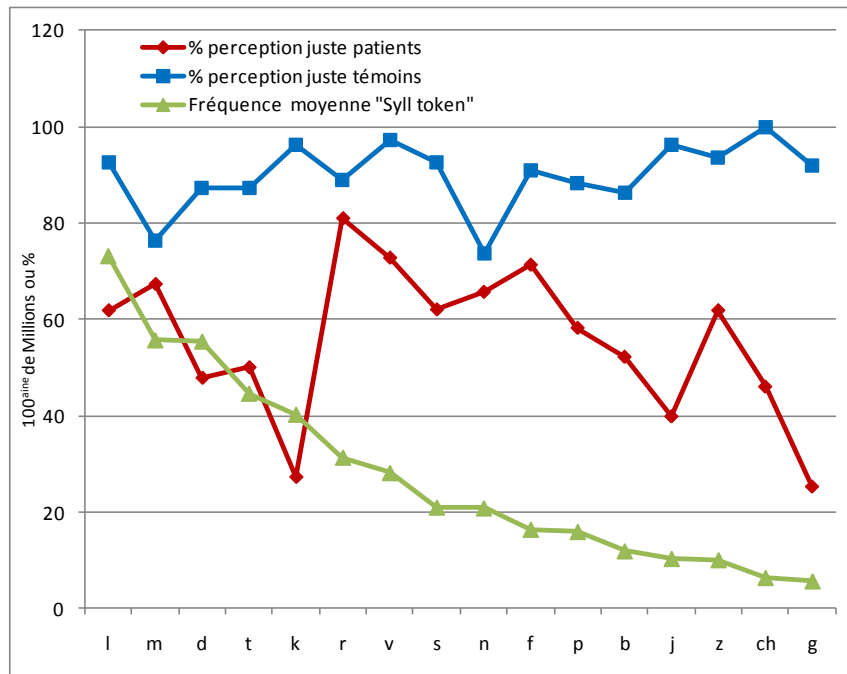


Figure n°5.4.3 . Taux de perception juste et fréquence moyenne « Syll token » des consonnes par ordre décroissant de la fréquence des consonnes

La comparaison des courbes dessinées par les taux de perception juste pour les locuteurs et la fréquence des consonnes (figure n°5.4.3), n'exclut pas une incidence de la fréquence pour certaines consonnes.

De même, l'analyse de l'influence de la voyelle de la syllabe, sur les taux de perception par le jury et le taux de reconnaissance par le système, révèle des différences (figure n°5.4.4).

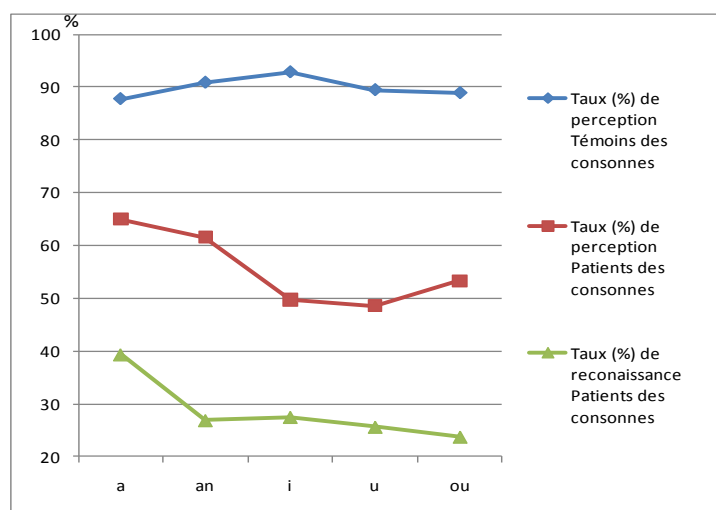


Figure n°5.4.4. Taux en Pourcentage de perception juste par les auditeurs et le système de reconnaissance des consonnes en fonction du contexte vocalique de la syllabe

Alors que pour les témoins le contexte /i/ est associé à une meilleure perception de la consonne, pour les patients le contexte /a/ est maximalisant quelle que soit la méthode. Il existe des discordances entre les résultats obtenus à partir du test perceptif et ceux obtenus à partir du système automatique pour le /an/ et le /ou/. Un effet maximalisant du /a/ a déjà été décrit dans le bruit (Hazan & Simpson 2008). Les moins bons résultats dans un environnement vocalique comme le /u/ sont classiques.

En prenant en considération les précautions justifiées par les remarques ci-dessus, les résultats obtenus par la méthode proposée sont suffisamment cohérents pour conclure à une différence de contribution entre les consonnes dans le système phonologique du français. Ainsi /n/ et /v/ semble contribuer de manière importante à l'intelligibilité contrairement au /s/. Une contribution « normale » ou moyenne concernerait les liquides /l/, /r/, la nasale /m/ et la fricative palatale /j/. Les phénomènes de compensation et les limites du système automatique rendent plus délicate l'interprétation pour les autres unités.

#### 5.4.3.2. LA RECHERCHE D'INFORMATIONS POUVANT CORROBORER NOS RESULTATS

---

Nous n'avons pas retrouvé dans la littérature d'étude similaire, ni de travail ayant le même objectif. De ce fait, nous avons cherché si les consonnes que nous avons repérées, étaient remarquables dans d'autres situations.

Ainsi, nous n'avons pas retrouvé de liens évidents avec :

- La fréquence des consonnes en français (www.Lexique.org, New B et Pallier C),
- La fréquence des consonnes dans les langues du monde enregistrées dans la base de données UPSID (UCLA Phonological Standard Inventory Data, 451 langues, I. Maddieson et K. Precoda, Escudier & Schwartz 2000),
- Les consonnes communes aux différentes langues du monde,
- L'ordre d'acquisitions des phonèmes en français.

Seules les consonnes /n/ et /v/ sont particulièrement fréquentes en français, mais le /s/ est alors en 4<sup>ème</sup> position (figure n°5.4.4).

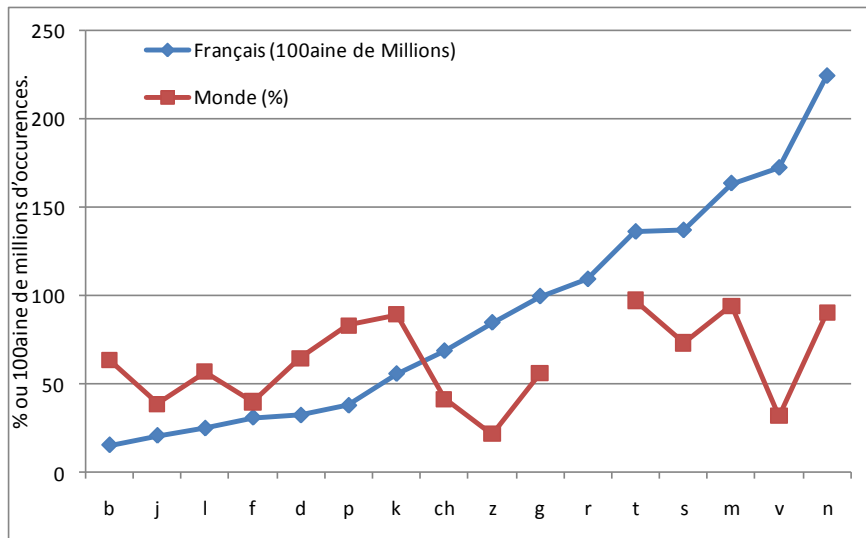


Figure n°5.4.4 : Courbe de tendance avec le nombre d'occurrences pour le français par centaine de millions et le pourcentage relatif des consonnes dans les langues du monde d'après la base de donnée UPSID (A noter le polymorphisme du /r/ n'est pas compatible avec un calcul simple de sa fréquence)

Nous avons également regardé si les phonèmes robustes à la perception présentaient un profil particulier en richesse de combinaison de traits (Dell 1985, Durand 2005) ou en richesse d'indices acoustiques (Delattre 1946), sans succès.

## 5.5. CONCLUSION

Ce travail est en faveur d'une contribution inégale des différentes consonnes pour l'intelligibilité de la parole. La méthode utilisée paraît pertinente et faisable. Mais, une analyse plus approfondie des résultats obtenus par le système automatique de reconnaissance est indispensable avant de continuer à caractériser les consonnes par cette méthode.

*Il m'est arrivé dans les sciences ce qui arriverait à un homme qui, s'étant levé de grand matin, attendrait avec impatience que l'aurore et le jour vinssent dissiper les ténèbres ; et qui, lorsque le soleil aurait paru, se trouverait aveuglé par l'éclat de ses rayons.*

*Goethe*  
*(Maximes et réflexions (3ème partie), trad. Sigismond Sklover, p.80, Brockhaus et Avenarius, 1842)*

## 6. CONCLUSIONS ET PERSPECTIVES

---

6. Conclusions et perspectives	287
6.1. Implication sur la définition des différents concepts.....	289
6.1.1. Intelligibilité et sévérité	289
6.1.2. Essai de définition des notions de sévérité, intelligibilité et compréhensibilité dans le cadre conceptuel de la CIF	290
6.1.3. Implications pour la mesure de l'intelligibilité et perspectives pour l'évaluation dans les TPPP	294
6.1.3.1. indice du déficit structurel	294
6.1.3.2. Indices d'impact sur l'activité et la participation communicative	295
6.1.3.2.1. Echanger de l'information	295
6.1.3.2.2. agir sur l'autre	297
6.1.3.2.3. Entretenir l'interaction	297
6.2. Implications sur la détermination des unités linguistiques minimales pertinentes pour les TPPP .....	298
6.2.1. L'intrication des différentes informations comprises dans le signal	299
6.2.2. Les compensations observées dans les TPPP	299
6.2.3. La notion de robustesse et les unités minimales perceptives	300



## 6.1. IMPLICATION SUR LA DEFINITION DES DIFFERENTS CONCEPTS

---

De manière générale, les concepts de sévérité, intelligibilité, compréhensibilité, largement utilisés dans la littérature souffrent d'un manque de précision dans leur définition.

---

### 6.1.1. INTELLIGIBILITE ET SEVERITE

---

Plusieurs arguments plaident en faveur d'une différence entre sévérité et intelligibilité, remettant en question le principe de l'évaluation de la sévérité par une mesure de la dégradation de l'intelligibilité. Différentes études relèvent l'absence de corrélation entre l'impression générale d'intelligibilité ou de sévérité et les taux d'identification d'unités linguistiques, principalement les mots et les phrases (Walshe 2008). Cette observation est probablement majorée par l'absence de standardisation de la consigne. Par exemple Dagenais et al (2006) demandent aux juges « comment cotez-vous la capacité de cette personne à parler ? », alors que Walshe (2008) demande « A quel point est ce facile ou difficile de comprendre la parole de cette personne ? ». Cette différence de consigne induit des processus cognitifs différents chez une partie des auditeurs, si un jugement de compréhension est sollicité, comme le démontre les résultats du chapitre 4.

Cette remise en question conduit à s'interroger sur la notion de sévérité dans les TPPP.

Les résultats du chapitre 3 sont en faveur d'une notion de sévérité définie par le degré de déviance par rapport à la normalité. Mais dans la quête d'un indice de sévérité, il est fondamental de déterminer « l'objet » déviant. En l'absence de définition de la normalité, nous citerons Van Riper & Emerik (1984 p 111) : *"Speech is impaired when it deviates so far from the speech of other people that it calls attention to itself, interferes with communication, or provokes distress in the speaker or the listener"*. Dans un contexte d'écoute, cette normalité se concentre sur le signal acoustique sans préjuger de son contenu. Notamment, cela ne justifie pas de concentrer l'évaluation sur l'information linguistique véhiculée par le signal. Une consigne, faisant référence au degré d'altération du signal de parole, paraît donc indiquée.

Or, un certain nombre d'auteurs utilisent la sévérité en référence au déficit du trouble de production, en faisant l'hypothèse que le déficit de production de la parole est mesurable par la perception de la parole produite (Darley 1969b, Kent et al 1989, Auzou 2007). D'autres considèrent plutôt la sévérité à travers l'impact de l'altération de la production de la parole sur la communication en utilisant les

mêmes méthodes (Yorston & Beukelman 1980, 1999, Hustadt & Beukelman 2002). Souvent, l'objet n'est pas explicité. Cependant cette conception est en pleine évolution (Hustadt 2008, Kim et al 2009). Les résultats du chapitre 5 démontrent que l'altération du signal n'est pas systématiquement associée à une perte de la reconnaissance d'une unité linguistique. Les auditeurs peuvent attribuer un signifié à un son présentant une forte distorsion. Le mécanisme de cette compensation reste hypothétique : repérage d'autres indices suffisamment stables, restauration favorisée par un ou des indices résiduels ? Il existe aussi des arguments pour penser qu'en fonction de la structure phonologique d'une langue, certains sons s'ils ne sont pas exposés à des confusions et s'ils sont suffisamment spécifiques de leur catégorie seront plus « résistants » perceptivement. Ces dernières observations peuvent contribuer aux difficultés pour mettre en évidence les corrélations entre les échelles de déficits moteurs et le taux d'identification de mots comme cela a été récemment rapporté par De Letter et al 2005 dans la maladie de Parkinson.

Ainsi les résultats de notre travail démontrent un niveau de compensation dans les TPPP qui peut être qualifié de phonologique. Ils incitent à revisiter la notion de la sévérité et le principe de son évaluation par la mesure de l'intelligibilité.

### 6.1.2. ESSAI DE DEFINITION DES NOTIONS DE SEVERITE, INTELLIGIBILITE ET COMPREHENSIBILITE DANS LE CADRE CONCEPTUEL DE LA CIF

En prenant comme cadre théorique les concepts développés par l'Organisation Mondiale de la Santé avec le modèle de Wood (Classification internationale du handicap : CIH) puis la classification internationale du fonctionnement (CIF) (figure n°6.2.1, Fougeyrollas 2002), la distribution de ces différentes notions en fonction des différents niveaux de fonctionnement peut déboucher sur une organisation plus adaptée de la prise en charge et des évaluations préalables.

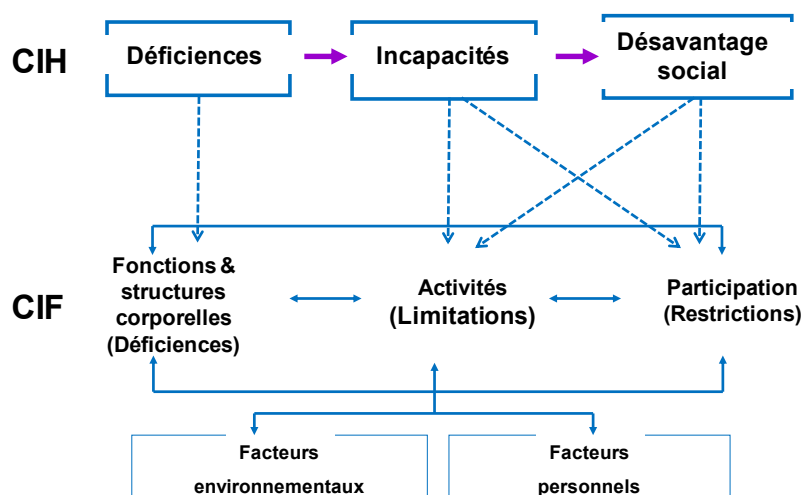


Figure n°6.2.1 : De la classification internationale du handicap (CIH) à la classification internationale du fonctionnement (ICF) – Source = Centre Technique National d'Etudes et de Recherches sur les Handicaps et les Inadaptations

En abordant le sujet par la déficience, la sévérité peut donc faire référence :

- aux structures corporelles, c'est à dire dans le contexte des TPPPP, aux structures cérébrales ou aux structures du carrefour aérodigestif ;
- à la fonction de production de la parole « parler » (les fonctions étant ici définies comme fonctions physiologiques des systèmes organiques).

Si ces deux notions étaient indifférenciées dans la classification internationale du handicap, elles ont été explicitement séparées dans la classification internationale du fonctionnement. Ceci renforce la conclusion du paragraphe précédent quant à l'importance de lever l'ambiguïté qui demeure dans la littérature, sur l'objet mesuré, lorsque l'on utilise le terme de sévérité dans les TPPP.

Cependant, si la fonction physiologique renvoie à l'acte de parler et donc à des modèles linguistiques, quelle fonction linguistique de cet acte faut-il utiliser ? Pour essayer de débrouiller cette question, nous allons l'aborder par l'autre extrémité du schéma de la CIF : le concept de participation.

Plusieurs auteurs s'intéressent à l'intégration de la prise en charge des TPPP dans le cadre de l'ICF – (Walshe 2009, Yorkston et al 2007). Yorkston et al (2008) est en train de développer des instruments de mesure de la participation « communicative ». En adoptant son point de vue, les limitations d'activité et les restrictions de participation peuvent exister le long d'un continuum opposé de telle manière qu'ils représentent des concepts complètement distincts. Les activités peuvent être pensées comme l'exécution d'une tâche de parole indépendamment de tout contexte personnel et environnemental, alors que la participation est l'utilisation de la parole dans le contexte de la « vraie vie » avec ses besoins et ses buts.

Ainsi, le domaine de la participation renvoie à des modèles de communication différents en fonction de la réalité de chacun, que l'on s'intéresse au profil des activités d'un individu ou aux préoccupations d'un organisme de tutelle. La diversité de ces modèles est illustrée dans le chapitre 2, section 2.3.1 avec en fonction des domaines d'application, des modèles sociométriques, interactionnistes ou situationnels. Par contre, l'activité communicative dans la mesure où elle entre dans le cadre d'un « environnement standardisé » devrait être plus facile à mesurer. Mais encore faut-il pouvoir définir cet environnement standardisé et la nature de l'activité communicative ! Cela sous-entend également un continuum entre la fonction de parler et les fonctions du langage dans les activités.



Que cela soit en termes d'activité ou de fonction, les modèles théoriques concernant les actes de langage sont pluriels laissant différentes options pour mesurer l'activité en ciblant un ou plusieurs actes. A la suite du travail de la catégorisation libre, le modèle de Jakobson (section 2.3.1, figure n°1.3.3) nous paraît le plus adapté bien que d'autres modèles plus pragmatiques (issus des actes du langage de Searle(1972) ou des modèles de conversation (Bloch 2009) soient également envisageables.

Jusqu'à présent l'utilisation du modèle de Shannon (section 2.2.1.2. figure n°1.2.4) était prépondérante dans la mesure où il est utilisé pour évaluer le déficit structurel sur l'hypothèse que le déficit de la fonction est le reflet du déficit structurel. Cette modélisation d'une théorie de l'information est alors parfaitement adaptée même si, dans son application avec des jurys d'écoute à la place du récepteur, est introduit le problème de l'accès au sens. En revanche, ce choix, comme l'a démontré Hustad (2008), n'est pas adapté à la mesure de l'impact sur la « compréhensibilité » d'un message en pré-requis à une interaction. Alors que Hustad (2008) fait référence au cadre théorique de la compréhension des énoncés défini par Kintsch (1983), nous pensons que le choix du modèle de communication de référence est un pré-requis à la poursuite de la réflexion. L'acte de langage retenu après l'étude de la catégorisation libre au chapitre 3 vise le destinataire et relève de la fonction conative, fonction qui répond à l'acte illocutoire de la théorie des actes du langage de Searle (1972).

Ainsi, en termes d'activité, le but est de pouvoir être compris lors des interactions avec les autres, constat décliné sous la forme de différentes théories en fonction des disciplines (Searle 1972, Sperber & Wilson 1989, Lindlom 1990). La standardisation de cette situation dans un objectif d'évaluation pour les TPPP doit savoir s'affranchir de ces différentes théories pour choisir une situation répondant aux besoins élémentaires de tout être vivant comme par exemple l'interaction suivante :

-« J'ai faim »

- « Que veux tu manger ? » ou « Veux tu manger » en fonction de la situation (délai par rapport au repas par exemple).

C'est-à-dire une conversation avec un seul interlocuteur dont l'assertion du premier locuteur va générer un énoncé de la part du second locuteur suivant un tour de parole. Mais ici, encore, plusieurs choix sont possibles en fonction de ou des étapes du processus de compréhension retenu, entre la perception de la parole et l'interprétation du discours en interaction.

Pour terminer cette réflexion et proposer des définitions adaptés aux TPPP, nous avons repris :

- le schéma synthétique du chapitre 2, section 2.3.2 (figure n°1.3.7)

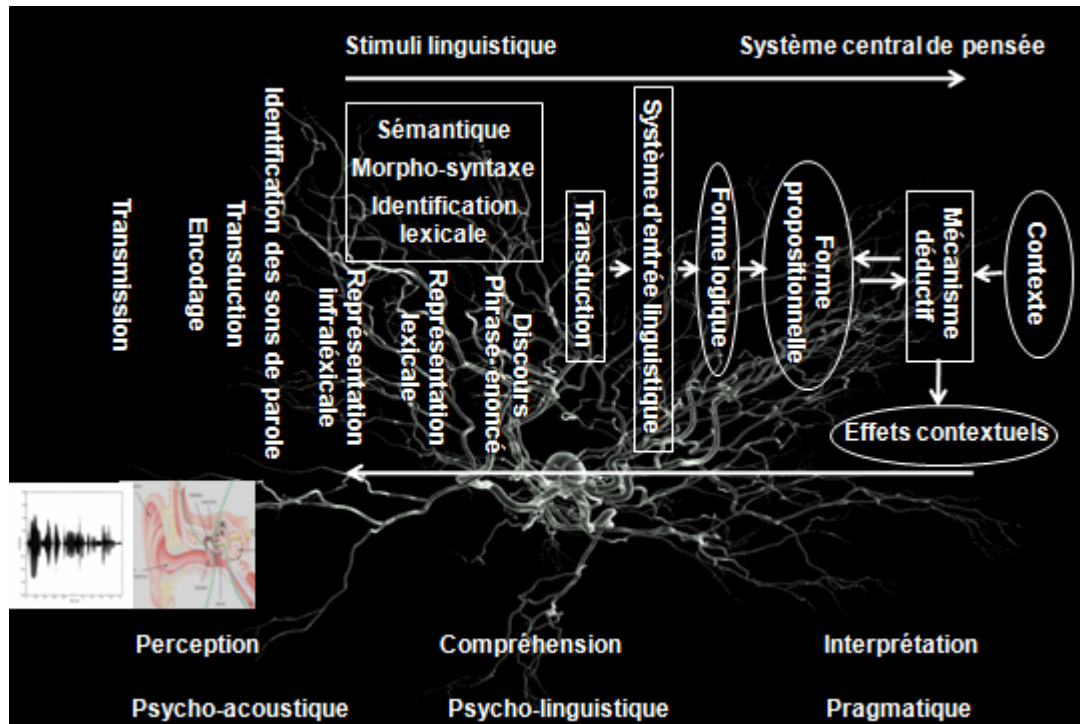


Figure n°1.3.7. Schéma des différentes étapes de la perception du son à l'interprétation des énoncés

- les définitions de la norme Française (NF EN 60849 - traduction CEI 60849 :1998) promulguée puis modifiée en 2004 (AFNOR 2004 NF EN 60268-16) pour les systèmes de sonorisation et de diffusion utilisés pour provoquer la mise en alerte,

- la proposition de Meyer 2005 dans le cadre d'un travail sur l'intelligibilité des phrases sifflées sur le parallèle entre les différents niveaux de représentation perceptive et les différentes notions définies par les instances de normalisations (figure n°6.2.2).

Sensation sonore	Perception phonétique	Perception des mots	Perception des phrases
Audibilité	Netteté	Intelligibilité	Compréhension

Figure n°6.2.2 : Parallèle entre les notions d'audibilité, netteté, intelligibilité et compréhension et le processus de perception d'après Meyer 2005

Les définitions suivantes sont ainsi proposées :

- *Intelligibilité : décodage des unités linguistiques élémentaires jusqu'au niveau lexical en référence à la perception de la parole. Cette définition recouvre la netteté définie par Meyer 2005, notion proche de la clarté définie dans la norme AFNOR 2004.*

- *Compréhensibilité : accès au sens d'un énoncé (phrase au minimum) en tant que stimuli linguistique à partir du décodage acoustico-phonétique et de la construction du stimuli comprenant les effets top down.*
- *Interprétabilité : accès au sens dans une conception plus pragmatique*

L'Audibilité reste une notion indépendante de la nature linguistique potentielle du signal. Il s'agit de la: propriété des sons leur permettant d'être entendus parmi d'autres sons.

---

### 6.1.3. IMPLICATIONS POUR LA MESURE DE L'INTELLIGIBILITE ET PERSPECTIVES POUR L'EVALUATION DANS LES TPPP

---

En retenant de la discussion précédente que la mesure de l'intelligibilité doit être précisée en fonction de son objectif, nous allons proposer des solutions en fonction des buts que nous avons déterminés dans le cadre de la CIF.

Ceci est d'autant plus justifié qu'une certaine ambiguïté est entretenue par des publications importantes sur le sujet. Ainsi, le travail de Kim Y et al (2009), qui démontre la valeur prédictive de la mesure du F2 et débouche sur un modèle statistique de la pente de F2 en relation avec la sévérité de la dysarthrie, est basé sur la mesure de F2 et une mesure direct d'amplitude d'intelligibilité et donc sur la perception du signal et pas sur le déficit structurel.

#### 6.1.3.1. INDICE DU DEFICIT STRUCTUREL

---

Il s'agit de déterminer un indice qui permettrait de mesurer la dégradation du signal produit par le locuteur. Le fait que le signal véhicule des informations non linguistiques mais fondamentales pour la communication (entre autre l'identité du locuteur) conduit à s'affranchir ici de la recherche d'indices plus spécifiques à la substance sous-tendant la nature symbolique des unités linguistiques.

De nature exclusivement acoustique, la pertinence d'un tel indice devrait être validée :

1/ par des corrélations suffisantes avec les scores de déficit adaptés au contexte pathologique, plus ou moins focalisés si nécessaire sur les structures anatomiques impliquées dans la production de la parole,

2/ en terme de prédictibilité sur les différents niveaux d'accès au sens, c'est à dire sur l'intelligibilité, la compréhensibilité et l'interprétation du message.

Dans la littérature, de nombreux paramètres sont disponibles. Une première étape comprendrait la recherche dans les autres domaines d'applications, des paramètres susceptibles de contribuer à la construction d'un tel indice. Dans ce travail nous avons déjà repéré un certain nombre de paramètres au chapitre 2, section 2.5.3. De plus, le chapitre 3 confirme l'importance d'y intégrer une dimension reflétant la vitesse de la production. La confrontation aux indices déjà validés dans les domaines pathologiques comme le DSI (Dysphonia Severity Index, Wuyts et al 2000) pour la dysphonie et les paramètres relatifs à l'intelligibilité dans les dysarthries serait la seconde étape.

En restant dans un modèle correspondant à la théorie de l'information de Shannon, cet indice pourrait être une traduction du déficit moteur, décelant des altérations infra cliniques et infra linguistiques. Il pourrait permettre de caractériser la relation entre le déficit moteur et la production de l'onde sonore, puis de connaître la prédictibilité de ce déficit sur la perception de la parole. L'intérêt d'un tel indice pourrait être évalué pour les TPPP au sens large, y compris les dysphonies. Il permettrait de quantifier l'altération pour la recherche et d'assurer la surveillance de l'évolution en pratique clinique.

En attendant, au vu de nos résultats, nous déconseillons l'utilisation de l'impression générale d'intelligibilité mesurée par une échelle d'amplitude directe au profit d'une impression générale de sévérité. La consigne ne doit pas contenir de référence à la compréhension du message mais définir la sévérité comme le degré d'altération globale du signal sonore.

### 6.1.3.2. INDICES D'IMPACT SUR L'ACTIVITE ET LA PARTICIPATION COMMUNICATIVE

---

L'évaluation de cet impact peut être mesurée à différents niveaux des représentations du processus de compréhension. Mais aussi, comme le souligne Kerbrat-Orecchioni (1970), « Parler, c'est sans doute échanger des informations; mais c'est aussi effectuer un acte régi par des règles précises qui prétend transformer la situation du récepteur et modifier son système de croyances et/ou son attitude comportementale; corrélativement, comprendre un énoncé c'est comprendre son contenu informationnel, sa visée pragmatique [...]. »

#### 6.1.3.2.1. ECHANGER DE L'INFORMATION

Echanger de l'information est donc le premier niveau et correspond à la bonne transmission du code, mais à travers la perception par un locuteur. Dans une situation standardisée, il s'agirait d'un auditeur « idéal » en termes de capacité de perception et de contrôle de celle-ci. C'est le domaine des tests

d'intelligibilité. Cette situation renvoie au modèle de Shanon mais avec toute la complexité des tests utilisant un jury d'écoute.

S'affranchir d'un jury en utilisant un système automatique avec une normalisation des techniques pour assurer un traitement satisfaisant des résultats permettrait de garantir une reproductibilité des tests.

Cette approche est en cours dans plusieurs laboratoires (Pouchoulin et al 2008, Maier et al 2009, Van Nuffelen et al 2009). Cela réduirait aussi, la problématique de la tâche qui, quelque soit le choix méthodologique pour assurer l'indentification de l'unité linguistique choisie, du fait de la consigne peut être une source de biais par elle-même.

En clinique, les tests reposent soit sur la transcription soit sur un choix forcé (cf chapitre 2 section 2.4.2)

La transcription même en l'absence de TPPP conduit à un taux d'erreur minimal non négligeable. Le choix forcé impose obligatoirement une réponse qui peut avoir été aléatoire. Cela a donné lieu, depuis les travaux de Miller et Nicely (1955) à une réflexion sur le traitement des données issues des matrices de confusion. En effet, le taux de perception juste ne recouvre qu'une partie du problème de l'identification des unités linguistiques par le locuteur. La quantité et la distribution des confusions (si elle n'est pas également répartie) interviennent sur la capacité du locuteur à la parole dégradée comme l'a décrit Liss (2007) (Chapitre 2, section 2.3.2.1.5). Ainsi une mesure pondérée par le taux de confusion pourrait être plus pertinente comme nous le proposons au chapitre 5 avec la F mesure. Ce type de mesure serait à valider en testant leur prédictibilité sur les déficits de communication à des niveaux supérieurs. L'utilisation du degré de vraisemblance à partir d'un système automatique est également une piste.

Demeurera toujours le choix du corpus, pour lequel les résultats de notre travail sont insuffisants pour faire des propositions. La poursuite de la recherche des unités linguistiques les plus contributives pour l'intelligibilité aidera à définir les critères déterminants la création de corpus.

Pour différencier, dans une situation de test, une évaluation de l'activité communicative d'une évaluation de la participation communicative, un changement de statut de l'auditeur est nécessaire. Ainsi, suite à cette remarque de Hustad (2008), au sujet de la discordance entre test d'intelligibilité et compréhension : « Perhaps most noteworthy is that generalization of an intelligibility score to conclusions about listeners ability to comprehend a speaker would like be inaccurate », il paraît fondamental de lever certaines ambiguïtés sur le rôle de l'auditeur. Tester la capacité de quelqu'un à

comprendre une parole détériorée est différent d'utiliser quelqu'un pour mesurer le degré d'altération de la production de la parole. Dans un test d'intelligibilité, dont l'objectif est de mesurer l'impact sur l'activité communicative, la standardisation correspond à la deuxième situation, alors que ce même test utilisé pour évaluer la participation, correspondrait à la première situation.

#### 6.1.3.2.2. AGIR SUR L'AUTRE

A un niveau au dessus : l'évaluation de la capacité à agir sur l'autre relève des tests de compréhensibilité comme ceux utilisés par Beukelman puis Hustad, (Hustad KC, Beukelman DR. 2002; ;Hustad 2008). Suite au travail de Hustad (2008), l'intelligibilité ne paraît pas être un bon indicateur de comment les auditeurs sont capables d'interpréter le message du locuteur. Cet auteur suggère qu'une théorie compréhensive de l'impact de la dysarthrie sur les capacités de communication doit interroger les processus utilisés par auditeurs aux multiples niveaux de représentation. La méthode utilisée repose sur l'évaluation de la compréhension d'un texte narratif par des questions avec un cotation de 0 à 2 en fonction du niveau d'inférence. Cependant, cette méthode adaptée des tests de compréhension de lecture est difficile à utiliser en clinique en raison de la mémorisation du ou des textes par les examinateurs.

Les conditions de standardisation d'une méthode d'évaluation en français sont en cours avec le travail de Fontan et al (2011). Ce travail devrait aboutir à une procédure d'évaluation de la compréhensibilité de la parole implémentée dans un logiciel EloKanz praticable en milieu clinique. Le principe repose sur l'enregistrement d'une consigne simple consistant à décrire le placement d'un objet par rapport à un autre puis dans un deuxième temps à demander à un auditeur de réaliser la consigne enregistrée. Cette procédure pourrait faire l'objet d'une normalisation.

#### 6.1.3.2.3. ENTRETENIR L'INTERACTION

L'étude des interactions dans le cadre des TPPP est très récente et son intégration dans le cadre d'une évaluation clinique serait novatrice (Bloch & Wilkinson 2009). La méthode décrite est celle de l'analyse de conversation telle qu'elle a été décrite au chapitre 2, section 2.3.2.2.2. Déclinée en fonction du profil de chaque patient en utilisant un modèle de communication situationnel comme celui proposé par Mucchielli (2000) (figure n°1.3.6), Cette évaluation pourrait être appliquée en fonction des besoins spécifiques des patients en se référant à d'autres modèles de communication (cf section 2.3.1).

Une tentative de synthèse de cette réflexion est représentée par la figure n°6.2.3

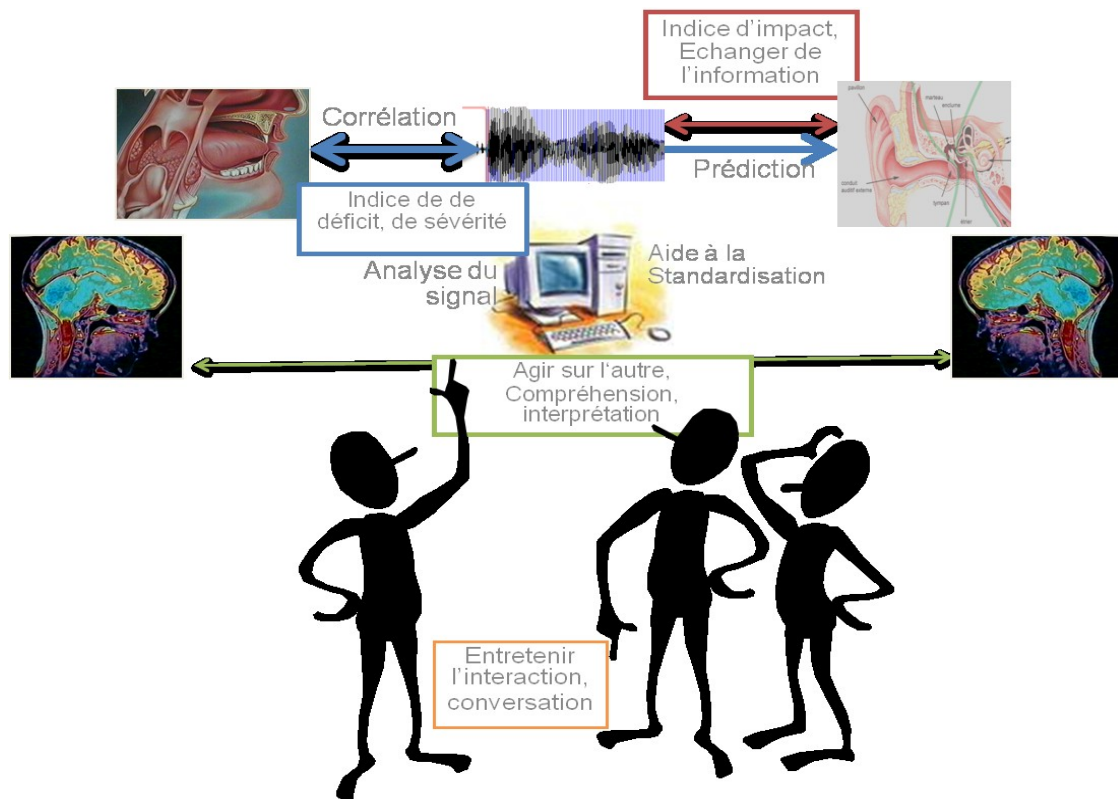


Figure n°6.2.3 : Illustration des différents niveaux d'évaluation des TPPP dans le contexte de la CIF

## 6.2. IMPLICATIONS SUR LA DETERMINATION DES UNITES LINGUISTIQUES MINIMALES PERTINENTES POUR LES TPPP

Dans un éditorial consacré aux perspectives de recherche pour le contrôle moteur de la parole et ses troubles, Kent (2000) posait la question suivante : « What is (are) the primary variables(s) controlled by the nervous system ? An answer to this question is central to theories of speech motor control and also can be highly important in assessing and treating disorders of speech motor control. ». Il listait ensuite les différentes propositions de primitives contrôlées par le système nerveux que sont les cibles acoustiques, les paramètres aérodynamiques, les positions de structures anatomiques isolées, la configuration générale du tractus vocal, et le degré de contraction musculaire. Il souligne dans cet article le fait que finalement, plus d'une variable pouvaient être nécessaire pour comprendre le contrôle neurologique de la parole et que la robustesse de la parole pouvait dériver en partie de la flexibilité du locuteur à choisir parmi ces variables. Notre travail ne permet pas de contribuer à la réponse à cette question. Mais il peut aider à préciser la problématique.

---

### 6.2.1. L'INTRICATION DES DIFFÉRENTES INFORMATIONS COMPRISES DANS LE SIGNAL

---

Les résultats du travail de catégorisation libre (chapitre 3) révèlent que même dans le cadre des TPPP, la voix reste un élément fondamental par les informations qu'elle véhicule sur l'identité et les caractéristiques du locuteur. Cette observation corrobore différentes sources d'information contribuant à expliquer le lien entre la voix et la compréhension. Goldinger (1996) apporte des arguments pour un stockage par les auditeurs de traces de voix spécifiques dans la mémoire épisodique. Ces souvenirs détaillés aident les auditeurs à percevoir la parole produite par la personne dans les interactions suivantes. Il est possible que ce type d'informations soit particulièrement utile quand la parole a quelque chose d'inhabituel ou est dégradée. Meunier (1999, 2010) associe les caractéristiques du locuteur aux sources de variation de la parole. Ces variations, inhérentes à la production de la parole, ne sont pas interprétées, mais attendues par les auditeurs. Hallé et al (1998) ont ainsi montré que les auditeurs identifiaient un phonème en fonction de leurs attentes plutôt que d'après les réalisations produites. De ce fait, l'absence de ces variations peut entraîner un coût perceptif. Cette augmentation du coût perceptif peut contribuer à l'échec du processus de compréhension par la réduction des capacités d'identification. Elle peut être mise en parallèle avec la perte du caractère naturel de la parole décrit dans les dysarthries. Elle laisse supposer la nécessité de changer de stratégie ou de mobiliser des ressources supplémentaires de la part de l'auditeur face à la flexibilité du locuteur à choisir parmi plus d'une variable au niveau du contrôle neurologique de la parole.

---

### 6.2.2. LES COMPENSATIONS OBSERVEES DANS LES TPPP

---

Les résultats du chapitre 5 démontrent qu'une altération importante du signal n'est pas systématiquement associée à une perte de la reconnaissance d'une unité linguistique et que, en plus des compensations articulatoires, existent des compensations phonologiques.

Les compensations articulatoires ont été particulièrement bien décrites dans les fentes palatines (Borrel 1931 pour le français). Dans ce cadre pathologique (insuffisance vélopharyngée dans le contexte des pathologies cranio-faciales), une théorie a été développée par Warren et al (1990) proposant la variable pression de l'air pharyngée comme une des primitives contrôlées par le système nerveux. Pour Kent (2000), cette théorie a une valeur unificatrice pour de nombreuses observations cliniques, incluant la sélection par le locuteur de compensations articulatoires. Ainsi, même si nous



n'avons pas dans notre travail cherché à les identifier, de nombreux arguments attestent leur existences dans le contexte de modification structurel y compris les chirurgies de la cavité buccale et de l'oropharynx (Couprie 1992, Fauvet 2006)

Les compensations phonologiques sont moins souvent décrites (cf chapitre 5, section 5.4.2). Leur description en parallèle de compensations articulatoires, contribue à la théorie Hypo et Hyper de Lindlom (1990, Hypo and Hyper Speech theory) et à son modèle adaptatif de la production et de la perception de la parole. Ce modèle ne postule plus un invariant physique. La communication parlée doit répondre à deux exigences : le besoin d'intelligibilité pour l'auditeur et le principe d'économie articulatoire pour le locuteur. La situation de communication d'un patient présentant un TPPP, conduit dans cette perspective le locuteur à produire son énoncé selon un principe de variabilité adaptative de façon à ce que les deux exigences soient respectées avec la particularité de solliciter les capacité d'adaptation phonologiques de l'auditeur pour l'intelligibilité en balance avec les adaptations les plus économiques possibles pour le locuteur.

Le processus qui équilibrerait cette balance pour un patient reste «mystérieux » même pour les thérapeutes. Une analyse des compensations phonologiques à travers la théorie dispersion-focalisation (Schwartz et al 1997b) pour les voyelles permettraient de valider une stratégie déterminée par une dispersion maximale dans le nouvel espace créé par la modification de la structure, assurant les contrastes suffisants pour une intelligibilité optimale. Les enregistrements de notre corpus et la segmentation isolant les voyelles devrait permettre ce travail.

---

### 6.2.3. LA NOTION DE ROBUSTESSE ET LES UNITES MINIMALES PERCEPTIVES

---

Les discussions autour de la notion d'unité minimale de perception se sont traduites dans notre travail (chapitre 5) par un choix méthodologique imparfait concernant les segments de production du corpus et les unités linguistiques pour l'analyse perceptive. Ainsi nous avons opté pour une production « factice » de mots dans la mesure où il s'agit de mots monosyllabiques dont une faible proportion correspond à des non-mots. Nous les avons ensuite traité comme des syllabes et avons ciblé notre analyse sur la consonne de la syllabe, un phonème. Ce choix a été en grande partie conditionné par la pratique clinique. Or, l'utilisation du phonème dans le traitement perceptif est discutée. Parmi les arguments justifiant ce questionnement nous citerons Meunier C (2010) : « la plupart des locuteurs

d'une langue sont capables d'identifier les phonèmes présents au sein d'un mot. Toutefois, les recherches sur l'acquisition du langage semblent montrer que cette capacité ne serait pas effective avant l'âge de 6 ou 7 ans. [...] La syllabe s'avère moins variable et présente des caractéristiques plus robustes que celles du phonème. [...] Cependant, cette question fait toujours débat car certains travaux ne répliquent pas les effets de l'avantage de la syllabe sur le phonème » (cf chapitre 2, section 2.3.2.1.2).

Mais nos résultats, malgré le fait qu'ils soient incomplets, démontrent une robustesse perceptive distribuée différemment sur les consonnes, en faveur d'une contribution inégale des phonèmes pour l'intelligibilité dans un système linguistique donné. Au-delà du fait que la poursuite de ce travail permettra d'optimiser les bilans en pratique clinique et de déterminer les cibles les plus efficaces pour la réhabilitation, ce résultat interroge sur la raison d'une différence de contribution sur l'intelligibilité dans un système. Ainsi, si une unité minimale de perception existait, une faible différence de contribution entre les différents éléments la composant ne serait elle pas un argument ? Cela pourrait être notre contribution la plus directe à la question.



## **7. ANNEXES**

---

7. Annexes	303
7.1. Annexes en rapport avec le chapitre 3 (catégorisation libre).....	305
7.1.1. Les textes utilisés pour les stimuli	305
7.1.1.1. Pour les TPPP : Pierrot par Guy de MAUPASSANT	305
7.1.1.2. Pour les dysphonies : La chèvre de M. Seguin	305
7.1.2. Annexe Logiciel TCL	305
7.1.2.1. Présentation de TCL-LabX	305
7.1.2.2. Les quatre étapes d'une expérimentation	306
7.1.3. Annexe EVA	307
7.1.3.1. Magnétophone numérique	307
7.1.3.2. Analyse de la prosodie	308
7.1.3.2.1. Mode d'emploi	308
7.1.3.2.2. Un exemple d'analyse : sujet FOM	308
7.1.4. Annexes en rapport avec la section 3.2.3. : l'étude 1	310
7.1.4.1. La description des locuteurs	310
7.1.4.2. Exemples de commentaires	310
7.1.4.2.1. Exemples de commentaires de 3 experts pour chaque stimulus	311
7.1.4.2.2. Exemples de commentaires de 3 NON experts	312
7.1.4.2.3. Exemples de commentaires DE 3 proches	313
7.1.4.3. Arbres annotés avec les résultats de l'analyse acoustique des paramètres prosodiques	314
7.1.4.3.1. Arbres pour les experts	314
7.1.4.3.2. Arbres pour les non experts	315
7.1.4.3.3. Arbres pour les proches	315
7.1.5. Annexes en rapport avec la section 3.3.3. : Etude 2	316
7.1.5.1. Description des stimuli vocaux	316
7.1.5.1.1. Les paramètres acoustiques	316
7.1.5.1.1.1. Les paramètres de fréquences	316
7.1.5.1.1.1.1. en fonction du genre	316
7.1.5.1.1.1.2. Les paramètres de fréquences en fonction de la pathologie	316
7.1.5.1.1.1.3. Les paramètres de fréquences en fonction de la sévérité (G)	317
7.1.5.1.1.2. Les paramètres d'intensité	317
7.1.5.1.1.3. Les paramètres des pauses	318
7.1.5.1.1.3.1. en fonction du genre	318
7.1.5.1.1.3.2. en fonction de la pathologie	318
7.1.5.1.1.3.3. en fonction de la sévérité (G)	319
7.1.5.2. Répartition des paramètres acoustiques en fonction des classes	319
7.1.5.2.1. Pour les experts	319
7.1.5.2.1.1. Répartition de la fréquence selon les classes des arbres	319
7.1.5.2.1.2. Répartition de la dynamique de l'intensité selon les classes des arbres	319
7.1.5.2.1.3. Répartition des pauses et données temporelles selon les classes des arbres	320
7.1.5.2.2. Pour les non experts	320
7.1.5.2.2.1. Répartition de la fréquence selon les classes des arbres pour les non experts	320
7.1.5.2.2.2. Répartition de la dynamique de l'intensité selon les classes des arbres pour les non experts	321
7.1.5.2.2.3. Répartition des pauses et données temporelles selon les classes des arbres	321
7.1.5.3. Extraits des commentaires pour les dysphonies	322
7.2. Annexes du chapitre 4 (Intelligibilité et sévérité) .....	323
7.2.1. Les Phrases du Hint utilisées pour le corpus de stimuli différents	323
7.2.2. Tableaux avec les valeurs des corrélations entre les paramètres des différentes tâches.	324
7.3 Annexes en rapport le chapitre 5 (Robustesse et contribution à l'intelligibilité). .....	324
7.3.1. Annexes en rapport avec la section.5.3.	324
7.3.1.1. Annexe en rapport avec la section 5.3.1.1 : Description de la technique chirurgicale	325
7.3.1.2. Annexe en rapport avec la section 5.3.2 : tableaux contenant les valeurs qui ont permis la construction des représentations graphiques	325
7.3.3 Annexes en rapport avec la section 5.3.3 concernant le rAng médian et la description de la répartition sur la bissectrice	332
7.3.2 Annexes en rapport avec la section 5.4 : la fréquence moyenne des consonnes	335

## 7.1. ANNEXES EN RAPPORT AVEC LE CHAPITRE 3 (CATEGORISATION LIBRE)

---

### 7.1.1. LES TEXTES UTILISES POUR LES STIMULI

---

#### 7.1.1.1. POUR LES TPPP : *PIERROT* PAR GUY DE MAUPASSANT

---

Conte parut le 09 octobre 1882, dans le journal *Le Gaulois*

Les deux femmes habitaient une petite maison à volets verts, le long d'une route, en Normandie, au centre du pays de Caux.

Comme elles possédaient, devant l'habitation, un étroit jardin, elles cultivaient quelques légumes.

Or, une nuit, on lui vola une douzaine d'oignons.

Dès que Rose s'aperçut du larcin, elle courut prévenir Madame, qui descendit en jupe de laine.

Ce fut une désolation et une terreur. On avait volé, volé Mme Lefèvre ! Donc, on volait dans le pays, puis on pouvait revenir.

Et les deux femmes effarées contemplaient les traces de pas, bavardaient, supposaient des choses : "Tenez, ils ont passé par là. Ils ont mis leurs pieds sur le mur ; ils ont sauté dans la plate-bande".

Et elles s'épouvantaient pour l'avenir. Comment dormir tranquilles maintenant !

#### 7.1.1.2. POUR LES DYSPHONIES : LA CHEVRE DE M. SEGUIN

---

La Chèvre de monsieur Seguin est une des nouvelles issues des Lettres de mon moulin d'Alphonse Daudet.

L'extrait suivant a été enregistré pour le corpus utilisé dans l'expérience de catégorisation libre pour les dysphonies.

« M. Seguin n'avait jamais eu de bonheur avec ses chèvres. Il les perdait toutes de la même façon : Un beau matin, elles cassaient leur corde, s'en allaient dans la montagne, et là-haut le loup les mangeait. Ni les caresses de leur maître, ni la peur du loup, rien ne les retenait. »

### 7.1.2. ANNEXE LOGICIEL TCL

---

#### 7.1.2.1. PRESENTATION DE TCL-LABX

---

TCL-LabX est un progiciel spécifiquement développé afin de préparer rapidement et facilement des expérimentations de catégorisation libre et de recueillir les données finales sous une forme susceptible d'être traitée par des logiciels de statistiques (comme Partition, AddTree).

L'informatisation de ce type de tests permet de produire des expérimentations répliquables et identiques pour chaque sujet et par conséquent fiables en terme de résultats. Ce qui suit s'appuie sur la version 0.3.0X de TCL-2 LabX. Les évolutions futures de ce progiciel peuvent amener certaines modifications mineures à ces explications. Une page web est dédiée à TCL-LabX. On peut y télécharger les dernières versions du logiciel et la documentation attenante.<http://w3.lordat.univ-tlse2.fr/p-gaillard/tcl-labx/>

## 7.1.2.2. LES QUATRE ETAPES D'UNE EXPERIMENTATION

---

### a. Préparation de l'expérimentation

La première étape consiste à rassembler les différents stimuli et éventuellement « prototypes induits ». Cette étape concerne largement le cadre théorique de l'expérimentation en cours (bruits, sons musicaux, parole, etc...). Le nombre et la durée de chacun des stimuli dépend directement des orientations de l'expérimentation en cours : TCL-LabX n'est pas limité dans ce domaine. Une fois ces éléments disponibles, il suffit de lancer TCL-LabX et de choisir l'option « File / New Config... », puis de donner un nom à cette configuration, et enfin de l'enregistrer dans un nouveau dossier dans un lieu au choix. Automatiquement, TCL-LabX créera trois dossiers : « sounds », « protosounds » et « primingsound ». Ces dossiers correspondent respectivement au dossier où l'on glissera les fichiers audios utilisés comme stimuli, les fichiers audios utilisés comme prototypes induits (douze maximums) et enfin le fichier audio utilisé comme amorce auditive. Dans le dossier « protosounds », seuls les douze premiers fichiers seront considérés. Dans le dossier « primingsound », seul le premier fichier sera considéré. Dans chacun des dossiers, les fichiers sont évoqués par ordre alphabétique. De ce fait, si l'on veut un ordre particulier dans les stimuli et/ou dans les sons prototypiques induits, il faudra nommer les fichiers en conséquence. Enfin, par mesure de compatibilité multiplateforme, on s'efforcera de donner des noms de fichiers courts, sans espaces et sans caractère accentué. Comme expliqué dans la documentation disponible avec l'application, on pourra également régler des options telles que la taille et la couleur des boutons, ou encore le texte de l'amorce ou de la consigne. Enfin, il sera possible de cacher le fond de l'écran de l'ordinateur pendant le test.

### b. Passation de l'expérimentation

Lorsque l'expérimentation est configurée et sauvegardée pour un usage ultérieur, l'expérimentateur n'a plus qu'à lancer le test via la commande « Experiment / Run experiment » et rentrer un nom de sujet. Pour les mêmes raisons que précédemment, il est recommandé de rentrer un nom sans espaces et sans caractères accentués. Le participant doit se tenir prêt pour la consigne ou l'amorce, puis pour la classification. Lorsqu'il est prêt, l'expérimentateur clique sur le bouton « Go » et s'efface devant le participant. Comme expliqué sur l'écran des classifications, le participant peut écouter les sons (en double-cliquant sur les boutons) et déplacer les boutons à l'écran (en maintenant un clic gauche et en déplaçant la souris). Nous rappelons que chacune de ces actions est enregistrée au moment où elle a lieu, et pourra faire l'objet d'une étude statistique sur le comportement du participant durant le test lui-même.

### c. Définition des classes / catégories

Lorsque le sujet a terminé sa classification, la troisième étape importante consiste à définir chacun des éléments de chaque classe et d'y associer un commentaire. Si l'expérimentateur ne souhaite pas recueillir de commentaire, il pourra laisser la fenêtre des commentaires vide. A ce stade de l'expérimentation, il est toujours possible de réécouter les sons, mais il n'est plus possible de les déplacer à l'écran. Avant chaque définition des membres d'une classe, il est recommandé de faire réécouter les sons associés à cette classe au participant afin qu'il se remémore les caractéristiques de cette classe. Une fois que le participant se souvient bien de la classe en cours de définition, l'expérimentateur choisit « Classes / Start Class Definition ». Il clique ensuite sur chacun des membres d'une classe en collaboration avec le participant. Chaque bouton cliqué prend alors une couleur différente. Il convient de faire très attention, car cette opération de sélection est irréversible : un bouton sélectionné ne peut plus être désélectionné. L'expérimentateur choisit enfin la commande « Classes / End Class Definition » : une petite fenêtre de saisie de texte

s'affiche alors. L'expérimentateur ou le participant lui-même peut rentrer librement les commentaires associés à cette classe. Un clic sur le bouton « OK » en bas de cette fenêtre de saisie marque la fin de la définition de la classe en cours.

L'expérimentateur recommence l'opération pour chacune des classes faites par le participant.

#### d. Sauvegarde et traitements des données

Enfin, lorsque l'ensemble des commentaires est saisi, l'expérimentateur sauvegarde la totalité des données (cf. plus haut pour la définition des données de sortie) via la commande du bouton « Save Data & Close Experiment ». Les données sauvegardées le seront automatiquement dans un dossier créé par l'application, dans le répertoire du fichier de configuration (au même niveau que les dossiers « Sounds », « protosounds » et « primingsound ») et nommé « results ». Si le bouton n'est pas disponible, cela signifie que des fichiers du même nom que ceux proposés automatiquement par TCL-LabX existent déjà (souvent parce que le sujet a le même nom qu'un sujet précédent). Il suffit alors à l'expérimentateur de modifier chacun des quatre noms de fichier et de vérifier à nouveau ces noms via la commande du bouton « Verify Files Names (again) ». Si aucun fichier préalablement enregistré ne porte le nouveau nom, la commande « Save Data & Close Experiment » devient disponible. Chacun des fichiers de données ainsi sauvegardés pourra alors faire l'objet soit d'un traitement individuel (cas du fichier contenant les actions du participant) ou collectif (cas des données des classes, dont les différents fichiers individuels pourront être concaténés afin d'être traités par des outils de statistiques et de représentations graphiques, en arbres de similarité par exemple). Nous évoquerons brièvement ce dernier mode de traitement dans la section suivante.

---

### 7.1.3. ANNEXE EVA

---

#### 7.1.3.1. MAGNETOPHONE NUMERIQUE

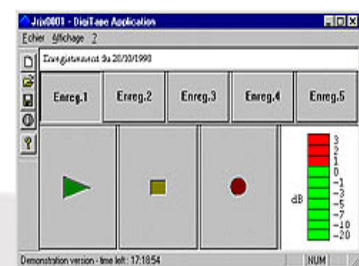
---



#### Magnétophone numérique (pour EVA2, DIANA)

**Assiste l'évaluation perceptuelle des troubles de la voix et de la parole, particulièrement dans un suivi.**

Le sujet est enregistré lors de différents examens. Le manipulateur dispose de 5 pistes pour réécouter les enregistrements successifs et évaluer la progression avec une approche comparative.





## 7.1.3.2. ANALYSE DE LA PROSODIE

### 7.1.3.2.1. MODE D'EMPLOI

EVA 2

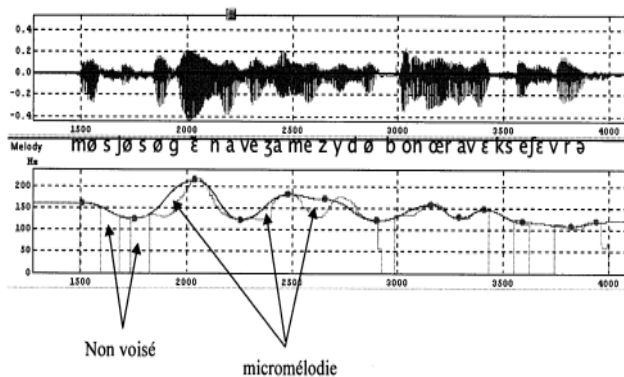
EVA 2

N

#### Note technique

#### Modélisation de la mélodie

La Modélisation de la mélodie est fondée sur la technique MOMEL<sup>\*</sup>. Les points cible correspondent aux variations locales pertinentes de la courbe mélodique. Ils sont reliés par une fonction d'interpolation de type spline quadratique. Ils permettent de retrouver le profil suprasegmental dans lequel n'interviennent pas les variations micromélodiques et les interruptions dues aux segments non voisés.



L'analyse est effectuée sur les données modélisées contenues dans la zone d'observation. Sont exclues les parties présentes dans les pauses silencieuses.

<sup>\*</sup> Hirst D.J. & Espesser R. (1993), « Automatic modelling of fundamental frequency using a quadratic spline function ». Travaux de l'Institut de Phonétique d'Aix, vol.15, 71-85

Prosodie

#### Prosodie

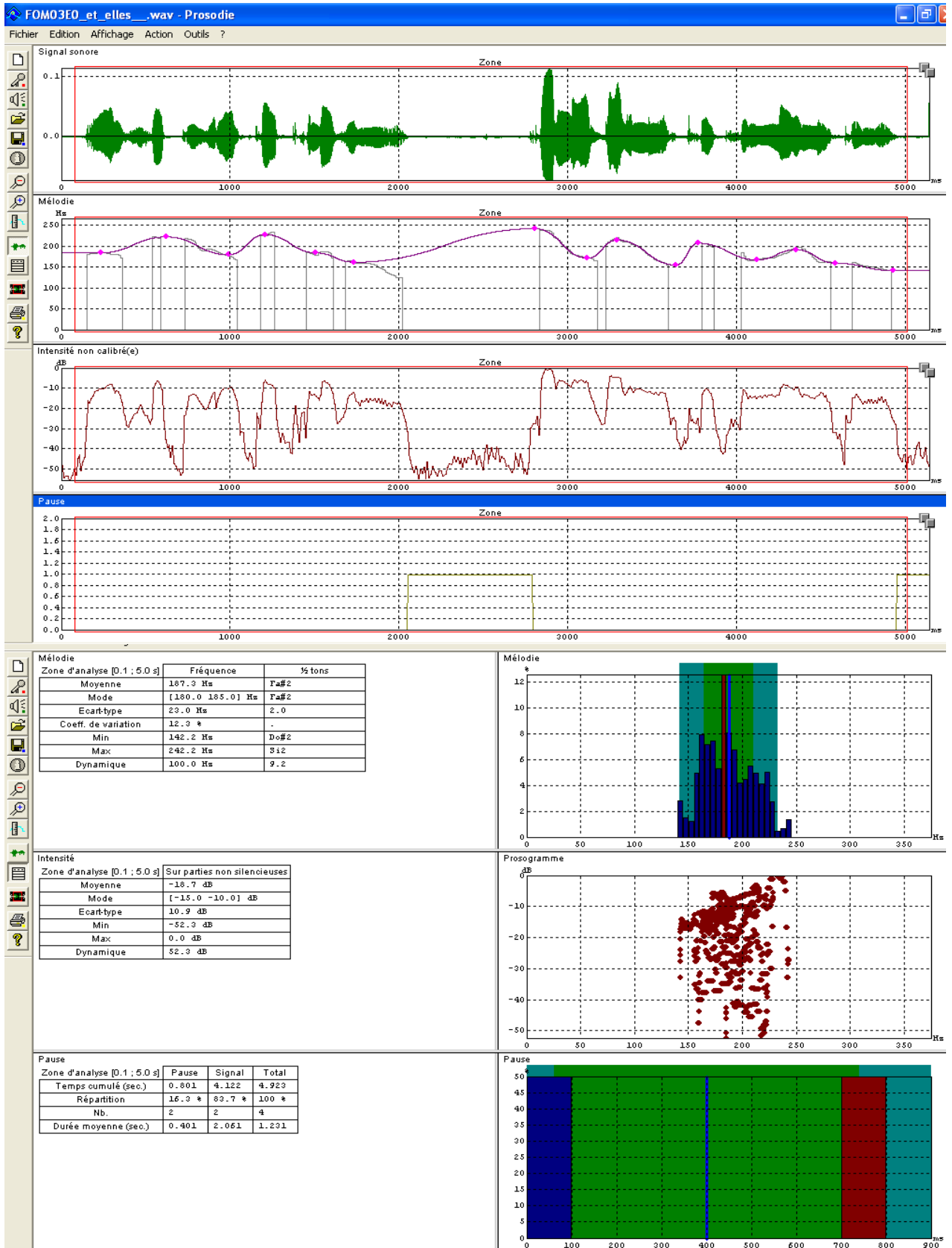


Le but de cette application est d'évaluer la prosodie, particulièrement dans le cas de troubles d'origine neurologique (ex : Parkinsoniens) ou dans le but d'évaluer l'expressivité de la parole (ex : laryngectomisés)

#### Principe

1. Le sujet est enregistré en parole spontanée, en lecture de texte...
2. Quatre fenêtres correspondent aux courbes sonores, de fréquence fondamentale (Hz), d'intensité (dB) et de détection de pauses silencieuses.
3. Il est possible d'évaluer :
  - les formes mélodiques
  - le contour d'intensité
  - la durée des énoncés, le débit syllabique
  - le placement et la quantification des pauses silencieuses

### 7.1.3.2.2. UN EXEMPLE D'ANALYSE : SUJET FOM



---

7.1.4. ANNEXES EN RAPPORT AVEC LA SECTION 3.2.3. : L'ETUDE 1

---

7.1.4.1. LA DESCRIPTION DES LOCUTEURS

---

Nom fichier	sévérité_moyenne	Altération intell_moyenne	Niveaux_intell_sev	Sexe	Groupe Patho Neuro/struct/temoin	Diagnostic	diag-court...	codes1_arbres
BEB	1	1	1	2	T	N0	N0	2TN0
CAA	1	1	1	1	T	N0	N0	1TN0
DAJ	1	1	1	2	T	N0	N0	2TN0
LEF	1	1	1	2	T	N0	N0	2TN0
ORL	1	1	1	2	T	N0	N0	2TN0
PEJ	1	1	1	1	T	N0	N0	1TN0
PIM	1	1	1	2	T	N0	N0	2TN0
PUM	1	1	1	2	T	N0	N0	2TN0
RAF	1	1	1	2	T	N0	N0	2TN0
ROC07	1	1	1	2	T	N0	N0	2TN0
SOB	1	1	1	2	T	N0	N0	2TN0
MAC17	1,4	1	2	1	N	Extrap	Extrap	1N\Extrap
FOM	1,6	1	2	2	T	N-SV	N-SV	2TN-SV
MAF	1,6	1,4	3	1	T	N-SV	N-SV	1TN-SV
SUM	2,6	2	3	2	N	Extrap	Extrap	2N\Extrap
ARA	4	3,4	3	1	S	OP	OP	1S\OP
FOY	4,8	3,2	3	1	N	Extrap	Extrap	1N\Extrap
HIM	9,6	6	3	1	N	Extrap	Extrap	1N\Extrap
CAG	9,8	6,4	3	2	S	CB-OP	CB-OP	2S\CB-OP
DAP	10	6,4	3	1	N	Extrap	Extrap	1N\Extrap
LEJ	13,8	10,6	4	1	S	CB-OP	CB-OP	1S\CB-OP
VAS	14,2	8,6	4	1	S	OP	OP	1S\OP
MAC13	14,6	11,6	4	2	S	VOILE-Fente	VoiFen	2S\VoiFen
and	15	10	4	2	N	SEP	SEP	2N\SEP
ROC05	18	12,8	4	1	N	Instab-Dystonie	InsDys	1N\InsDys
BAP	18,8	13	4	1	N	VOILE-Paralysie	VoiPar	1N\VoiPar
cof	18,8	14,4	4	2	S	CB-OP	CB-OP	2S\CB-OP
DAC	20,8	14	5	2	N	Instab-Cérébelleux	InsCer	2N\InsCer
WAL	24,6	24,4	6	1	S	CB-OP	CB-OP	1S\CB-OP
REA	26,4	25,2	6	1	S	CB-OP	CB-OP	1S\CB-OP
SAJ	27,4	27	6	1	N	Extrap	Extrap	1N\Extrap
PIP	27,6	27,4	6	1	S	CB-OP	CB-OP	1S\CB-OP
CoM	29,8	29,8	6	2	N	Extrap	Extrap	2N\Extrap

7.1.4.2. EXEMPLES DE COMMENTAIRES

---

### 7.1.4.2.1. EXEMPLES DE COMMENTAIRES DE 3 EXPERTS POUR CHAQUE STIMULUS

ARA	léger trouble de la parole sans altération de l'intelligibilité	patient avec altération du rythme dyspnéique	trouble de l'articulation
BAP	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	déperdition nasale	Nasalité
BEB	parole normale	sujets normaux prosodie moins marquée	Femmes témoins
CAA	parole normale	sujets normaux prosodie moins marquée	hommes témoins
CAG	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	déperdition nasale	trouble de l'articulation
COF	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	déperdition nasale	Nasalité
COM	important trouble de la parole responsable d'une inintelligibilité ou d'une importante altération de l'intelligibilité	Dysarthrie severe ** SAL ? ** Cérébelleux ?	Trouble du débit
DAC	trouble de la parole par altération de la prosodie (débit, rythme...) et/ou altération vocale, diminution de l'intelligibilité légère à modérée	problème de sphincter buccale et d'explosion	Trouble du débit
DAJ	parole normale	sujets normaux prosodie moins marquée	Femmes témoins
DAP	trouble de la parole par altération de la prosodie (débit, rythme...) et/ou altération vocale, diminution de l'intelligibilité légère à modérée	patient avec altération du rythme dyspnéique	Trouble vocal
FOM	léger trouble de la parole sans altération de l'intelligibilité	Asymptomatique et respect prosodie	lenteur de débit hommes et femmes
FOY	léger trouble de la parole sans altération de l'intelligibilité	patient avec altération du rythme dyspnéique	Trouble vocal
HIM	trouble de la parole par altération de la prosodie (débit, rythme...) et/ou altération vocale, diminution de l'intelligibilité légère à modérée	patient avec altération du rythme dyspnéique	Trouble vocal
LEF	parole normale	Asymptomatique et respect prosodie	Femmes témoins
LEJ	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	déperdition nasale	trouble de l'articulation
MAC 13	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	déperdition nasale	Nasalité
MAC 17	parole normale	patient avec altération du rythme dyspnéique	lenteur de débit hommes et femmes
MAF	léger trouble de la parole sans altération de l'intelligibilité	sujets normaux prosodie moins marquée	hommes témoins
ORL	parole normale	Asymptomatique et respect prosodie	Femmes témoins
PEJ	parole normale	Asymptomatique et respect prosodie	hommes témoins
PIM	parole normale	Asymptomatique et respect prosodie	Femmes témoins
PIP	important trouble de la parole responsable d'une inintelligibilité ou d'une importante altération de l'intelligibilité	Dysarthrie severe ** SAL ? ** Cérébelleux ?	Nasalité
PUM	parole normale	sujets normaux prosodie moins marquée	Femmes témoins
RAF	parole normale	sujets normaux prosodie moins marquée	Femmes témoins
REA	important trouble de la parole responsable d'une inintelligibilité ou d'une importante altération de l'intelligibilité	BPTM et patients ORL	Nasalité
ROC 05	trouble de la parole par altération de la prosodie (débit, rythme...) et/ou altération vocale, diminution de l'intelligibilité légère à modérée	déperdition nasale	Trouble du débit
ROC 07	parole normale	Asymptomatique et respect prosodie	Femmes témoins
SAJ	important trouble de la parole responsable d'une inintelligibilité ou d'une importante altération de l'intelligibilité	Dysarthrie severe ** SAL ? ** Cérébelleux ?	Trouble du débit
SOB	parole normale	Asymptomatique et respect prosodie	lenteur de débit hommes et femmes
SUM	léger trouble de la parole sans altération de l'intelligibilité	patient avec altération du rythme dyspnéique	lenteur de débit hommes et femmes
VAS	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	BPTM et patients ORL	trouble de l'articulation
WAL	trouble de l'articulation de la parole (modification ou absence de phonèmes), diminution de l'intelligibilité légère à modérée	BPTM et patients ORL	trouble de l'articulation

#### 7.1.4.2.2. EXEMPLES DE COMMENTAIRES DE 3 NON EXPERTS

AND	clair mais fatigue ou zozotement	voix âgées, chevrotantes et fatiguées	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
ARA	clair mais fatigue ou zozotement	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
BAP	difficultés pour certaines consonnes incompréhensible	voix nasonnées	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
BEB	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
CAA	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
CAG	clair mais fatigue ou zozotement	voix avec défauts d'articulation et de prononciation	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
COF	difficultés pour certaines consonnes incompréhensible	voix avec défauts d'articulation et de prononciation	Troubles articulaires, personnes n'arrivant pas à parler, à prononcer, à articuler ... le discours est incompréhensible.
COM	incompréhensible	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Troubles articulaires, personnes n'arrivant pas à parler, à prononcer, à articuler ... le discours est incompréhensible.
DAC	difficultés pour articuler	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
DAJ	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
DAP	tres clair	voix avec défauts d'articulation et de prononciation	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
FOM	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
FOY	clair mais fatigue ou zozotement	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
HIM	clair mais fatigue ou zozotement	voix avec défauts d'articulation et de prononciation	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
LEF	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
LEJ	difficultés pour certaines consonnes incompréhensible	voix avec défauts d'articulation et de prononciation	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
MAC 13	difficultés pour certaines consonnes incompréhensible	voix nasonnées	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
MAC 17	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
MAF	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
ORL	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
PEJ	tres clair	voix nasonnées	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
PIM	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
PIP	difficultés pour certaines consonnes incompréhensible	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
PUM	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques

RAF	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
REA	difficultés pour articuler	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
ROC 05	difficultés pour articuler	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
ROC 07	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
SAJ	difficultés pour articuler	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
SOB	tres clair	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
SUM	clair mais fatigue ou zozotement	voix masculines et féminines non pathologiques aux accents, prosodies, intensités différents	Voix "normales" masculines et féminines, totalement intelligibles et non pathologiques
VAS	difficultés pour certaines consonnes incompréhensible	voix âgées, chevrotantes et fatiguées	Voix pathologiques, troubles articulaires ... mais le discours reste intelligible et compréhensible.
WAL	difficultés pour articuler	voix de personnes dysarthriques, difficultés d'articulation ... ( peut-être IMC), certaines sont inintelligibles	Troubles articulaires, personnes n'arrivant pas à parler, à prononcer, à articuler ... le discours est incompréhensible.

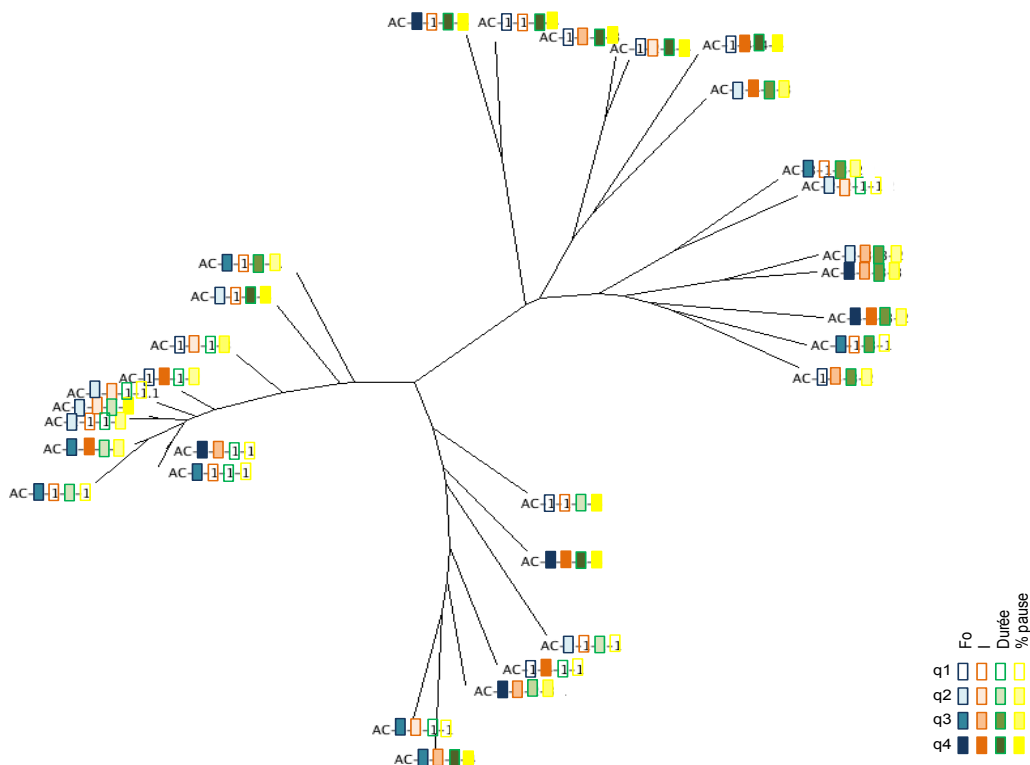
#### 7.1.4.2.3. EXEMPLES DE COMMENTAIRES DE 3 PROCHES

AND	On comprend pas	fatiguée qui arrive plus a parler	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
ARA	bien aussi	normaux	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
BAP	fort mais compréhension plus difficile	voix de personne âgés	voix ressemblants à celles des résidants - connaissance, capable de lire ou de répéter mais émotionnellement cela va altérer la compréhension
BEB	très audible et un peu plus faible comme son, comme tonalité	a peu pres normal	Pas de pb à les écouter - ne soulèvent pas de questions
CAA	bien	normaux	Pas de pb à les écouter - ne soulèvent pas de questions
CAG	fort mais compréhension plus difficile	zozotement un peu	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
COF	inaudible ou presque	dentier	inaudible
COM	On comprend pas	incompréhensible	voix ressemblants à celles des résidants - connaissance, capable de lire ou de répéter mais émotionnellement cela va altérer la compréhension
DAC	Flou, moins bien	défaut de quoi ?	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
DAJ	bien	voix calme	Pas de pb à les écouter - ne soulèvent pas de questions
DAP	pas mal mais faible bien prononcé	l'air fatigué	Pas de pb à les écouter - ne soulèvent pas de questions
FOM	très audible et un peu plus faible comme son, comme tonalité	voix calme	Pas de pb à les écouter - ne soulèvent pas de questions
FOY	pas mal mais faible bien prononcé	l'air fatigué	Pas de pb à les écouter - ne soulèvent pas de questions
HIM	pas mal mais plus faible toujours bien prononcé	voix plus calme	Pas de pb à les écouter - ne soulèvent pas de questions
LEF	c'est très bien	femmes plus énergiques	Pas de pb à les écouter - ne soulèvent pas de questions
LEJ	pas mal mais faible bien prononcé	quelqu'un de calme	Pas de pb à les écouter - ne soulèvent pas de questions
MAC 13	Flou, moins bien	ne sait pas	voix ressemblants à celles des résidants - connaissance, capable de lire ou de répéter mais émotionnellement cela va altérer la compréhension
MAC 17	bien aussi	normaux	Pas de pb à les écouter - ne soulèvent pas de questions
MAF	pas mal mais faible bien prononcé	a peu pres normal	Pas de pb à les écouter - ne soulèvent pas de questions
ORL	bien aussi	femmes plus	Pas de pb à les écouter - ne soulèvent pas de questions

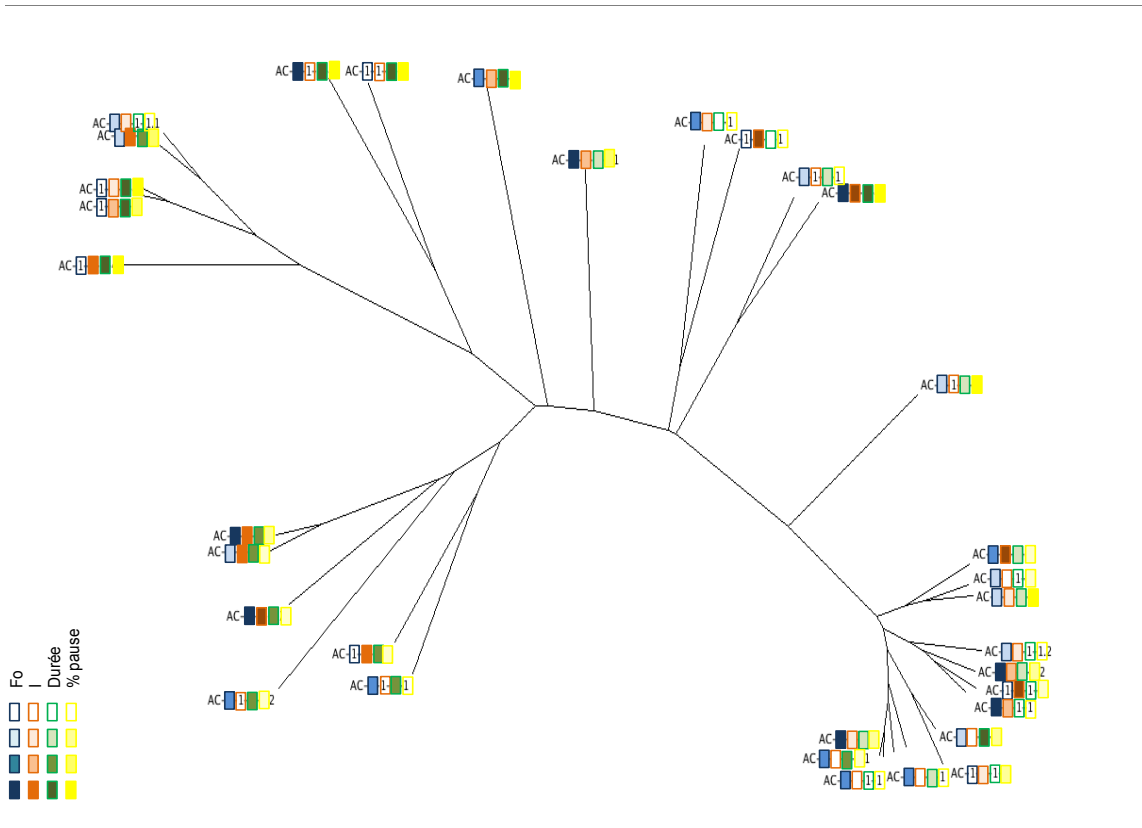
		énergiques	
PEJ	bien aussi	normaux	Pas de pb à les écouter - ne soulèvent pas de questions
PIM	pas mal mais plus faible toujours bien prononcé	voix plus calme	Pas de pb à les écouter - ne soulèvent pas de questions
PIP	On comprend pas	on dirait qu'il plaisante ou ils ne peuvent pas parler	voix ressemblants à celles des résidents - connaissance, capable de lire ou de répéter mais émotionnellement cela va altérer la compréhension
PUM	pas mal mais plus faible toujours bien prononcé	voix plus calme	Pas de pb à les écouter - ne soulèvent pas de questions
RAF	bien aussi	femmes plus énergiques	Pas de pb à les écouter - ne soulèvent pas de questions
REA	On comprend pas	on dirait qu'il plaisante ou ils ne peuvent pas parler	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
ROC 05	inaudible ou presque	sait pas	voix ressemblants à celles des résidents - connaissance, capable de lire ou de répéter mais émotionnellement cela va altérer la compréhension
ROC 07	pas mal mais faible bien prononcé	voix calme	Pas de pb à les écouter - ne soulèvent pas de questions
SAJ	On comprend pas	quelqu'un qui se réveille	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
SOB	bien aussi	femmes plus énergiques	Pas de pb à les écouter - ne soulèvent pas de questions
SUM	très audible et un peu plus faible comme son, comme tonalité	voix calme	Pas de pb à les écouter - ne soulèvent pas de questions
VAS	inaudible ou presque	voix de personne âgées	Non hésitant mais appareillage prothèse - gens qui ont toujours su parler
WAL	On comprend pas	voix de personne âgées	voix ressemblants à celles des résidents - connaissance, capable de lire ou de répéter mais émotionnellement cela va altérer la compréhension

### 7.1.4.3. ARBRES ANNOTES AVEC LES RESULTATS DE L'ANALYSE ACOUSTIQUE DES PARAMETRES PROSODIQUES

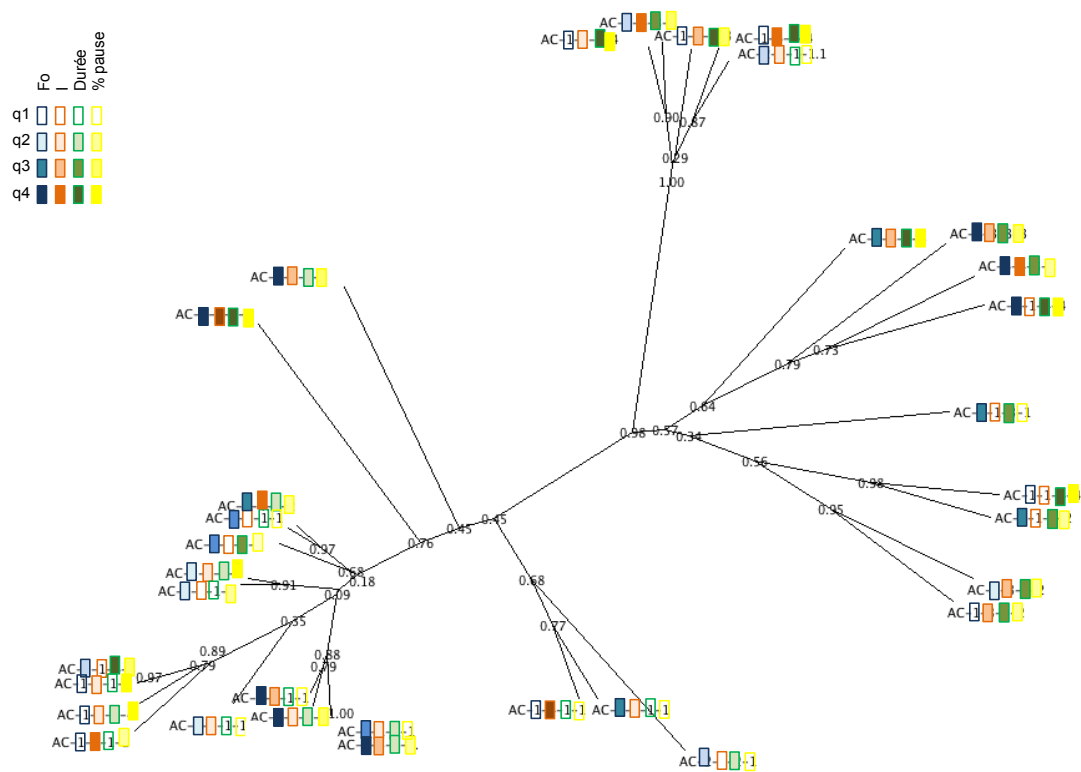
#### 7.1.4.3.1. ARBRES POUR LES EXPERTS



### 7.1.4.3.2. ARBRES POUR LES NON EXPERTS



### 7.1.4.3.3. ARBRES POUR LES PROCHES



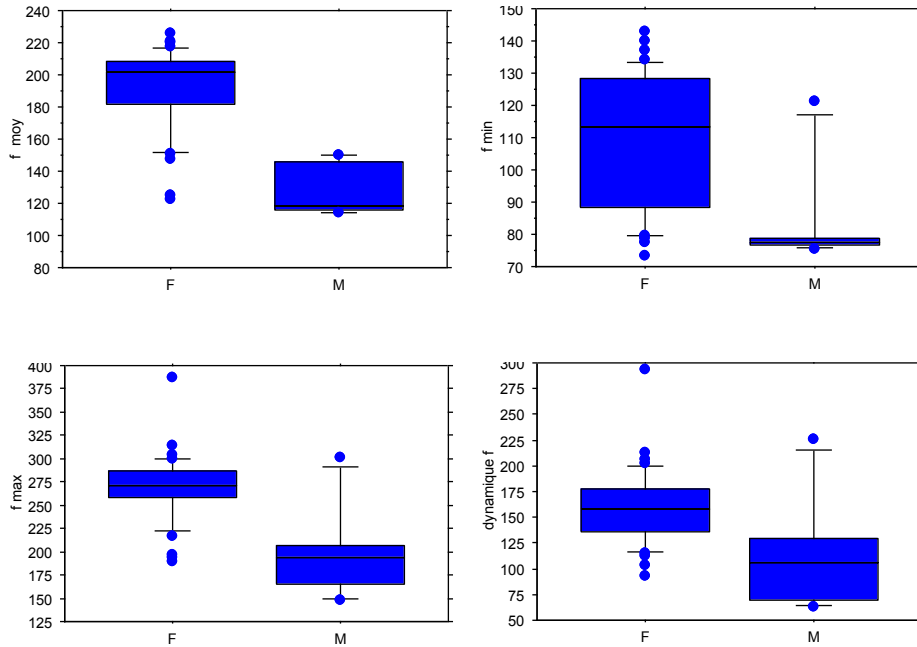


7.1.5.1. DESCRIPTION DES STIMULI VOCAUX

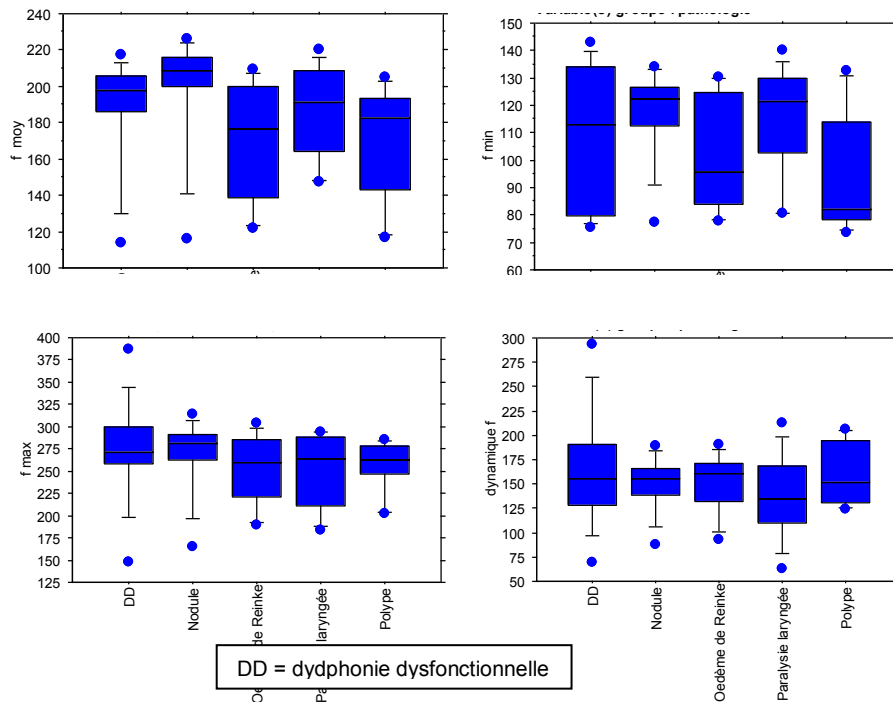
7.1.5.1.1. LES PARAMETRES ACOUSTIQUES

7.1.5.1.1.1. LES PARAMETRES DE FREQUENCES

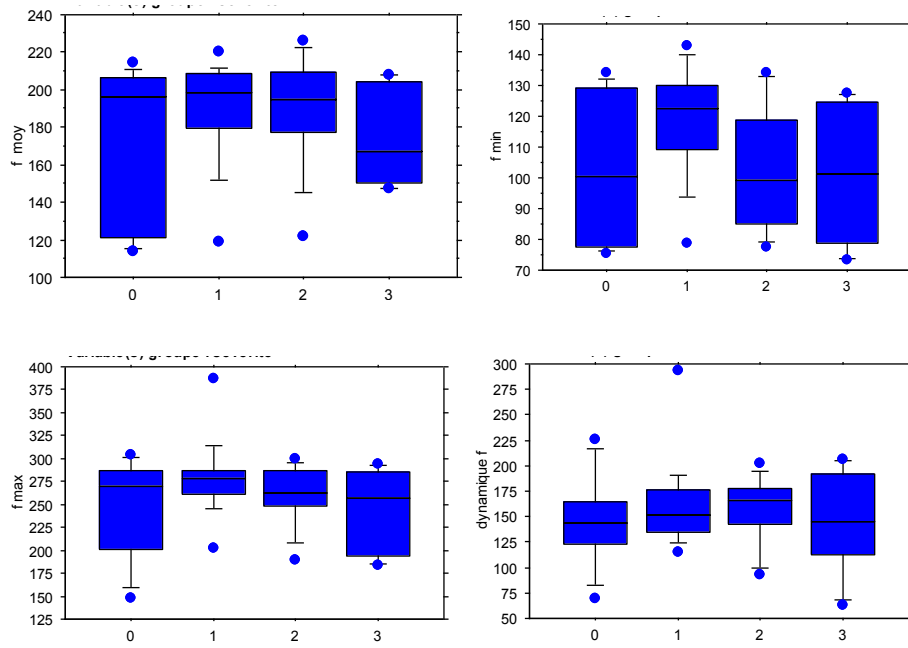
7.1.5.1.1.1.1. EN FONCTION DU GENRE



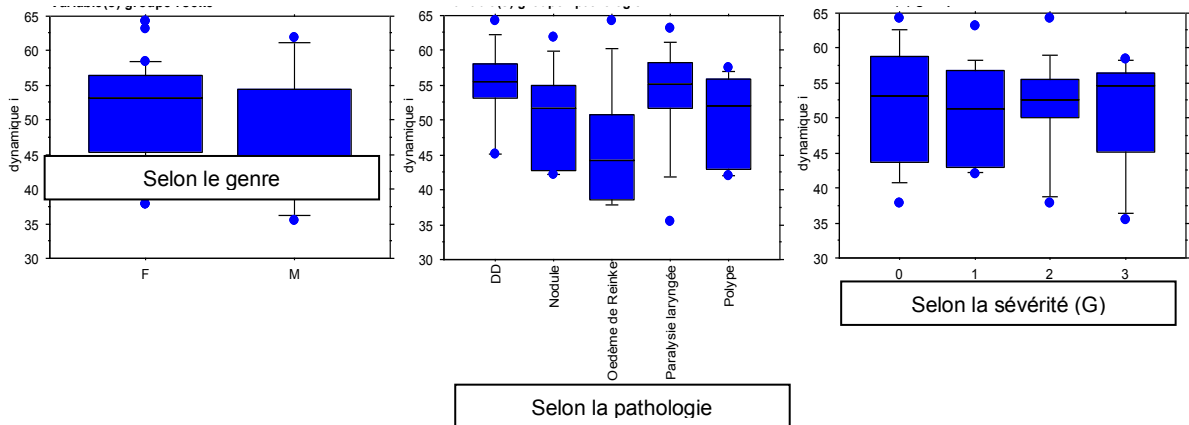
7.1.5.1.1.1.2. LES PARAMETRES DE FREQUENCES EN FONCTION DE LA PATHOLOGIE



### 7.1.5.1.1.3. LES PARAMETRES DE FREQUENCES EN FONCTION DE LA SEVERITE (G)

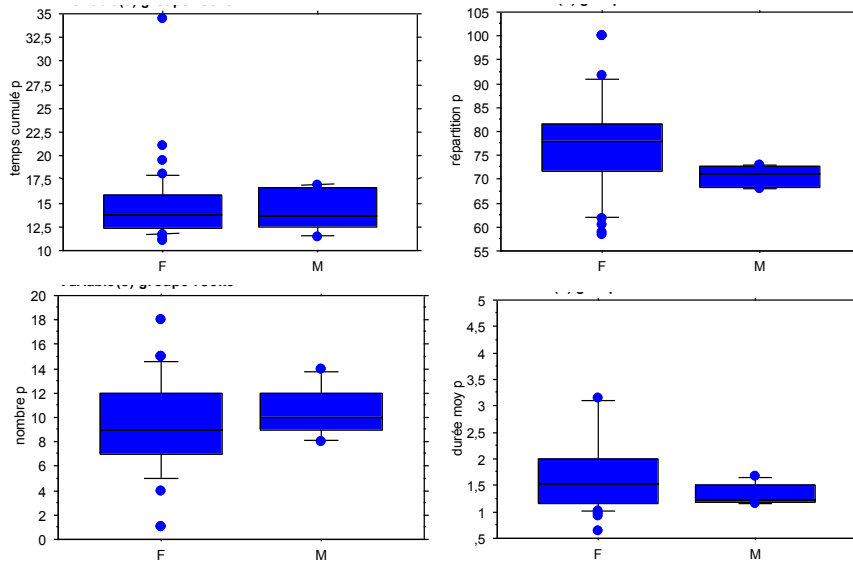


### 7.1.5.1.1.2. LES PARAMETRES D'INTENSITE

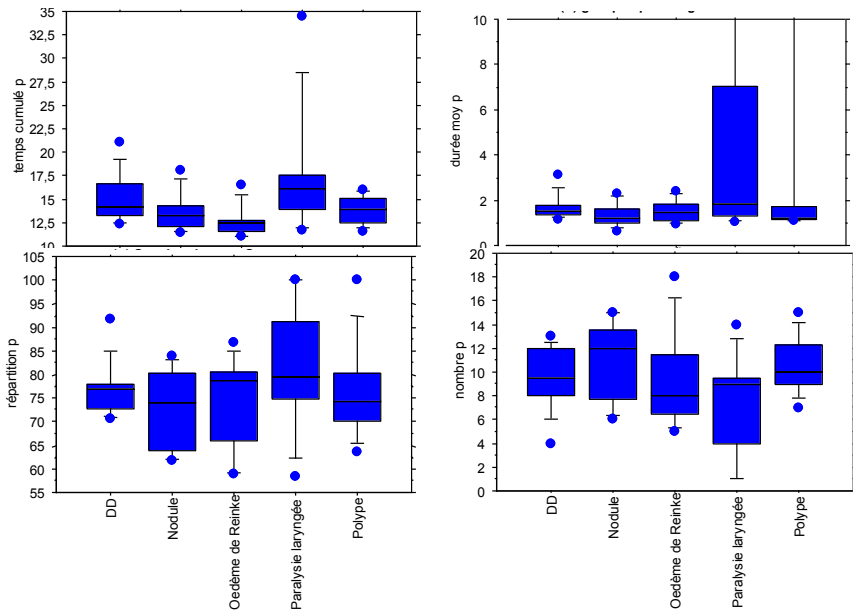


### 7.1.5.1.1.3. LES PARAMETRES DES PAUSES

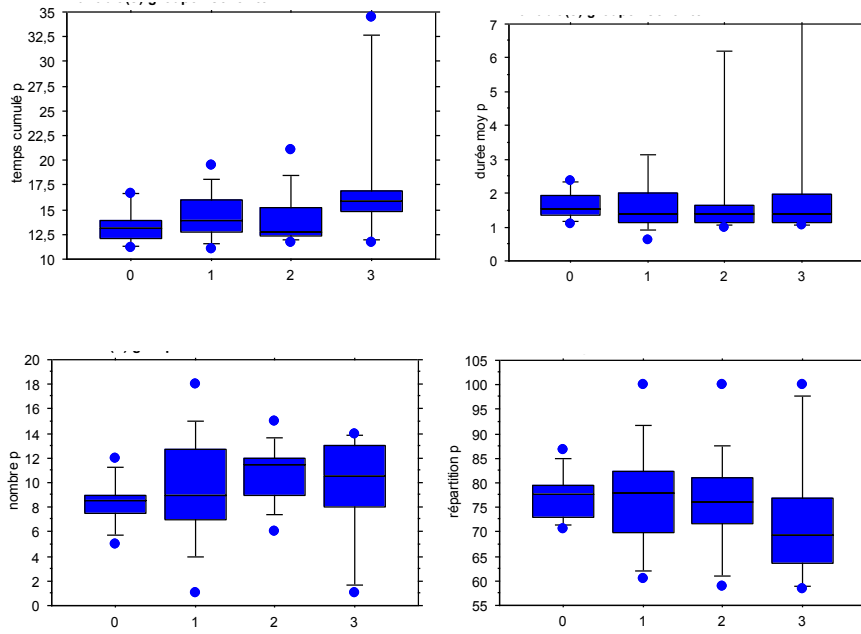
#### 7.1.5.1.1.3.1. EN FONCTION DU GENRE



#### 7.1.5.1.1.3.2. EN FONCTION DE LA PATHOLOGIE



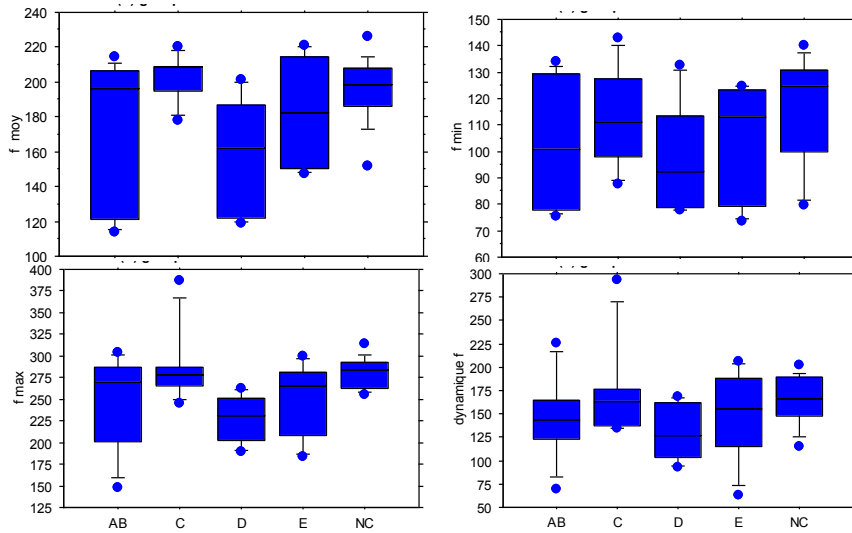
### 7.1.5.1.1.3.3. EN FONCTION DE LA SEVERITE (G)



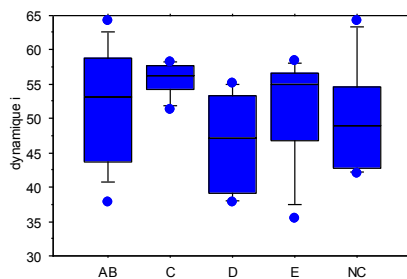
## 7.1.5.2. REPARTITION DES PARAMETRES ACOUSTIQUES EN FONCTION DES CLASSES

### 7.1.5.2.1. POUR LES EXPERTS

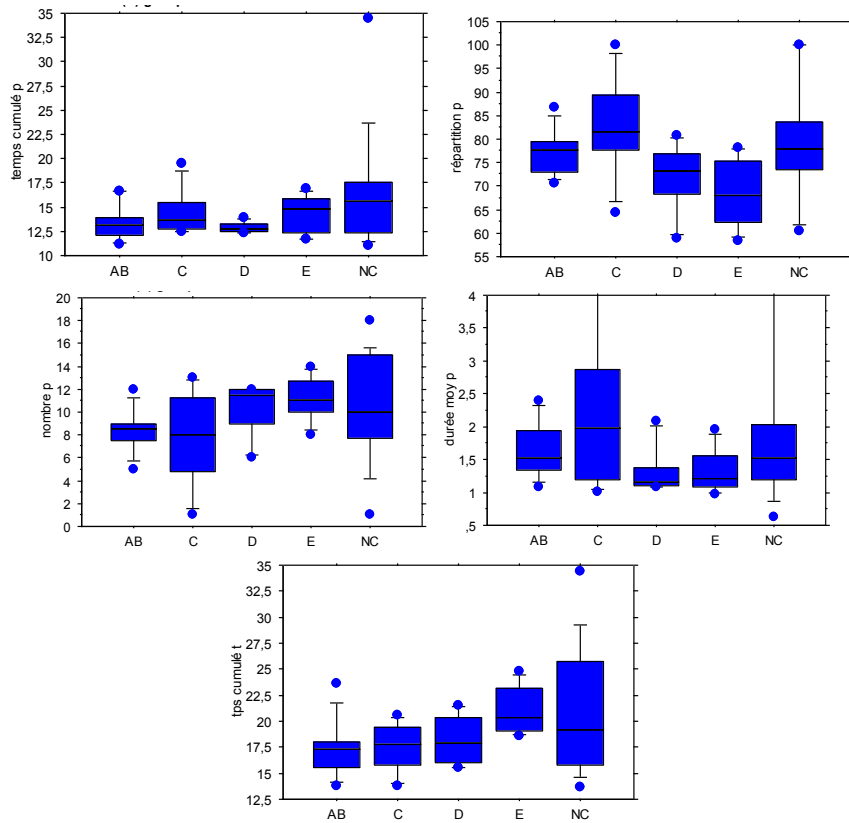
#### 7.1.5.2.1.1. REPARTITION DE LA FREQUENCE SELON LES CLASSES DES ARBRES



#### 7.1.5.2.1.2. REPARTITION DE LA DYNAMIQUE DE L'INTENSITE SELON LES CLASSES DES ARBRES

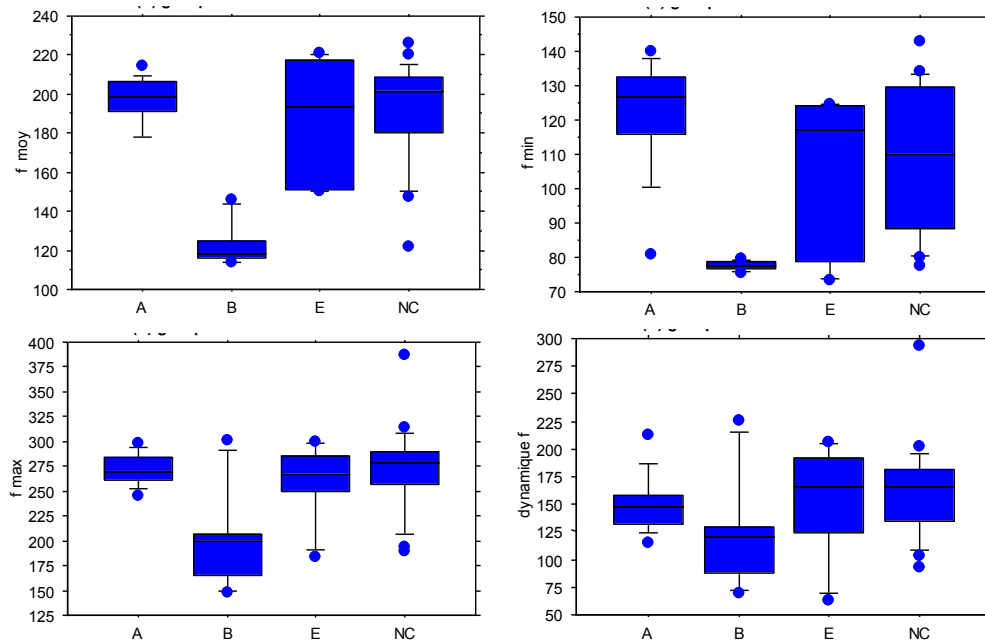


### 7.1.5.2.1.3. REPARTITION DES PAUSES ET DONNEES TEMPORELLES SELON LES CLASSES DES ARBRES

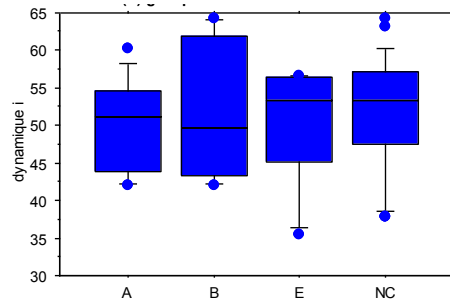


### 7.1.5.2.2. POUR LES NON EXPERTS

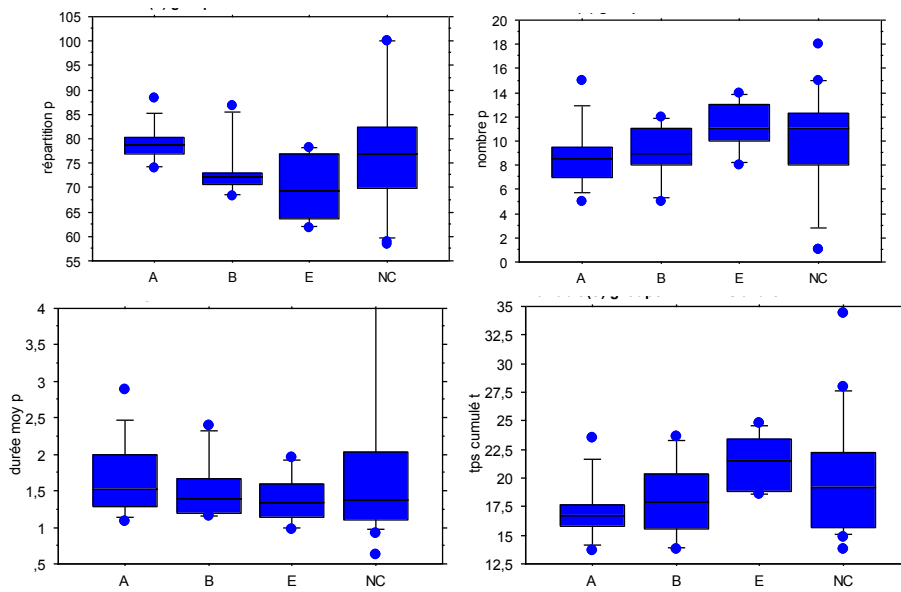
#### 7.1.5.2.2.1. REPARTITION DE LA FREQUENCE SELON LES CLASSES DES ARBRES POUR LES NON EXPERTS



7.1.5.2.2. REPARTITION DE LA DYNAMIQUE DE L'INTENSITE SELON LES CLASSES DES ARBRES POUR LES NON EXPERTS



7.1.5.2.3. REPARTITION DES PAUSES ET DONNEES TEMPORELLES SELON LES CLASSES DES ARBRES



### 7.1.5.3. EXTRAITS DES COMMENTAIRES POUR LES DYSPHONIES

Mots ou segments pertinants EXPERTS	Mots ou segments pertinants NON EXPERTS
""voix - gros forçage/début de - douloureuse - raucité - voilée - souffle - placement vocal - débit (inadéquation, rapidité) - nasonnement - pathologique de type neuro (type dysarthrie) - normale - gestion respiratoire - de conteur""	""rocaillement léger/très très rocailleuse/plus fort - gêne, force sur la voix - masculine - grave/très/plus grave - jeune/relativement/très - femme - accent étranger - voix posée - sans problème - garçon (jeune) - fillette de 7 ans - mature""
""voix - normale - rauque - éraillée - mal "	""trouble articulaire, langue ""gênante"" - ""nasalisation - effort pour produire un son, voix douloureuse à l'écoute - homme - femme - ""usée"" (de fumeur)- ""fatiguée"" - sans problème - âge - ""granuleuse"" - abimées - amélioration ""
voix - préservée - normale/subnormale - "	""normales, c'est-à-dire claires et limpides - "
""voix - harmonieuse, normale - éraillée - "	""voix - grave - parle avec le nez - on entend "
""voix - altération : très/assez/peu/pas""	""pathologie légère/lourde/limite - problème de voix plus prononcé - rhume - "
"" normale - fatiguée - soufflée - aggravée - "	""voix - plus aigüe que ce qu'elle devrait être - nasillarde - forcée sur les syllabes - essouffement - rauque (fumeur)/très rauque - enfant - sans rien à signaler - aphonie""
""féminin - masculin - normale - éraillée/légèrement - "	""personne qui donne/ne donne pas envie "
""voix de nodule voire polype ou sulcus glottidis, ressemble à de la voix après cordectomie - voix de fumeur : éraillée, cordes vocales boudinées, timbre mat, souffle - fuite d'air (mauvaise fermeture des cordes)/petite fuite - serrage laryngé/leger serrage - nasalisation - normale /dysphonie/légère dysphonie - ondulation des cordes vocales - forçage vocal - bitonalité - asthénie""	""voix - très éraillée/légèrement - homme - femme - normale - différente des autres""
""serrage/serrage plus marqué - coups de glotte - articulation tendue, voix ""coincées"" dans un registre aigu - désonorisations justement dans les aigus - registre medium - raucité dans les graves - registre grave - désonorisation - asthénie - souffle - registre aigu - compensation par une tendance au serrage et utilisation d'un timbre métallique (twang) - sans compensation donc voix assourdie - aggravée""	""voix - enfant - timbre- aigus - grave - qui "
""pas de patho/pas forcément patho - légère raucité/raucité - voix qui ne nécessite pas forcément une prise en charge - souffle - serrage - instabilité - asthénie (fatigabilité)""	""voix - enfant - féminin - agréable/pas totalement agréable/pas désagréable/désagréable - fluide/très fluide sur la lecture/absence de fluidité - difficilement classable - modifiée - peu souvent entendue - normalité - subie = on sait qu'ils ne font pas exprès - masculin - sortie d'aphonie - tabagisme chronique - douce - profonde - posée - accent""
""voix - altération : très/assez/peu/pas""	""voix - normale - nasonnée - rauque (= de fumeur) - chevrotante - enrouée (qui tire, qui se serre)""
""voix dysphoniques de personnes âgées,"	""voix - enfant - homme - femme - nasonnée "
""accord pneumo-phonique : pauses inspiratoires bruyantes, mal à l'aise par rapport au geste de lecture, ne savent pas comment utiliser leur articulation. pas d'homogénéité des consonnes et des voyelles. - normale, sans altération, timbre/voix équilibrée dans les caractéristiques, on ne le sent pas essoufflé - dysarthrie (ralentissement du débit, mais bonne articulation, ""voix de mémé"" sans altération de l'intensité ni quoi que ce soit - problème de hauteur, mal placée - bitonalité - dysphonie, timbre altéré, voilée, éraillée, rauque, type nodule ou polype, ""voix d'enseignant"" - de fumeurs, de bandes ventriculaires, serrée dans la gorge, on sent un serrage""	""voix - femme - âge plutôt jeune/plus âgé - pathologie - enfant - homme""
""femme - homme - normale/subnormale""	""voix - nasonnée - pharyngée enrouée - "
""voilée/un peu voilée, désonorisée/seulement par moments, soufflée/un peu soufflée, ""rigide""/assez rigide/pas rigide(fluctuation mais la vibration y est) = pas de souplesse de vibration, d'ondulation de la muqueuse, fixée, forçage/petit forçage - vibration irrégulière qui donne un aspect éraillé/on sent l'éraillage mais moins dramatiques que d'autres/moins d'éraillage - nodule, dysphonie de l'enfant - pas ordinaire ou normale mais pas dramatique/normale - ""graillonnant"" , vibration postérieure, pas de son, limite chuchotée, voix ""acier"" (un peu métallique)""	""voix - très claire - vibrations - normalité""

## 7.2. ANNEXES DU CHAPITRE 4 (INTELLIGIBILITE ET SEVERITE)

### 7.2.1. LES PHRASES DU HINT UTILISEES POUR LE CORPUS DE STIMULI DIFFERENTS

Seules les phrases soulignées ont été sélectionnées pour cette expérience

#### Liste 1

Ø Tous les poissons vivent dans l'eau.  
Elle est très gentille.  
Mes vacances sont finies.  
Nous allons à la banque.  
L'arbre est bien décoré.  
Le souper était chaud.  
Elle a perdu son bas.  
Ce garçon mange ta poire.  
La planète est grosse.  
Tes frères jouent au base-ball.

#### Liste 3

Mon parapluie est jaune.  
Le chien joue avec son os.  
Ils ont trouvé un sapin.  
Elle jouera de la flûte.  
Il a eu mal au dos.  
Elle a dit une menterie.  
Il raconte une belle histoire.  
Elle prend un bain chaud.  
Cette église est très vieille.  
Ces dragons crachent du feu.

#### Liste 6

Sa cabane est dans l'arbre.  
La musique était très forte.  
Mon frère ramasse ces poubelles.  
Ta soeur aime chanter.  
Les champignons sont bruns.  
L'ours est pris dans le piège.  
Le chocolat a fondu.  
Il écoute de la musique.  
Elle se lève tôt ce matin.  
Ils jouaient à la cachette.

#### Liste 8

L'avion est dans les nuages.  
Ce bébé mange souvent.  
Ils regardent ce spectacle.  
Tes roses blanches sont belles.  
La souris est un rongeur.  
L'oiseau a perdu des plumes.  
Elle porte un chandail.  
La dame lave mon visage.  
Cette histoire est triste.  
Il cherche ce panier de pommes.

#### Liste 2

Il joue aux quilles avec vous.  
Des enfants vont à l'école.  
Elle mange ta soupe.  
Mon bateau est très rapide.  
Elle se salit les mains.  
La machine fait du bruit.  
Il prend soin de ma mère.  
Votre maison est grise et bleue.  
Ce petit garçon chante bien.  
Ta coiffeuse coupe les cheveux.

#### Liste 4

Le voleur est dans la banque.  
Elles prennent une marche.  
L'oiseau est sur une branche.  
Il était très patient.  
Ils ont marché sur ce pont.  
Elles font des bonhommes de neige.  
Ils ont cassé tous les oeufs.  
Ma fille lave ses mains.  
Tes souliers sont en cuir.  
La salle était vide.

#### Liste 9

Je suis tombé sur la glace.  
Elle regarde par sa fenêtre.  
L'enfant dessine un cochon.  
Ta soupe était délicieuse.  
Ce chien dormait dehors.  
Il joue avec ma poupée.  
Elle joue aux billes avec moi.  
Ta fenêtre est ouverte.  
Le vent fait bouger des feuilles.  
Il est parti à la pêche.

#### Liste 11

Il est né à l'hôpital.  
Des enfants courent dehors.  
Elle a compté jusqu'à dix.  
Ce casse-tête est difficile.  
Elle a caché ta plume.  
Tu as caché son jouet.  
La mère berce son enfant.  
Ce marchand vend les bonbons.  
La souris mange du fromage.  
Ces grenouilles sont vertes.

#### Liste 5

Elle a fait son lit.  
Il ne faut pas manger vite.  
Il porte des boucles d'oreille.  
Mes lunettes sont rondes.  
Ce chat joue avec ma balle.  
Elle est fâchée contre ton père.  
Il mange une salade aux fruits.  
Ce lapin ronge une carotte.  
Elle est partie en voyage.  
Le savon nettoie bien.

#### Liste 7

Elle dessine une fleur bleue.  
Mes vaches sont dans l'étable.  
L'ours trouve du miel.  
La fille dort dans son lit.  
Ce pirate cherche un trésor.  
Ils aimaient tes raisins secs.  
Des enfants sautent dans la neige.  
Je voyage par autobus.  
Elle a écrit au tableau.  
Il porte plusieurs bagues.

#### Liste 10

J'ai un livre à colorier.  
Papa tirait ce chariot.  
Elle mange de la crème glacée.  
Tout le monde est en classe.  
Il faut lire ta page.  
Elle est perdue dans la ville.  
Le soleil se couche tôt.  
Il boit du jus d'orange.  
Ce bonbon est très sucré.  
Elle nage dans la rivière.

#### Liste 12

Elle vit dans la jungle.  
J'ai fabriqué cette cabane.  
L'éléphant a une longue trompe.  
Ce camion est rouge.  
Il saute sur mon trampoline.  
La marmotte creuse un trou.  
Elle achète des légumes frais.  
Sa femme joue du piano.  
Mes cheveux sont blonds.  
Ton cerf-volant est jaune.



7.2.2. TABLEAUX AVEC LES VALEURS DES CORRELATIONS ENTRE LES PARAMETRES DES DIFFERENTES TACHES.

Expérience 1	voix		résonance		prosodie		production phonémique	
	Sev	Intel	Sev	Intel	Sev	Intel	Sev	Intel
jugé 1	0,70	0,72	0,85	0,85	0,73	0,75	0,88	0,86
jugé 2	0,76	0,69	0,54	0,45	0,59	0,67	0,74	0,68
jugé 3	0,62	0,72	0,32	0,46	0,58	0,46	0,78	0,82
jugé 4	0,80	0,75	0,84	0,84	0,86	0,84	0,87	0,86
jugé 5	0,54	0,46	0,68	0,61	0,40	0,42	0,85	0,87

Expérience 2	voix		résonance		prosodie		production phonémique	
	Sev	Intel	Sev	Intel	Sev	Intel	Sev	Intel
jugé 1	0,57	0,63	0,83	0,83	0,69	0,63	0,92	0,91
jugé 2	0,75	0,74	0,86	0,86	0,71	0,71	0,91	0,91
jugé 3	0,72	0,53	0,65	0,54	0,84	0,79	0,89	0,89
jugé 4	0,64	0,61	0,83	0,78	0,59	0,57	0,95	0,95
jugé 5	0,77	0,76	0,80	0,78	0,40	0,35	0,89	0,95

Tableau 4.3.6 : Corrélations entre la Sévérité et l'intelligibilité d'une part et les tâches d'autre part (couleur d'autant plus intense, que la corrélation est forte)

Expérience 1	voix	résonance	prosodie	production phonémique
jugé 1	-0,27	-0,06	-0,23	0,08
jugé 2	-0,02	0,09	-0,12	0,14
jugé 3	-0,16	-0,07	-0,02	-0,17
jugé 4	-0,06	-0,16	-0,08	-0,05
jugé 5	0,22	-0,01	-0,03	-0,08

Expérience 2	voix	résonance	prosodie	production phonémique
jugé 1	-0,35	-0,14	-0,07	0,15
jugé 2	-0,09	0,01	-0,17	0,11
jugé 3	0,25	0,10	-0,06	-0,17
jugé 4	0,35	0,46	0,25	0,19
jugé 5	0,004	0,04	0,14	-0,22

Tableau 4.3.7 : Corrélations entre les scores de l'analyse préalable et la différence de score entre sévérité et altération de l'intelligibilité (sévérité – intelligibilité) pour les expériences 1 et 2.

7.3 ANNEXES EN RAPPORT LE CHAPITRE 5 (ROBUSTESSE ET CONTRIBUTION A L'INTELLIGIBILITE).

7.3.1. ANNEXES EN RAPPORT AVEC LA SECTION.5.3.

### 7.3.1.1 .ANNEXE EN RAPPORT AVEC LA SECTION 5.3.1.1 : DESCRIPTION DE LA TECHNIQUE CHIRURGICALE

Description de la technique chirurgicale de l'oropharyngectomie par voie de buccopharyngectomie transmandibulaire

Le principe de cette voie d'abord est d'assurer l'accès à l'oropharynx en pratiquant une incision cutanée cervicale prolongée en région labiomentonnière après avoir contourné la houppe du menton . La lèvre inférieure est transfixiée. Une mandibulotomie est réalisée juste en avant du nerf mentonier qui est souvent traumatisé (donnant lieu à une paralysie de la lèvre inférieure et du menton du côté opéré) La branche montante de la mandibule est découverte après section du masséter. . Le plancher buccal est ensuite sectionné et la mandibule luxée du côté de la lésion de manière à exposer la région où se situe la tumeur.

La résection de la tumeur est réalisée en fonction de sa topographie en essayant de passer à distance. Elle peut donc intéresser la langue mobile et ou la base, la loge amygdalienne, le voile du palais, le plancher buccal, la mandibule

Quand le défaut est volumineux, un comblement par un lambeau est nécessaire. Le choix de ce lambeau dépend de la forme et du volume à combler. Les techniques les plus fréquemment utilisées sont le transfert d'un lambeau musculocutané de grand pectoral ou d'un lambeau libre antébrachial.

Quand une résection de la mandibule est pratiquée du fait d'un envahissement tumoral ou d'une trop grande proximité de la tumeur, la reconstruction est quasi systématique soit par un lambeau libre de fibula, soit par une plaque en titane.

La fermeture lorsque la mandibule est conservée est assurée par la suture des plans musculaires et muqueux restant du plancher buccal et du vestibule puis par une ostéosynthèse de la mandibule. Enfin la fermeture des plans musculaires et cutanés sont réalisées de la région labiomentonnière puis cervicale.

### 7.3.1.2. ANNEXE EN RAPPORT AVEC LA SECTION 5.3.2 : TABLEAUX CONTENANT LES VALEURS QUI ONT PERMIS LA CONSTRUCTION DES REPRESENTATIONS GRAPHIQUES

Cette annexe regroupe tous les tableaux contenant les valeurs qui ont permis la construction des représentations graphiques de la section 5.3.2.

Pour faciliter le lien entre le tableau de l'annexe et la figure de la section 5.3.2 lui correspondant, la numérotation du tableau est la même que celui de la figure.

Auditeurs	ok	err	xxx	Nombre total de cibles	Taux de bonne perception %	Taux d'erreurs %	Taux de non réponse %
AG	599	345	16	960	62,40	35,94	1,67
AR	578	365	17	960	60,21	38,02	1,77
BW	609	343	8	960	63,44	35,73	0,83
CD	597	351	12	960	62,19	36,56	1,25
EP	571	388	1	960	59,48	40,42	0,10
JB	584	375	1	960	60,83	39,06	0,10
LM	587	368	5	960	61,15	38,33	0,52
LU	610	341	9	960	63,54	35,52	0,94
MR	579	375	6	960	60,31	39,06	0,63
NG	562	386	12	960	58,54	40,21	1,25
PM	603	349	8	960	62,81	36,35	0,83
Totaux/taux moyens	6479	3986	95	10560	61,35	37,75	0,90

Tableau n°5.3.2.1 : Description des réponses au test de perception pour les consonnes par auditeur à l'origine du graphique de la figure n°5.3.2.1

Cible	Perçu																	Total
	p	t	k	f	s	ch	b	d	g	v	j	z	m	n	l	r	xxx	
p	97	1	3	2			5								1		1	110
t	9	96	1	1				3										110
k	1		106	1					1								1	110
f				100	3	5				1							1	110
s				3	103	4												110
ch						110												110
b	4						95	5	2	1	1			1			1	110
d		6						96	7								1	110
g			1	1			5		101				1				1	110
v										107	1	1			1			110
j									1	2	106						1	110
z					1			1		4	1	103						110
m							15			1			84	5	4		1	110
n							1	5	2				8	81	13			110
l							4						1	3	102			110
r	1		1	1			4			3			1			98	1	110
Total	112	103	112	109	107	119	129	110	116	120	109	104	95	90	121	98	21	3520

Tableau n°5.3.2.1 bis : Matrice de confusion des deux locuteurs témoins en rapport avec le Tableau n°5.3.2.1 de la section 5.3.2.1.1.2. concernant l'analyse des confusions

Locuteurs	ok	err	xxx	Nombre total de cibles	Taux de bonne perception %	Taux d'erreurs %	Taux de non réponse %
Témoins	1584	167	9	1760	90,00	9,49	0,51
Patients	4895	3819	86	8800	55,63	43,40	0,98
Total général	6479	3986	95	10560	61,35	37,75	0,90

Tableau n°5.3.2.2 : Description des réponses au test de perception pour 2 groupes de locuteurs à l'origine du graphique de la figure n°5.3.2.2

Locuteur	BAD		PRM		Témoins		
Consonne	ok	Taux de bonne perception %	ok	Taux de bonne perception %	ok	Taux de bonne perception %	Nombre de cible pour un patient
p	52	94,55	45	81,82	97	88,18	55
t	53	96,36	43	78,18	96	87,27	55
k	53	96,36	53	96,36	106	96,36	55
b	45	81,82	50	90,91	95	86,36	55
d	49	89,09	47	85,45	96	87,27	55
g	51	92,73	50	90,91	101	91,82	55
s	53	96,36	49	89,09	102	92,73	55
z	54	98,18	49	89,09	103	93,64	55
ch	55	100,00	55	100,00	110	100,00	55
j	51	92,73	55	100,00	106	96,36	55
f	48	87,27	52	94,55	100	90,91	55
v	55	100,00	52	94,55	107	97,27	55
m	53	96,36	31	56,36	84	76,36	55
n	43	78,18	38	69,09	81	73,64	55
l	50	90,91	52	94,55	102	92,73	55
r	52	94,55	46	83,64	98	89,09	55
Total	817	92,84	767	87,16	1584	90,00	880

Tableau n°5.3.2.3 : Nombre de réussite et taux de bonne perception des consonnes par les locuteurs témoins (plus la couleur est intense, plus le pourcentage est élevé) à l'origine de la figure n°5.3.2.3

Locuteur	Témoins				BAD		PRM		
Lieu	ok	err	xxx	Taux de bonne perception %	ok	Taux de bonne perception %	ok	Taux de bonne perception %	Total pour un locuteur
Labial	483	63	4	87,82	253	92,00	230	83,64	275
Alvéolaire	580	79	1	87,88	302	91,52	278	84,24	330
Vélopalatal	423	14	3	96,14	210	95,45	213	96,82	220
Uvulaire	98	11	1	89,09	52	94,55	46	83,64	55
Nasalité									
Oui	165	54	1	75,00	96	87,27	69	62,73	110
Non	1419	113	8	92,14	721	93,64	698	90,65	770
Interrompu									
Non	993	103	4	90,27	514	93,45	479	87,09	550
Oui	591	64	5	89,55	303	91,82	288	87,27	330
Voisé									
Non	611	46	3	92,58	314	95,15	297	90,00	330
Oui	973	121	6	88,45	503	91,45	470	85,45	550
Total	1584	167	9		817		767		880

Tableau n°5.3.2.4 : Nombre de réussite et taux de bonne perception par traits pour les locuteurs témoins (plus la couleur verte est intense, plus le pourcentage est élevé) à l'origine de la figure n°5.3.2.4

Voyelles	BAD				PRM				Total pour un locuteur
	ok	err	xxx	Taux de bonne perception %	ok	err	xxx	Taux de bonne perception %	
a	171	5		97,16	170	4	2	96,59	176
an	135	40	1	76,7	117	57	2	66,48	176
i	173	1	2	98,3	173	1	2	98,3	176
ou	172	4		97,73	162	13	1	92,05	176
u	175	0	1	99,43	158	15	3	89,77	176

Tableau n°5.3.2.5 : Nombre de réussite et taux de bonne perception des voyelles pour les locuteurs témoins (plus la couleur verte est intense, plus le pourcentage est élevé) à l'origine de la figure n°5.3.2.5.

	nb	Fréquence syllabe « token »		Fréquence Phonétique du mot		Fréquence orthographique du mot	
		moyenne	DS	moyenne	moy	moyenne	DS
OK	1584	2958		1760	3946	1515	4444
err	167	3236	4242	1486	3455	1653	4206

Tableau n°5.3.2.6 : Comparaison de la moyenne des fréquences s de la syllabe ou du mot contenant la consonne en fonction de la réussite pour le groupe témoins représentée sur la figure n°5.3.2.6

	nb	Fréquence syllabe « token »		Fréquence Phonétique du mot		Fréquence orthographique du mot	
		moyenne	DS	moyenne	moy	moyenne	DS
OK	6479	3149	4242	1891	4161	1637	4334
err	3986	2728	3482	1486	3428	1350	4037

Tableau n°5.3.2.6 bis : Comparaison de la moyenne des fréquences de la syllabe ou du mot contenant la consonne en fonction de la réussite pour le groupe des patients représentée sur la figure n°5.3.2.6

Locuteurs	ok	err	xxx	Nombre total de cibles	Taux de bonne perception %	Taux d'erreurs %	Taux de non réponse %
alj	436	438	6	880	49,55	49,77	0,68
bac	602	271	7	880	68,41	30,80	0,80
bof	502	367	11	880	57,05	41,70	1,25
boj	267	595	18	880	30,34	67,61	2,05
coj	393	476	11	880	44,66	54,09	1,25
dec	748	130	2	880	85,00	14,77	0,23
eup	392	480	8	880	44,55	54,55	0,91
gaj	257	611	12	880	29,20	69,43	1,36
rac	670	202	8	880	76,14	22,95	0,91
wif	628	249	3	880	71,36	28,30	0,34
Patients	4895	3819	86	8800	55,63	43,40	0,98

Tableau n°5.3.2.7 : Description des réponses au test de perception par patients

	ok	Total	Taux % de perception
<b>Délai</b>			
Jusqu'à 3 ans	2685	4400	61,02
Plus de 3 ans	2210	4400	50,23
<b>site</b>			
Amygdaglosse	1690	2640	64,02
Base de langue	1104	1760	62,73
langue	2101	4400	47,75
<b>T</b>			
2	2671	4400	60,70
3	1957	3520	55,60
4	267	880	30,34
<b>Voie d'abord</b>			
BPTM	3497	6160	56,77
Naturel	1398	2640	52,95
<b>Résection mandibulaire</b>			
Non	3397	6160	55,15
Oui	1498	2640	56,74

Tableau n°5.3.2.8 : Impact des facteurs en rapport avec la tumeur sur le taux de perception juste

	ok	Total	Taux % de perception
<b>Grade</b>			
2	2461	4400	55,93
3	2167	3520	61,56
4	267	880	30,34
<b>Sangle labio-jugale</b>			
2	703	1760	39,94
3	2942	5280	55,72
4	1250	1760	71,02
<b>Langue</b>			
0	502	880	57,05
1	393	880	44,66
2	2624	5280	49,70
3	1376	1760	78,18
<b>Isthme oropharynx</b>			
2	2127	3520	60,43
3	2768	5280	52,42
<b>Mandibule</b>			
1	895	1760	50,85
2	1451	2640	54,96
3	2549	4400	57,93

Tableau n°5.3.2.9 : Impact des paramètres recueillis lors de l'évaluation clinique sur le taux de perception juste.

	Total par patient	ALJ	BAC	BOF	BOJ	COJ	DEC	EUP	GAJ	RAC	WIF	Moy des % patient
p	55	76,4	52,7	65,5	16,4	16,4	89,1	60,0	34,6	85,5	85,5	58,18
t	55	27,3	65,5	83,6	27,3	30,9	72,7	10,9	1,8	90,9	89,1	50
k	55	43,6	61,8	3,6	0,0	0,0	87,3	9,1	21,8	45,5	0,0	27,27
b	55	47,3	52,7	72,7	7,3	29,1	74,6	72,7	40,0	52,7	72,7	52,18
d	55	47,3	90,9	60,0	0,0	27,3	54,6	14,6	1,8	85,5	96,4	47,82
g	55	45,5	49,1	30,9	0,0	3,6	89,1	9,1	1,8	21,8	1,8	25,27
s	55	49,1	70,9	69,1	36,4	85,5	92,7	16,4	1,8	98,2	100,0	62
z	55	45,5	78,2	90,9	34,6	74,6	72,7	14,6	12,7	94,5	100,0	61,82
ch	55	14,6	67,3	14,6	7,3	56,4	98,2	50,9	1,8	78,2	70,9	46
j	55	9,1	41,8	34,6	30,9	38,2	92,7	40,0	3,6	58,2	49,1	39,82
f	55	45,5	50,9	50,9	85,5	47,3	94,6	98,2	81,8	81,8	76,4	71,27
v	55	61,8	81,8	81,8	72,7	63,6	90,9	78,2	74,6	63,6	58,2	72,73
m	55	67,3	78,2	52,7	34,6	65,5	78,2	67,3	85,5	72,7	70,9	67,27
n	55	69,1	94,6	74,6	18,2	72,7	92,7	27,3	14,6	96,4	96,4	65,64
l	55	45,5	76,4	72,7	25,5	78,2	85,5	63,6	1,8	92,7	76,4	61,82
r	55	98,2	81,8	54,6	89,1	25,5	94,6	80,0	87,3	100,0	98,2	80,91
Moy par loc.		49,82	70,68	56,8	34,16	40,52	85,14	46,98	35,56	79,33	76,15	

Tableau n°5.3.2.12 : Taux de bonne perception par consonne et locuteur en rapport avec la figure n°5.3.2.12 de la section 5.3.2.1.3.1

	ok	Total	Taux de bonne perception %
<b>Lieu</b>			
Labial	1769	2750	64,33
Alvéolaire	1920	3300	58,18
Vélopalatal	761	2200	34,59
Uvulaire	445	550	80,91
<b>Nasalité</b>			
Oui	731	1100	66,45
Non	4164	7700	54,08
<b>Interrompu</b>			
Non	3461	5500	62,93
Oui	1434	3300	43,45
<b>Voisé</b>			
Non	1731	3300	52,45
Oui	3164	5500	57,53

Tableau n°5.3.2.13 : Taux de perception juste par trait

Consonne perçue	Nb de confusion ou erreurs	Total Perçu	% de confusion
b	102	389	26,22
ch	154	407	37,84
d	143	406	35,22
f	599	993	60,32
g	155	294	52,72
j	203	423	47,99
k	123	273	45,05
l	232	572	40,56
m	140	510	27,45
n	112	475	23,58
p	66	386	17,1
r	219	664	32,98
s	255	596	42,79
t	103	380	27,11
v	898	1298	69,18
z	308	648	47,53

Tableau n°5.3.2.18 : Taux de confusion des consonnes (Nombre d'erreur /Nombre total de consonnes perçues) à la base de la figure n°5.3.2.18 section 5.3.2.2.2

	Taux de confusion
Lieu	
Labial	39,65
Alvéolaire	28,05
Vélopalatal	34,86
Uvulaire	32,98
Nasalité	
Nasal	10,86
Oral	2,69
Interrompu	
non interrompu	20,65
Interrompu	10,24
Voisement	
Non voisé	2,6
Voisé	5,58

Tableau n°5.3.2.19 : Taux de confusion des traits

	F mesure
b	0,6112886
ch	0,52873563
d	0,55020921
f	0,51069345
g	0,32938389
j	0,45220966
k	0,36452005
l	0,60606061
m	0,69811321
n	0,70829268
p	0,68376068
r	0,73311367
s	0,59511344
t	0,59569892
v	0,43290043
z	0,56761269

Tableau n°5.3.2.2.25 : Calcul de la F mesure pour les consonnes listées par ordre décroissant.



	F mesure
Lieu	
Labial	0,6823
Alvéolaire	0,6944
Vélopalatal	0,506
Uvulaire	0,7331
Nasalité	
Nasal	0,8422
Oral	0,9749
Interrompu	
non interrompu	0,8648
Interrompu	0,7038
Voisement	
Non voisé	0,9332
Voisé	0,9593

Tableau n°5.3.2.2.26 : F mesure estimée pour les traits.

### 7.3.3 ANNEXES EN RAPPORT AVEC LA SECTION 5.3.3 CONCERNANT LE RANG MEDIAN ET LA DESCRIPTION DE LA REPARTITION SUR LA BISSECTRICE

---

Cette annexe comprend le complément d'analyse réalisé pour justifier le choix du rang médian de la cible ou consigne comme indice de distorsion dans l'expérience de recherche de la robustesse perceptive (figure n°7.3.3.1 à 7.3.3.3).

Elle est suivie du tableau précisant la valeur moyenne de le l'indice choisi : le rang médian, pour chaque consonne (tableau n°7.3.3.1).

Enfin le dernier tableau contient la valeur des consonnes répartis sur la bissectrice du graphe rang médian des consonnes reconnu par le système en fonction du rang des consonnes par le système correspondant à une percption juste par les auditeurs puisque à ce moment là, la cible = le perçu (tableau n°7.3.3.2).

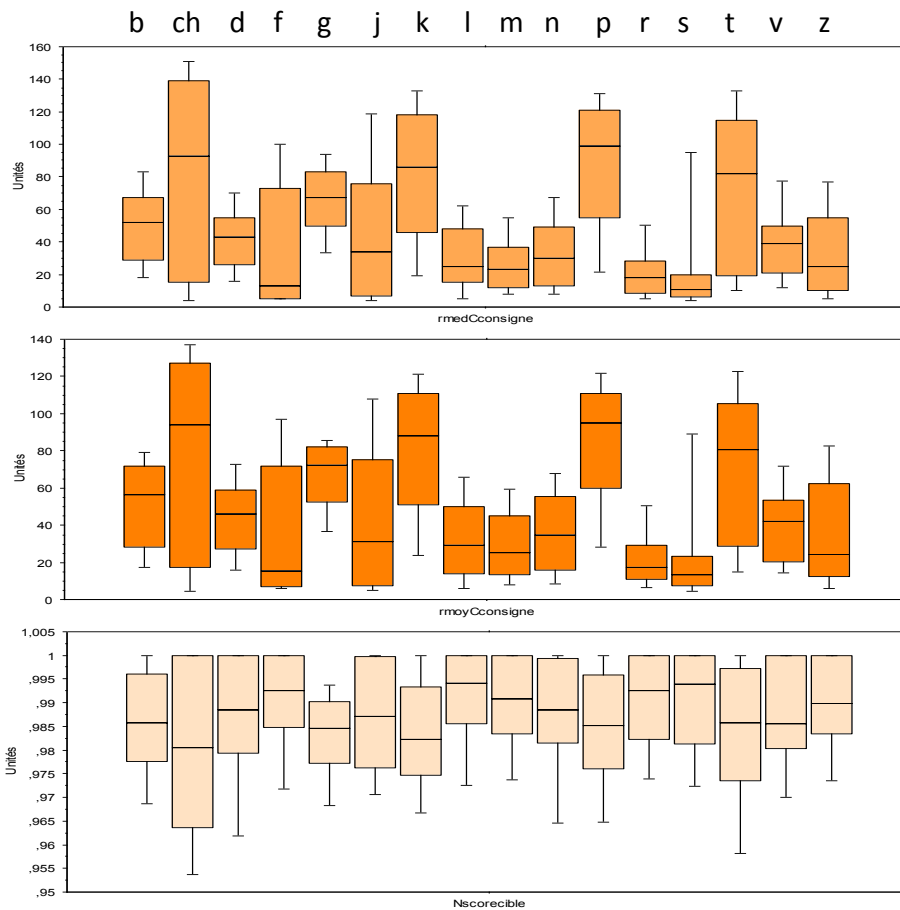


Figure n°7.3.3.1. : Indice de distorsion en utilisant les différents modes de calcul en fonction des consonnes. De haut en bas, la médiane du rang, la moyenne du rang et le score normé moyen

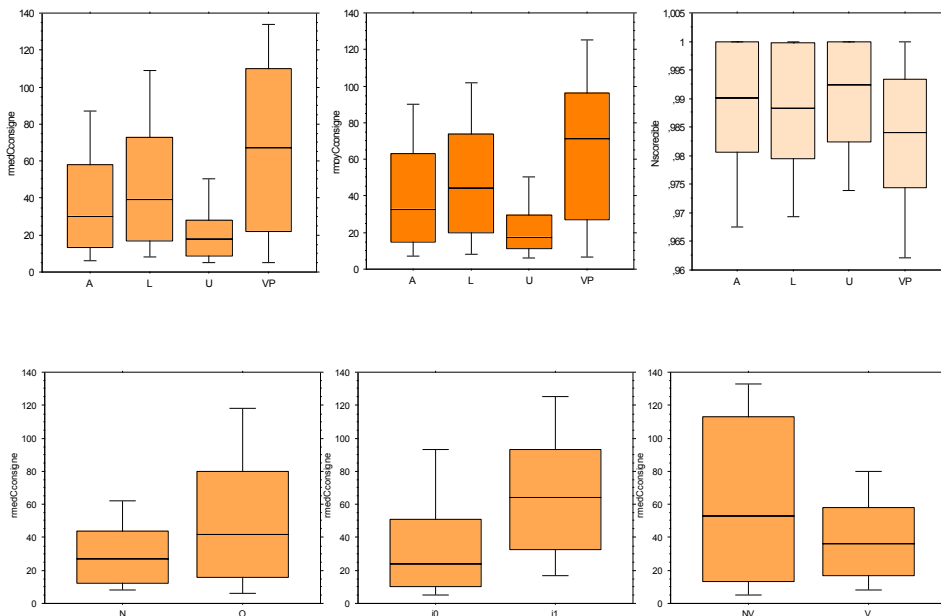


Figure n°7.3.3.2. : Indice de distorsion en utilisant les différents modes de calcul en fonction des traits. De haut en bas, la médiane du rang, la moyenne du rang et le score normé moyen

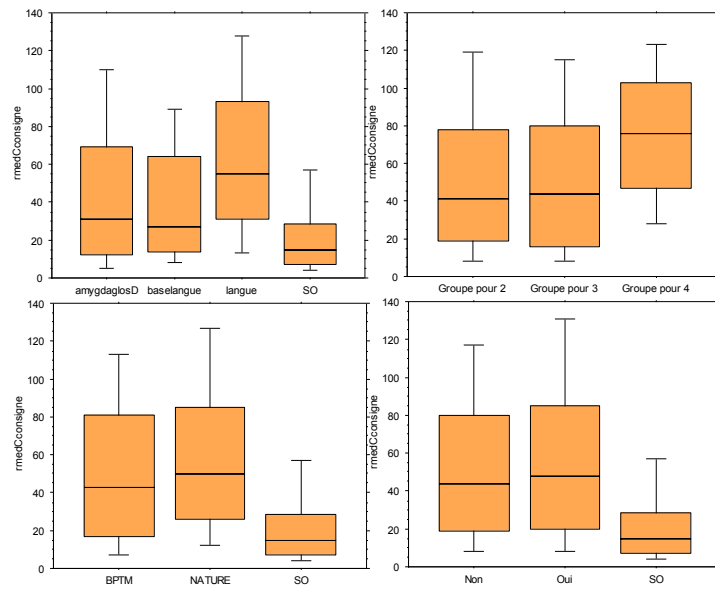


Figure n°7.3.3.3. : Indice de distorsion en utilisant les différents modes de calcul en fonction du site tumoral initial (sans objet = les 2 témoins).

	Moyenne de la médiane du rang		
	témoins	Patients	Différence
ch	4,3	91	-86,7
k	32,6	92,58	-59,98
t	28,8	78,48	-49,68
j	7,1	54,86	-47,76
p	56,2	93,08	-36,88
f	6,6	42,08	-35,48
g	42,2	70,36	-28,16
n	13,7	40,24	-26,54
d	21,8	47,02	-25,22
z	15,1	39,28	-24,18
s	9,5	30,3	-20,8
v	24,9	42,74	-17,84
m	16,5	30,9	-14,4
b	38,3	52,62	-14,32
l	21,3	34,98	-13,68
r	22,1	22,42	-0,32

Tableau n°7.3.3.1 : Moyenne de la médiane du rang en fonction de la consonne

	Rcible (classes)				
	0_20	20_40	40_60	60_80	>80
b	178	29	31	28	21
ch	146	29	8	0	70
d	157	11	29	52	14
f	312	15	16	0	49
g	50	24	9	11	45
j	158	14	5	15	27
k	53	18	0	10	69
l	224	56	11	9	40
m	266	58	21	11	14
n	235	26	27	34	39
p	139	37	0	19	125
r	337	46	38	21	3
s	274	8	1	10	48
t	206	5	21	1	42
v	218	83	41	9	49
z	284	3	19	0	34

Tableau n°7.3.3.2 : Recherche du nombre de points par cible sur la bissectrice en la découpant de 0 à 19 puis 20 à 39 puis 40 à 59 puis 60 à 79 puis >80

---

7.3.2 ANNEXES EN RAPPORT AVEC LA SECTION 5.4 : LA FREQUENCE MOYENNE DES CONSONNES

---

Consonnes	Fréquence moyenne de la consonne
l	7317,348
m	5571,252
d	5545,86
t	4464,8
k	4027,868
r	3137,144
v	2815,976
s	2101,286
n	2083,952
f	1641,64
p	1594,7975
b	1193,708
j	1035,158
z	1009,282
ch	640,164
g	569,37

Tableau n°5.4.4. : Fréquences moyennes des consonnes calculées à partir de la fréquence "Syll token" des syllabes du corpus en nombre d'occurrences par millions présentées par ordre décroissant représentées figure 5.4.4 de la section 4.4.3.2.



## 8. REFERENCES

---

<a href="#">8. Références</a>	337
<a href="#">8.1. Bibliographie</a> .....	339
<a href="#">8.2. Logiciels</a> .....	359
<a href="#">8.3. Adresse des Sites web</a> .....	360



## 8.1. BIBLIOGRAPHIE

---

### A

AFNOR. Association française de normalisation. Saint Denis, 2004

Allard ER, Williams DF. Listeners' perceptions of speech and language disorders. *J Communication disorders* 2008;41:108-123

Allen JB. Consonant recognition and the articulation index. *J. Acoust. Soc. Am.* 2005;117:2212-2223

Andruski JE, Blumstein S, Burton M. "The effect of subphonetic differences on lexical access" *Cognition* 1994;52:163-187

ANSI S3.5. American National Standard Methods for the Calculation of the speech Intelligibility Index. New York, 1997

Arai T, Greenberg S. Speech intelligibility in the presence of cross-channel spectral asynchrony, IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, 1998,933-936

Auran C, Bouzon C, Hirst D, Lévy C, Nocéra P. Algorithme de prédiction d'élisions de phonèmes et influence sur l'alignement automatique dans le cadre du projet Aix-MARSEC. In Journées d'études de la Parole, JEP, Fez, Maroc, 2004

Auzou P, Gaillard MJ, Ozsancak C, Leonardon S, Jan M, Hannequin D. Evaluation clinique de la dysarthrie. Isebergues: Ortho Edition, 1998

Auzou P, Rolland Monnoury V. Batterie d'évaluation clinique de la dysarthrie. Isebergues:Ortho Edition, 2002

Auzou P. Les objectifs du bilan de la dysarthrie. Dans: Les dysarthries. Auzou P, Rolland Monnoury V, Pinto S, Özancak C. Marseille: Solal, 2007:189-195.

### B

Bagou O, Frauenfelder UH. Alignement lexical et segmentation de la parole. *Revue française de linguistique appliquée.* 2002;7:7-32

Baken R. Clinical measurement of speech and voice. Singular Publishing Group, San Diego. 2000

Bard EG, Shillcock RC, Altmann GTM. (). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context, *Perception and Psychophysics*, 1988;44 : 395-408

Barthelemy JP, Guénoche A. Les arbres et les représentations de proximités. Paris: Mason, 1988



- Barthélemy JP. Similitude, arbres et typicalité. In D. Dubois (Ed.), *Sémantique et cognition - Catégories, prototypes, typicalité*. Paris: CNRS, 1993 :205-224
- Bashford J, Reiner K, Warren R. Increasing the intelligibility of speech through multiple phonemic restorations. *Perception & psychophysics*. 1992;51:211-217
- Baylon C: *Sociolinguistique. Société, langue et discours*. Nathan. Beaud, 1991
- Beaudet C. Clarté, lisibilité, intelligibilité des textes : un état de la question et une proposition pédagogique ». *Recherches en rédaction professionnelle* 2001;1:1-17
- Benki JR. Analysis of English nonsense syllable recognition in noise. *Phonetica* 2003;60:129-157
- Benoît C, Mohamadi T, Kandel S. Effects of phonetic context on audio-visual intelligibility of French. *J Speech Hear Res*. 1994;37:1195-1203
- Benoît C, Guiard Marigny T, Legoff B, Adjoudani A. Which components of the face do humans and machines best speechread ? In *speechreading by humans and machines*. Stork & Hennecke (eds). Springer Verlag, Berlin. 1996, 351-372
- Beukelman DR, Yorkston KM. The relationship between information transfer and speech intelligibility of dysarthric speakers. *Journal of Communication Disorders*. 1979;12:189-196
- Beukelman DR, Yorkston KM. Influence of passage familiarity on intelligibility estimates of dysarthric speech. *J Commun Disord*. 1980;13:33-41
- Billman KS 1986 Phonological processes and intelligibility of spontaneous utterance in young children Unpublished master's thesis San Diego state university. San Diego California. cité par Kent (1992)
- Blanchet A, Gotman A. *L'enquête et ses methods. L'entretien*. Paris: Armand Colin, 2009:126p
- Blanchet PG, Snyder GJ. Speech rate treatments for individuals with dysarthria: a tutorial. *Percept Mot Skills*. 2010;110:965-982
- Bloch S, Wilkinson R. The understandability of AAC: A conversation analysis study of acquired dysarthria. *Augmentative and alternative communication*. 2004;20:272-282
- Bloch S, Wilkinson R. Acquired dysarthria in conversation: identifying sources of understandability problems. *Int J Lang Commun Disord*. 2009;44:769-783
- Borel S. *Phonétique*. Dans *Division palatine*, Veau V. Masson, Paris 1931 : 465-519
- Borrell A, & Billières M. L'évolution de la norme phonétique en français contemporain. *La Linguistique*. 1989;25: 45-62
- Borsky S, Shapiro LP, Tuller B. The temporal unfolding of local acoustic information and sentence context. *J Psycholinguist Res*. 2000;29:155-168
- Bronckhorst AW, Bosman AJ, Smoorenberg GF. A model for context effects in speech recognition. 1993;1:499-509.
- Bradlow AR, Torretta G M & Pisoni DB. Intelligibility of normal speech: global and fine grained acoustic phonetic talker characteristics. *Speech Communication*. 1996;20:255-272
- Bradlow AR, Nygaard LC DB, Pisoni D. Effects on talker, rate and amplitude variation on recognition memory for spoken word. *J. Acoust. Soc. Am*. 1999;61:206-219

Bressmann T, Sader R, Whitehill TL, Samman N. Consonant intelligibility and tongue motility in patients with partial glossectomy. *J. Oral Maxillofac. Surg.* 2004;62:298-303

Bunton K, Kent RD, Kent JF, Rosenbek JC. Perceptuo-acoustic assessment of prosodic impairment in dysarthria. *Clin Ling Phon.* 2000;14:13-24

Bunton K, Weismer G. The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *J. Speech Lang. Hear. Res.* 2001 ;44:1215-1228

Bunton K, Kent RD, Duffy JR, Rosenbek JC, Kent JF. Listener agreement for auditory-perceptual ratings of dysarthria. *J Speech Lang Hear Res* 2007; 50:1481-95

## C

Caron J. Précis de psycholinguistique. Paris: Presses universitaires de France; 2001

Carré R, Divenyi PL. Modeling and perception of 'gesture reduction. *Phonetica.* 2000;57:152-169

Chomsky N. Aspects of the theory of syntax. Cambridge: M.I.T. Press; 1965

Chomsky N. Théorie du gouvernement et du liage : les conférences de Pise. Paris: Editions du Seuil; 1991

Coleman J. Phonetic representations in the mental lexicon. In: Phonetics, phonology and cognition. Durand J, Laks B. Oxford UK: Oxford university press; 2002:128

Combescuré P. 20 listes de phrases phonétiquement équilibrées, *Revue d'Acoustique* 1981;56:34-38

Content A & Frauenfelder UH. La syllabe comme unité de perception de la parole : état de la question. Actes des XXIVèmes Journées d'Etude sur la Parole, Nancy, 2002

Cole AR Listening for mispronunciations: a measure of what we hear during speech. *Perception and psychophysics* 1973 1:153-156

Coupie C. Analyse des compensations articulatoires après glossectomie totale Mémoire d'orthophonie Toulouse, 1992

Crevier-Buchman L, Vaissière J, Maeda S, Brasnu D. [Intelligibility of French consonants after partial supra-cricoid laryngectomy]. *Rev Laryngol Otol Rhinol (Bord).* 2002;123:307-310

Crevier-Buchman L, Maeda S, Bely N, Laccourreye O, Vaissière J, Brasnu D. Compensations articulatoires après laryngectomie partielle supracricoïdienne avec cricothyoïdoépiglottopexie. *Ann Otolaryngol Chir Cervicofac.* 2001;118:81-88

Crochemore E, Vannier F. Analyse phonétique de la parole dysarthrique. In : Auzou P, Özsancak C, Brun V. Les dysarthries. Masson 2001 ; 2 : 71-82

Cutler A., Mehler J., Norris D. Segui J. The syllable's differing role in the segmentation of french and english. *Journal of Memory and Language,* 1986;25:385-400

Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 1988;14:113-121

## D

Daniloff RG, Shrinder TH, Zemlin WR. Intelligibility of vowels altered in duration and frequency. *J acoust Soc Am* 1968; 44:700-707

Darley F, Aronson A, Brown J. Clusters of deviant speech dimensions in dysarthrias. *Journal of Speech, Language, and Hearing Research: JSLHR*. 1969b;12:462-496

Darley F, Aronson A, Brown J. Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research: JSLHR*. 1969a;12:246-269

Darley F, Aronson A, Brown J. *Motor speech disorders*. W.B. SAUNDERS. COMPANY, Philadelphia, PA 1975

Duez D. On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops. *J Phonetics* 1995; 23:407-427

Duez D. Restoration of deleted and assimilated consonant sequences in conversational French speech: effects of preceding and following context 2001; 31: 101-114

D'Arcy Wentworth Thompson *On Growth and Form: The Complete Revised Edition* Cambridge 2000

De Bodt MS, Hernández-Díaz Huici ME, Van De Heyning PH. Intelligibility as a linear combination of dimensions in dysarthric speech. *J Commun Disord*. 2002;35:283-292

De Bodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997;11:74-80

Dejonckere Ph, Crevier-Buchman L, Marie JP, Moerman M, L Remacle M, Woisard V. Implementation of the European laryngological society (ELS) basic protocol for assessing voice treatment effect. *Rev Laryngol Otl Rhinol (Bord)* 2003;124:279-83

Dejonckere Ph, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van de Heyning V, Remacle M, Woisard V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *Eur Arch Otorhinolaryngol* 2001;258:77-82

Dejonckere PH, Remacle M, Fresnel-Elbaz E, Woisard V, Crevier L, Millet B. Reliability and clinical relevance of perceptual evaluation of pathological voices. *Rev Laryngol Otol Rhinol (Bord)*. 1998;119(4):247-248

Dejonckere Ph, Obbens C, de Moor GM, Wieneke Gh. Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatr* 1993;45:76-83.

De Letter M, Santens P, Van Borsel J. The effect of Levodopa on word intelligibility in Parkinson's disease. *J of Communication disorders* 2005;38:187-196

Delattre P. *L'Enseignement de l'R français*. *The French Review*. 1946;1:428-433

- Dell F. Les règles et les sons. Introduction à la phonologie générative. Hermann, Paris, 1985
- Dittner J, Lepage B, Woisard V, Kergadallan M, Boisteux K, Robart E\*and al. Elaboration et validation d'un test quantitatif d'intelligibilité pour les troubles pathologiques de la production de la parole. Rev Laryngol Otol Rhinol (Bord) 2010;131,1:1-4
- Dos Santos Barreto S, Zazo Ortiz K. Intelligibility measurement in speech disorders : a critical review of the literature. Pro-phono revistade atualizaçao Cientifica 2009;20:201-6
- Doyle PC, Swift ER, Haaf RG. Effects of listener sophistication on judgments of tracheoesophageal talker intelligibility. Journal of Communication Disorders. 1989;22:105-113
- Doyle PC, Leeper HA, Kotler AL, et al. Dysarthric speech: a comparison of computerized speech recognition and listener intelligibility. J Rehabil Res Dev. 1997;34(3):309-316
- Dubno JR, Levitt H. Predicting consonant confusions from acoustic analysis. JASA 1981 ;69 :249-261
- Duez D. Acoustic analysis of occlusive weakening in Parkinsonian French speech. International Congress of Phonetic Sciences, Saarbrücken : Allemagne 2007 [Internet]. [http://hal.archives-ouvertes.fr/hal-00244491\\_v1/](http://hal.archives-ouvertes.fr/hal-00244491_v1/)
- Duffy JR. Motor speech disorders: substrates, differential diagnosis, and management (2<sup>nd</sup> ed) Saint Louis, MO:Elsevier Mosby, 2005
- Duffy JR. History, current practice and future trends and goal. Dans Motor speech disorders, Essays for Ray Kent. Weimer G. 2007 Plural Publishing: San Diego :7-55
- Dupoux E, Pallier C, Kakehi K & Mehler J. New evidence for prelexical phonological processing in word recognition. Language and Cognitive Processes, 2001;5, 491-505
- Durand J. Les primitives phonologiques : des traits distinctifs aux elements. Dans Phonologie et phonétique : forme et substance, Nguyen N, Wauquier-Graveline S, Durand J. Paris : Lavoisier, 2005:65-93.
- Dumay N, Frauenfelder UH, Content A. The role of the syllable in lexical segmentation in French: word-spotting data. Brain Lang. 2002;81:144-161

## ***E***

- Enderby P. Frenchay Dysarthria assessment. College Hill Press. San Diego: 1983
- Escudier P. La parole : des modèles cognitifs aux machines communicantes. Paris: Hermès science publications; 2000

## ***F***

- FairbanksG, KodmanLa variabilité depend F. Worl intelligibility as a function of time compression. J acoust Soc Am 1957;29:636-641

Fauvet F, Schultz P, Debry C, Hirsch F, Sock R. Intelligibilité de la parole après glossectomie totale et réhabilitation orthophonique précoce. JEP Journées d'étude sur la parole, 2006

Fletcher H.. Speech and Hearing. Van Nostrand, New York, 1929

Flipsen P. Speaker-listener familiarity: parents as judges of delayed speech intelligibility. J of communication disorders 1995;28:3-19

Fodor J. La Modularité de l'esprit : essai sur la psychologie des facultés, collection Propositions, Les Éditions de Minuit, Paris, 1983.

Fónagy I. Le français change de visage. Revue Romane, 1989 ;24 :225-254

Fontan L, Tardieu J, Gaillard P, Woisard V, Legros C, Serrano E. Can intelligibility scores predict performance in real communication tasks ? Interspeech 2011, soumis

Fougeyrollas P. L'évolution conceptuelle internationale dans le champ du handicap : enjeux socio-politiques et contributions québécoises. Pistes, vol. 4, n° 2, 2002, 26 p., ill., réf. 3p., CAN <http://www.bdsp.ehesp.fr/Fulltext/309256/>

Fournier JE. Audiométrie vocale. Les épreuves d'intelligibilité et leurs applications au diagnostic , à l'expertise et à la correction prothétique des surdiés Paris: Maloine 1951

Frauenfelder UH, Kearns RK. Sequence monitoring. Language and Cognitive Processes 1996;11:665-673

Fry,D.B.,Abramson,AS, Eimas, PD, & Liberman, AM The identification and discrimination of synthetic vowels. Language & Speech, 1962, 5,171-189

Fujisaki, H., & Kawashima, T. A model of the mechanisms for speech perception quantitative analysis of categorical effects in discrimination. Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 1971;30: 59-68

Fujimura O. Syllables as concatenated demissyllables and affixes. Journal of Acoustical Society of America 1976;59, S55-S55.

## G

Gaillard MJ. Evaluation Clinique de la Dysarthrie (ECD) : analyse des organes. In : Auzou P, Özsancak C, Brun V. Les dysarthries. Masson 2001 ; 2 : 83-89

Gaillard P. Etude de la perception des transitoires d'attaque des sons de steeldrums : particularités acoustiques, transformation par synthèse et catégorisation. Musique, acoustique musicale et perception. Université de Toulouse II, 2000

Gaillard P. Castellengo M, Dubois D. Variabilité temporelle dans la perception des transitoires d'attaque des sons percussifs. Cahiers de la Société Québécoise de la recherche en musique. Numéro spécial, 2007

Garcia JM, Cannito MP. Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. J Speech Hear Res. 1996;39:750-760

Garric N & Calas F. Introduction à la pragmatique. Paris: Hachette; 2007

- Ganong, WF. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1980; 6:110-125
- Gatignol P, Tankere F, Clero D, Lobryau C, Soudant J, Lamas G. [Comparative evaluation of speech disorders and verbal and non verbal communication within two groups of patients: patients with facial paralysis (FP) and those who had undergone hypoglosso-facial anastomosis (HFA)]. *Rev Laryngol Otol Rhinol (Bord)*. 2003;124:291-297
- Gentil M. Phonetic intelligibility testing in dysarthria for the use of French language clinicians. *Clinical Linguistic & phonetics*. 1992;6:179-189
- Gerratt BR, Kreiman J. Measuring vocal quality with speech synthesis. *J. Acoust. Soc. Am.* 2001;110:2560-2566
- Giovanni A, Molines V, Nguyen N, Teston B, Robert D, Cannoni M, Pech A. A multiparameter method of computer-assisted objective vocal evaluation. *Ann Otolaryngol Chir Cervicofac* 1992;109:200-6
- Giovanni A, Robert D, Estublier N, Teston B, Zanaret M, Cannoni M. Objective evaluation of dysphonia : preliminary results of a device allowing simultaneous acoustic and aerodynamic measurements. *Folia Phoniatr* 1996 ; 48 : 175-185
- Goldberg E. *Les prestiges du cerveau ou comment l'esprit se bonifie avec l'âge* Paris Robert Laffont 2007 "334p
- Goldinger SD. Words and voices: episodic traces in spoken word identification and recognition memory. *J Exp Psychol Learn Mem Cogn*. 1996;22:1166-1183
- Goldinger SD, Kleider HM, Shelley E. The marriage of perception and memory: creating two-way illusions with words and voices. *Mem Cognit*. 1999;27:328-338
- Graesser A, Millis K, Zwaan R. Discourse comprehension. *Annual Review of Psychology*. 48:163-189
- Grataloup C, Hoen M, Meunier F, Pellegrino F, Veuillet E, Collet L. Cognitive Reconstruction of Reversed Speech in Normal and Dyslexic Adults. In proceedings of EuroCogSci 2007, the European cognitive science conference, Vosniadou, S., Kayser, D. & Protopapas, A. (eds), Lawrence Erlbaum Associates, 2007 :688-693
- Gordon-Salant S. Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *J. Acoust. Soc. Am.* 1986;80:1599-1607
- Gordon-Salant S, Fitzgibbons PJ. Temporal factors and speech recognition performance in young and elderly listeners. *J Speech Hear Res* 1993 36:1276-1285
- Grosjean F. The recognition of words after their acoustic offset: Evidence and implications, *Perception and Psychophysics*, 1985;38 : 299-310
- Grossberg S . Resonant neural dynamics of speech perception. *Journal of Phonetics* 2003;31:423-445
- Gremmo MJ, Holec H. La compréhension orale : un processus et un comportement. *Le Français dans le Monde, Recherches et Applications*. N° spécial Fév/Mars 1990
- Guastavino C. Categorization of environmental sounds. *Canadian Journal of experimental psychology*. 2007;61:54-63.

Guyot F. Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative : une approche par la catégorisation. Psychoacoustique, université du Maine, Paris, 1996

# H

Hallé P, Segui J, Frauenfelder U, Meunier C. The processing of illegal consonant clusters : A case of perceptual assimilation. *J of experimental psychology: Human perception and performance* 1998;24:592-608

Haton P. Lamotte M. Étude statistique des phonèmes et diphonèmes dans le français parlé. *Revue d'Acoustique* 1971;16 :

Hawkins S. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*. 2003;31:373-405

Hawkins S. Phonological features, auditory objects, and illusions. *Journal of Phonetics*. 2010;38:60-89

Hayes-Harbs R, Smith BL, Bent T, Bradlow AR. The interlanguage speech intelligibility benefit of native speakers of mandarin: production and perception of English word-final voicing contrasts. *J of phonetics* 2008;36:664-679

Hazan V, Simpson A. The effect of cue-enhancement on the intelligibility of nonsense word and sentence materials presented in noise. *Speech communication* 1998;24:211-226

Hazan V, Kim J, Chen Y. Audiovisual perception in adverse conditions: Language, speaker and listener effects *Speech Communication*, Volume 52, Issues 11-12, November-December 2010, 996-1009

Higginbotham DJ, Drazek AL, Kowarsky K, Scally C, Segal E. Discourse comprehension of synthetic speech delivered at normal and slow presentation rates. *Augmentative and Alternative Communication*. 1994;10:191-202

Hirano M. Psycho-acoustic evaluation of voice: GRBAS scale for evaluating the hoarse voice. *Clinical examination of voice* Sprieger Verlag, Wien 1981

Houde JF. There's more to speech perception than meets the ear. *Proc. Natl. Acad. Sci. U.S.A.* 2009;106:20139-20140

Houtgast T., Steeneken HJM. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Amer* 1985;77 :1069-1077

Hunter L, Pring T, Martin S. The use of strategies to increase speech intelligibility in cerebral palsy: an experimental evaluation. *Br J Disord Commun*. 1991;26:163-174

Hustad KC, Cahill MA. Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*. 2003;12:001-011

Hustad KC, Beukelman DR. Listener comprehension of severely dysarthric speech: effects of linguistic cues and stimulus cohesion. *Journal of Speech, Language, and Hearing Research*. 2002;45:545-558

Hustad KC, Dardis CM, McCourt KA. Effects of visual information on intelligibility of open and closed class words in predictable sentences produced by speakers with dysarthria. *Clin Linguist Phon.* 2007;21:353-367

Hustad KC. Effects of speech stimuli and dysarthria severity on intelligibility scores and listener confidence ratings for speakers with cerebral palsy. *Folia Phoniatr Logop.* 2007;59:306-317

Hustad KC. The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *J. Speech Lang Hear Res* 2008;51:562-573

Hugon B. Impact de l'environnement sur le signal de parole. In *Précis d'audioprothèse : production, phonétique acoustique et perception de la parole.* Renard X. Issy Les Moulineaux Elsevier Masson. 2008:309-327

Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive viewer for large phylogenetic trees, *BMC Bioinformatics.* 2007;22:460

## ***I***

Isaacs T. Towards Defining a Valid Assessment Criterion of Pronunciation Proficiency in Non-Native English-Speaking Graduate Students *The Canadian Modern Language Review / La revue canadienne des langues vivantes –* 2008; 64:555-580

## ***J***

Jakobson R. *Essais de linguistique générale*, Éditions de Minuit, Paris, 1963

Jilliot J, Vinet A et Le Her F. Impact de la perte auditive sur la perception de la parole. Altérations quantitatives. In *Précis d'audioprothèse : production, phonétique acoustique et perception de la parole.* Renard X. Issy Les Moulineaux Elsevier Masson. 2008:357-377

## ***K***

Kain AB, Hosom J, Niu X, van Santen JP, Fried-Oken M, Staehely J. Improving the intelligibility of dysarthric speech. *Speech Communication.* 2007;49:743-759

Kain A, Amano-Kusumoto A, Hosom J. Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility. *J. Acoust. Soc. Am.* 2008;124:2308-2319

Kayasith P, Theeramunkong T. Speech confusion index: A confusion-based speech quality indicator and recognition rate prediction for dysarthria. *Computers and Mathematics with Applications.* 2009;58:1534-1545



- Kearns KP, Simmons NN. Interobserver reliability and perceptual ratings: more than meets the ear. *J Speech Hear Res.* 1988;31:131-136
- Kent R, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. *J of Speech Hearing Disorders.* 1989;54:482-499
- Kent R. Intelligibility in speech disorders. Amsterdam/Philadelphia. John Benjamins; 1992
- Kent RD. Research on speech motor control and its disorders: a review and prospective. *J Commun Disord.* 2000;33:391-427
- Kintsch W. A cognitive architecture for comprehension. In: Pick HJ, van den Broek PW, editors. *Cognition: conceptual and methodological issues.* American Psychological Association; Washington, DC: 1992. pp. 143–163
- Kintsch W. *Comprehension : a paradigm for cognition.* Cambridge, MA: Cambridge university Press.1998
- Klatt DH. Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America* 1980;67 :971-995
- Knuuttila H, Pukander J, Maatta T, Pakarinen, Viikman. Speech articulation after subtotal glossectomy and reconstruction with myocutaneous flap. *Acta Otorhinolaryngologica.* 1999;119:621-626.
- Konstantinovic V.S, Dimic N.D. Articulatory function and tongue mobility after surgery followed by radiotherapy for tongue and floor of the mouth cancer patients. *Br J Plast Surg* 1998;51:589-93.
- Johnson-Laird PN. *Comprehension : Mental Models.* Cambridge, MA: Harvard university Press,1983
- Kerbrat-Orecchioni C. *Le discours en interaction.* Armand Colin. Paris: 2005
- Kerbrat-Orecchioni C. *La conversation. Le seuil "Mémo".* Paris: 1996
- Kerbrat-Orecchioni C. *L'énonciation de la subjectivité dans le langage.* Armand Colin. Paris: 1980
- Kim Y, Weismer G, Kent RD, Duffy JR. Statistical models of F2 slope in relation to severity of dysarthria. *Folia Phoniatr Logop.* 2009;61:329-335.
- Klatt DH. Software for a Cascade/Parallel Formant Synthesizer. *Journal of the Acoustical Society of America* 1980;67:971-995
- Kolinsky R. Spoken word recognition : a stage processing approach to language differences. *The European Journal of Cognitive Psychology* 1998;10:1-40
- Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. *J. Acoust. Soc. Am.* 2007;122:2354-2364
- Kreiman J, Gerratt BR. Sources of listener disagreement in voice quality assessment. *J. Acoust. Soc. Am.* 2000;108:1867-1876
- Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J. Acoust. Soc. Am.* 1998;104:1598-1608

Kreiman J, Gerratt BR. Perception of voice. In classics in voice and laryngology. Branski RC, Sulica L. Plural Publishing, San Diego 2009,585p

Kröger BJ, Kopp S, Lowit A. A model for production, perception, and acquisition of actions in face-to-face communication. *Cogn Process.* 2009 12;11:187-205

Kuhl PK. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept Psychophys.* 1991;50:93-107

## **L**

Labastrou E. A propos de sévère, ne sommes nous pas trop sévère ? *Pharmaterm.* 2000;11(3):1-8

Ladefoged P, Broadbent D. Informations conveyed by vowels J. *Acoust. Soc. Am.* 1957;29:98-104

Lado R. *Language Teaching: a scientific approach.* New York: McGraw-Hill, 1964

Lafon JC. *Le Test phonétique et la mesure de l'audition.* Dunod paris, 1964

Lahiri A, Reetz H. Underspecified recognition. In: C. Gussenhoven and N. Warner, Editors, *Labphon 7*, Mouton, Berlin, 2002: 637–676

Lahiri A, Reetz H. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics.* 2010;38:44-59

Lambour C. Évaluation de l'intelligibilité de la parole dans les établissements recevant du public sonorisés. Les méthodes normalisées sont-elles adaptées ? *Acoustique et techniques* 2002;29:9-16

Laures JS, Bunton K, Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions, *J Comm. Disord* 2003;36:449–464.

Laver J. *The gift of speech*, Edinburgh University Press., Edinburgh, UK. 1991

Lévy C. *Modèles acoustiques compactes pour les systèmes embarqués.* Thèse LIA informatique, université d'Avignon, 2006.

Lévy C, Linarès G, Nocéra P, Bonastre JP. Reconnaissance de chiffres isolés embarquée dans un téléphonie portable. Reconnaissance de chiffres isolés embarquée dans un téléphone portable. In *Journées d'études de la Parole*, JEP, Fez, Maroc, 2004

Li N, Loizou PC. Factors influencing glimpsing of speech in noise. *J Acoust Soc Am.* 2007;122:1165-1172

Lieberman AM, Delattre P, Cooper FS. The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants. *The American Journal of Psychology.* 1952;65:497-516

Liebermann A, Cooper E, Shankweiler D, Studdert-Kenedy M. The perception of the speech code. *Psychological review.* 1967;74:431-461

Lieberman AM, Mattingly IG. The motor theory of speech perception revised. *Cognition*. 1985;21:1-36.

Licklider JCR. Effects of amplitude distortion upon the intelligibility of speech. *J. Acoust. Soc. Amer.* 18;1946: 429-434

Licklider JCR. The intelligibility of amplitude dichotomized, time-quantized speech waves. *J. Acoust. Soc. Amer.* 1950;22:820-823

Lindblom B. Explaining phonetic variation: a sketch of the hypo- and hyperspeech theory. Dans: *Speech production and speech modelling*. Hardcastle WJ & Marchal A. Dordrecht: Kluwer Academic Publishers; 1990:403-439.

Liss JM, Spitzer SM, Caviness JN, Adler C, Edwards BW. Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *J. Acoust. Soc. Am.* 2000;107:3415-3424

Liss JM, Spitzer SM, Caviness JN, Adler C. The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *J. Acoust. Soc. Am.* 2002 ;112:3022-3030

Liss JM. Perception of dysarthric speech . In Weismer G (ed). *Motor speech disorders: essays for ray kent*. 2007 Plural Publishing: San Diego: 187-219

Luce PA, Pisoni DB. Recognizing spoken words: the neighbourhood activation model. *Ear and Hearing* 1998 ;19:1-36

Luce PA, Goldinger SD, Auer ET, Vitevitch MS. Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics* 2000;62:615-625

## *M*

Maassen B, Povel DJ. The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *J. Acoust. Soc. Am.* 1985;78:877-886

Maassen B, Povel DJ. The effect of correcting fundamental frequency on the intelligibility of deaf speech and its interaction with temporal aspects. *J. Acoust. Soc. Am.* 1984;76:1673-1681

McClelland JL, Elman JL. The TRACE model of speech perception. *Cogn Psychol* 1986;18:1-86

McQueen JM, Cutler A. Spoken word access processes: An introduction. *Language and Cognitive Processes* 2001;16:469-490

McQueen JM, Cutler A, Briscoe T, Norris D. Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes* 1995;10:309-331

McQueen JM, Otake T, Cutler A. Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language* 2001; 45:103-132

Maier A, Haderlein T, Eysholdt U, Rosanowski F, Batliner A, Schuster M, et al. PEAKS - A system for the automatic evaluation of voice and speech disorders. *Speech Communication*. 2009;51:425-437

- Maddieson I. Testing the universality of phonological generalizations with a phonetically-specified segment database: results and limitations. *Phonetica* 199;48:193-206.
- Maffiolo V. Caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain, Acoustique, Université du Maine, Paris, 1999
- Magnen C, Billières M, Gaillard P. Surdité phonologique et catégorisation. Perception des voyelles françaises par les hispanophones. *Revue Parole* 2005 ;33 :9-34
- Magnen C. Approche dynamique de la perception de la parole : catégorisation de la substance et de la variabilité phonétique en langue maternelle par les francophones et en langue étrangère par les hispanophones. Université de Toulouse II, Le Mirail, 2009.
- Malaurie P. L'intelligibilité des lois. *Pouvoirs* 2005 ; 3 :131-137
- Marin B, Legros D. Psycholinguistique cognitive Lecture compréhension et production de texte. De BoeckDehnière G & Baudet 1992. Lecture, compréhension de textes et science cognitive. Paris PUF
- Marslen-Wilson WD, Warren P. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review* 1994; 101:653-675
- Marslen-Wilson WD, Welsh A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 1978;10:29-63
- Massaro DW. Perceptual Units in Speech Recognition. *Journal of Experimental Psychology* 1974;102:349-353
- Massaro DW, Cohen MM. Categorical or continuous speech perception: A new test. *Speech Communication* 1983;2:15-35
- Mattys SL, White L, Melhorn JF. Integration of multiple speech segmentation cues: a hierarchical framework. *J Exp Psychol Gen.* 2005;134:477-500
- McNeil MR. Clinical speech management of Sensorimotor speech disorders . 2<sup>nd</sup> ed. Thieme, 2008
- Mehler J, Dommergues J, Frauenfelder UH, Segui, J. The syllable's role in speech segmentation. *Journal of verbal learning and verbal behavior* 1981;20 :298-305
- Ménard L, Schwartz JL, Aubin J. Invariance and variability in the production of the height feature in French vowels *Speech Communication* 2008;50:14-28
- Menendez-Pidal, X.; Polikoff, J.B.; Peters, S.M.; Leonzio, J.E.; Bunnell, H.T. The Nemours database of dysarthric speech. Dans Proc of ICSLP, Philadelphia PA 1996:1962 – 1965
- Metz DE, Schiavetti N, Sacco PR. Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutters. *J Speech Hearing Disorders* 1990;55:516-25
- Meunier F, Cenier T, Barkat M, Magrin-Chagnolleau, I. Mesure d'intelligibilité de segments de parole à l'envers en français, *Proceedings of 24th journée d'étude sur la parole* , 2002, 117-120
- Meunier C. Recovering missing phonetic information from allophonic variations. *Proceedings of the 14<sup>th</sup> international congress of phonetic sciences.* San Francisco, 1999: 845-848

Meunier C. Invariants et variabilité en phonétique. Dans Phonologie et phonétique : forme et substance, Nguyen N, Wauquier-Graveline S, Durand J. Paris : Lavoisier, 2005:349-374

Meyer J. Description typologique et intelligibilité des langues sifflées, approche linguistique et bioacoustique. 2005 Université Lumière Lyon 2

Meyer J. Intelligibilité de la parole à plusieurs distances dans un bruit naturel. 10<sup>ème</sup> congrès Français d'acoustique. Lyon Avril 2010. <http://hal.archives-ouvertes.fr/docs/00/55/09/02/PDF/000594.pdf>

Meyer G.; Morse R. The intelligibility of consonants in noisy vowel-consonant-vowel sequences when the vowels are selectively enhanced. *Speech Communication*, 2003;41:429-440

Miller GA, Licklider JCR. The intelligibility of interrupted speech. *J. Acoust. Soc. Am.* 1950;22:167-173

Miller JL, Liberman AM. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Percept Psychophys.* 1979;25:457-465

Miller JL. On the internal structure of phonetic categories: a progress report. *Cognition.* 1994 Juin;50(1-3):271-28

Moeschler J. & Reboul A. Compréhension, pragmatique et argumentation », in Sabah G. (éd.), Compréhension et langues et interaction. Paris, Hermès, 2006 :117-146

Moeschler J. Introduction à la linguistique contemporaine. 3 éd. Paris: Armand Colin; 2009

Morange S, Dubois D, Fontaine J. Perception of Recorded Singing Voice Quality and Expertise: Cognitive Linguistics and Acoustic Approaches. *J Voice.* 2010;24:450-457

Morton, J.M.: Interaction of information in word recognition. In: *Psychological Review* 1969, 76S:185-178

Mounin G. *Dictionnaire de la Linguistique.* PUF/Quadrages, rééd. 1993

Mucchielli A. les modèles de la communication. Dans « La communication : état en des savoirs. Editions Sciences Humaines 3<sup>ème</sup> édition. Cabin P. 2000 :65-78

## N

Nagafuchi M. Intelligibility of distorted speech sounds shifted in frequency and time in normal children; *Audiology* 1976 15:326-337

Nasir SM, Ostry DJ. Auditory plasticity and speech motor learning. *Proc. Natl. Acad. Sci. U.S.A.* 2009;106:20470-20475

Nguyen N. La perception de la parole. In Phonologie et phonétique : forme et substance, Nguyen N, Wauquier-Graveline S, Durand J. Paris : Lavoisier, 2005 :425-447

Norris D. Shortlist: A connectionist model of continuous speech recognition, *Cognition* 1994;52:189-234

Norris D, McQueen JM. Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev.* 2008;115:357-95

## O

Ohala JJ. Phonological evidence for top-down processing in speech perception. In *Invariance and variability in speech processes*, J. Perkell & D. Klatt. Hillsdale: Erlbaum. 1986: 386-397

Oudeyer P. The self-organization of speech sounds. *J. Theor. Biol.* 2005 ;233:435-449

Oudeyer P. The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies.* 2003;59:157-183

Ozsancak C, Parais AM, Auzou P. Analyse perceptive de la dysarthrie: présentation et validation d'une étude clinique : étude préliminaire. *Rev. Neurol. (Paris).* 2002;158:431-438

## P

Pallier C. Rôle de la syllabe dans la perception de la parole : études attentionnelles. Thèse de doctorat, EHESS, Paris, 1994

Parker M, Cunningham S, Enderby P, Hawley M, Green P. Automatic speech recognition and training for severely dysarthric users of assistive technology: the STARDUST project. *Clin Linguist Phon.* 2006 Mai;20:149-156

Pavlovic CV, Rossi M, Espesser R. Subjective assessment of acceptability, intelligibility and naturalness of text-to-speech synthesis, In *SIOA-1989, Vol.2*, 94-98

Peckels J.P, ROSSI M. Le test de diagnostic par paires minimales. *Revue d'acoustique* 1973;27: 245-262

Phatak SA, Lovitt A, Allen JB. Consonant confusions in white noise. *J. Acoust. Soc. Am.* 2008;124:1220-1233

Pisoni DB. The role of auditory short-term memory in vowel perception. *Memory. & Cognition*, 1975;3:7-18

Pisoni DB, Luce PA. Speech perception : reasearch, theory and the principal issues. Dans *Patterns recognition by humans and machines: speech perception*. Schwab EC et Nusbaum HC. New York:academic press1986 :1-50

Poitevineau J. Prendre ses distances : de l'usage raisonné (raisonnable) de quelques statistiques pour les arbres additifs. In D. Dubois (Ed), *Le sentir et le dire. Concepts et méthodes en psychologie et linguistique cognitives*. Paris: L'Harmattan 2009, 255-278

Pouchoulin G, Fredouille C, Bonastre JF, Ghio A, Giovanni A. Analyse phonétique dans le domaine fréquenteille pour la classification des voix dysphoniques. *JEP* 2008

Power MH, Braida LD. Consistency among speech parameter vectors: application to predicting speech intelligibility. *J. Acoust. Soc. Am.* 1996;100:3882-98

## R

Reboul A, Moeschler J. *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*, Paris, Armand Colin, 1998

Repp BH. Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychol Bull.* 1982;92:81-110

Rey A. *Dictionnaire historique de la langue Française. Le Robert*1998

Robert D, Bianco-Blache A, Spezza C, Verschuere A, Pouget J, Giovanni A. [Assessment of dysarthria and dysphagia in ALS patients]. *Rev. Neurol. (Paris).* 2006;162:445-453

Rosch E. Principles of categorization. In Rosch E, Lloyd BB. *Cognition and categorization.* 1978:27-48

Rosch E. Basic objects in natural categories. *Cognitive Psychology.* 1976;8:382-439

Rousseaux M, Dubuffet V, Tisne S et al. Evaluation de l'intelligibilité : présentation du Test Lillois de Dysarthrie (TLD). In : Auzou P, Özsancaç C, Brun V. *Les dysarthries.* Masson 2001;2:61-71

## S

Savin H & Bever T. The non-perceptual reality of the phoneme. *JVLVB* 1970,9:295-302

Sacks H, Schegloff E, Jefferson G. A simplest systematic for the organization of turn-taking for conversation. *Language* 1974;50:696-735

Sagi E, Svirsky MA. Information transfer analysis: a first look at estimation bias. *J. Acoust. Soc. Am.* 2008;123:2848-2857

Saitou N., & Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 1987,4(4), 406-425

Samar V, Metz D. Criterion validity of speech intelligibility rating-scale procedures for the hearing impaired population. *J Speech and Hearing Res.* 1988;31:307-316

Samuel AG. Phonemic restoration: insights from a new methodology. *J Exp Psychol Gen.* 1981;110:474-494

Samuel AG. Lexical activation produces potent phonemic percepts. *Cogn Psychol.* 1997 S32:97-127.

Samuel AG., & Pitt MA. Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language* 2003;48:416-434

- Sattah A. & Tversky A. Additive similarity Trees. *Psychometrika*, 1977;42, 19-345.
- Saussure F. *Cours de linguistique générale*. Paris: Payot; 1995.
- Schaeffer N, Eichorn N. The effect of differential vowel prolongations on perceptions of speech of naturalness. *J of fluency disorders* 2001;26:335-348
- Schiavetti N. Scaling procedures for the measurement of speech intelligibility. Dans: *Intelligibility in speech disorders*. Amsterdam/Philadelphia: John Benjamins; 1992. p. 11-34
- Schwartz JL, Boë LJ, Vallée N, Abry C. Major trends in vowel system inventories *Journal of Phonetics* 1997a;25:233-253
- Schwartz JL, Boë LJ, Vallée N, Abry C. The Dispersion-Focalization Theory of vowel systems *Journal of Phonetics* 1997b;25:255-286
- Scotto Di Carlo N. Contraintes de production et intelligibilité de la voix chantée. *Travaux interdisciplinaires du laboratoire Parole et Langage*, 2005 ;24 :159-179
- Searle R. *Les actes de langage*, 1972, Paris, Hermann (ed), rééd. 2009
- Segui J, Ferrand L. *Leçons de parole*. Odile Jacob. Paris: 2000
- Segui J, Dupoux E, Mehler J: The role of the syllable in speech segmentation, phoneme identification and lexical access. In Altmann GTM (ed.), *Cognitive models of speech processing, psycholinguistic and computational* Cambridge, Mass, MIT Press. 1990:263-280
- Shannon C. A mathematical theory of communication. *Bell System Technical Journal*. 1948;27:379-423,623-656
- Shannon RV, Zeng FG, Wygonski J., Kamath V, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270,303-304
- Schaeffer N, Eichorn N. The effect of differential vowel prolongations on perceptions of speech of naturalness. *J of fluency disorders* 2001;26:335-348
- Schegloff, G. Jefferson AH. Sacks. The preference for self- correction in the organization of repair in conversation, *Language* 1977;53:361-382
- Schegloff EA. *Sequence Organization in Interaction: A Primer in Conversation Analysis, Volume 1*, Cambridge: Cambridge University Press, 2007
- Shiller DM, Sato M, Gracco VL, Baum SR. Perceptual recalibration of speech sounds following speech motor learning. *J. Acoust. Soc. Am.* 2009 Fév;125:1103-1113
- Stelzle F, Maier A, Nöth E, Bocklet T, Knipfer C, Schuster M, et al. Automatic Quantification of Speech Intelligibility in Patients After Treatment for Oral Squamous Cell Carcinoma. *Journal of Oral and Maxillofacial Surgery* 2011 In Press, Corrected Proof. Available from: <http://www.sciencedirect.com/gate1.inist.fr/science/article/B6WKF-51WV6WX-7/2/33451fe3b63b2fd4454b9490043318ac>
- Schuster M, Maier A, Haderlein T, Nkenke E, Wohlleben U, Rosanowski F, et al. Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition. *Int. J. Pediatr. Otorhinolaryngol.* 2006;70:1741-1747



Singer M. Psycholinguistics: Discourse comprehension. In: Kazdin AE, editor. Encyclopedia of psychology. Vol. 6. American Psychological Association; Washington, DC: 2000. pp. 269–372

Southwood M, Weismer G. Listener judgments of the bizarreness, acceptability, naturalness, and normalcy of the dysarthria associated with amyotrophic lateral sclerosis. *J Med Speech Lang Pathol* 1993;1:151-161

Sperber D, Wilson D. *La pertinence : Communication et cognition*. Paris: Les Editions de Minuit; 1989

Stevens S. *Psychophysics*. New York: Wiley; 1975

Stevens KN. Evidence for the role of acoustic boundaries in the perception of speech sounds. In: Fromkin, *Phonetic Linguistics: Essays in the honor of Peter Ladefoged*, Academic Press, Orlando 1985

Stevens KN, Blumstein S.E. Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 1978;64, 1358-68

Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. *J acoust Soc Am* 2002;111:1872-1891

## *T*

Tabossi P, Collina S, Mazzetti M, Zopelo M : Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance* 2000;26:758-775

Tapiero I. *Situation Model and level of coherence: toward a definition of comprehension*. Hillsdale, Lawrence Erlbaum associates, 2007

Terai H, Shimahara M. Evaluation of speech intelligibility after a secondary dehiscence operation using an artificial graft in patients with speech disorders after partial glossectomy. *British Journal of Oral and Maxillofacial Surgery*. 2004;42:190-194

Teilhard de Chardin P. *Le Groupe zoologique humain. Structures et directions évolutives*, Albin Michel, 1956

Tjaden KK, Liss JM. The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics and Phonetics*. 1995;9:139–154

Tjaden K, Wilding GE. Rate and loudness manipulations in dysarthria: acoustic and perceptual findings. *J. Speech Lang. Hear. Res.* 2004;47:766-783

## *V*

Van Dommelen WA, Hazan V. Perception of English consonants in noise by native and Norwegian listeners. *Speech Communication* 2010;52:968-979

Van Dijk T, Kintsch W. Strategies of discourse comprehension. New York: Academic Press; 1983

Van Nuffelen G, De Bodt M, Guns C, Wuyts F, Van de Heyning P. Reliability and clinical relevance of segmental analysis based on intelligibility assessment. *Folia Phoniatr Logop.* 2008;60:264-268

Van Nuffelen G, Middag C, De Bodt M, Martens J. Speech technology-based assessment of phoneme intelligibility in dysarthria. *Int J Lang Commun Disord.* 2009;44:716-730

Van Riper C, Emerik L. Speech correction. Englewood Cliffs: Prentice Hall; 1984

Van Wieringen A, Pols LC. Frequency and duration discrimination of short first-formant speechlike transitions. *J. Acoust. Soc. Am.* 1994;95:502-511

## W

Wakumoto M, Ohno K, Imai S, Yamashita Y, Akizuki H, Michi K. Analysis of the articulation after glossectomy. *Journal of oral rehabilitation* 1996;23:764-770

Warren DW, Dalston RM, Mayo R. Hypernasality and velopharyngeal impairment. *The cleft palate Journal* 1990;27:53-60

Warren RM. Auditory illusions and perceptual processing of speech. Dans *Principles of experimental phonetics*. Lass JL Moby-Year Book, Saint Louis 1996: 435-469

Warren R. Perceptual restoration of missing speech sounds. *Science* 1970;167:392-393

Warren R, Obusek C. Speech perception and phonemic restorations. *Perception & psychophysics.* 1971;9:358-362

Walshe M, Peach R, Miller N. The dysarthria Impact Profile: Validation of an instrument to measure psychosocial impact in dysarthria. *International journal of language and communication disorders.* 2009;44:693-715

Weismer G, Liss JM. Age end speech motor control *Handbook on aging and communication* D Ripich San Diego : college Hill press, 1991

Weismer G, Laures JS. Direct magnitude estimates of speech intelligibility in dysarthria: effects of a chosen standard. *J. Speech Lang. Hear. Res.* 2002;45:421-433

Weimer G, Martin RE. Acoustic and perceptual approaches to the study of intelligibility. Dans: *Intelligibility in speech disorders*. Amsterdam/Philadelphia: John Benjamins; 1992 : 68-118

Weismer G, Jeng JY, Laures JS, Kent RD, Kent JF. Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatr Logop.* 2001;53:1-18

Weismer G. *Motor speech disorders: essays for ray kent*. Plural Publishing: San Diego, 2007

Weisz F. *Etude franco-italienne de l'évaluation perceptive de la sévérité de la dysphonie*. Ecole d'orthophonie de Marseille, 2010

Whitehill TL, Ciocca V. Perceptual-phonetic predictors of single-word intelligibility: a study of Cantonese dysarthria. *J. Speech Lang. Hear. Res.* 2000;43:1451-1465

- Whitehill TL. Assessing intelligibility in speakers with cleft palate: a critical review of the literature. *Cleft Palate Craniofac. J* 2002;39:50-58
- Whitehill TL, Ciocca V, Chan JC, Samman N. Acoustic analysis of vowels following glossectomy. *Clin Linguist Phon.* 2006;20:135-140
- Whitehill TL. Assessing intelligibility in speakers with cleft palate: a critical review of the literature. *Cleft Palate Craniofac. J.* 2002;39:50-58
- Whitehill TL. Studies of Chinese speakers with dysarthria: informing theoretical models. *Folia Phoniatr Logop* 2010;62:92-96
- Wieringen, A. van & Pols, L.C.W. Discrimination of single and complex consonant-vowel and vowel-consonant-like formant transitions'. *J. Acoust. Soc. Am.* 1995;98:1304-1312
- Williams SE, Watson JB. Differences in speaking proficiencies in three laryngectomees groups. *Arch Otolaryngol.* 1985;111:216-219
- Williams FD, Dietrich S. Effects of speech and language disorders on raters' perception. *J communication disorders* 1996;29:1-12
- Williams FD, Dietrich S. Perceptions of communicative disorders. Vérification and specification of rater variables. *J of communication disorders* 2001;34:355-366
- Wilson RH, Preece JP, Slamon DL, Sperry JL, Boerstein SP. Effects of time compression and time compression plus reverberation on the intelligibility of Northwestern University Auditory test No6. 1994 *Jam Acad Audil* 5:269-277
- Winograd T. *Understanding Natural Language.* New York Academic Press, 1972
- Wurm LH; Vakoch DA; Seaman SR. Recognition of spoken words: semantic effects in lexical access. *Language and speech* 2004;47:175-204
- Wuyts FL, De Bodt MS, Molenberghs G, Remacle M, Heylen L, Millet B, Van Lierde K, Raes J, Van de Heyning PH. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res.* 2000;43:796-809

## Y

- Yorkston KM, Beukelman DR. A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *J Commun Disord* 1980;13:15-31
- Yorkston KM, Beukelman DR. *Assessment of Intelligibility in Dysarthric Speakers.* Tigard OR: CC publications, 1981
- Yorkston KM, Dowden PA, Beukelman DR. Intelligibility measurement as a tool in the clinical management of dysarthric speakers. Dans: *Intelligibility in speech disorders.* Kent R D. Amsterdam/Philadelphia: John Benjamins 1992:265-286
- Yorkston KM, Strand EA, Kenedy MRT. Comprehensibility of dysarthric speech: implications for assessment and treatment planning. *Am J Speech Language Pathology* 1996;5:55-66
- Yorkston KM, Beukelman DR, Tice TR. *Sentence intelligibility test.* Lincoln, NE: Tice technology Services, 1996

Yorkston KM, Beukelman DR, Strand EA, Bell KR. Management of Motor Speech Disorders in Children and Adults. 2<sup>nd</sup> édition. Pro-ed. Austin, Texas, 1999

Yorkston KM, Baylor CR, Deitz J., Dudgeon BJ, Eadie T, Miller RM, Altmann DA. Satisfaction with communicative participation as defined by adults with multiple sclerosis. J of communication disorders 2007,40:433-451

Yorkston KM, Baylor CR, Deitz J., Dudgeon BJ, Eadie T, Miller RM, Altmann DA. Developing a scale of communicative participation: a cognitive interviewing study. Disability and rehabilitation 2008,30:425-33

Yu P, Ouakine M, Revis J, Giovanni A. Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements. J Voice 2001;15:529-42

## Z

Ziegler W, Zierdt A. Telediagnostic assessment of intelligibility in dysarthria: a pilot investigation of MVP-online. J Commun Disord. 2008;41:553-577

Zielinski BW. The listener: No longer the silent partner in reduced intelligibility. System 2008;36:69-84

Zyski BJ, Weisiger BE. Identification of dysarthria types based on perceptual analysis. J of communication disorders 1987;20:367-378.

Zuliani P. Intelligibilité de la parole dans des conditions d'écoute difficiles, Thèse de Doctorat de l'Université de Toulouse III, 1988

## 8.2. LOGICIELS

---

### PRAAT

Boersma P & Weenick D 2009 : Praat : doing phonetics by computer version 5.1.0 computer program] Retrieved May 1, 2009, from <http://www.praat.org/>

### TCL-LABX

Pascal Gaillard sur <http://w3.octogone.univ-tlse2.fr/lordat/p-gaillard/tcl-labx/>

### DENDROSCOPE

Daniel H Huson, Daniel C Richter, Christian Rausch, Tobias Dezulian, Markus Franz and Regula Rupp. Dendroscope: An interactive viewer for large phylogenetic trees sur <http://www-ab.informatik.uni-tuebingen.de/software/dendroscope>

## PARTITION AND ADDTREE

Partition & AddTree software Tree calculation, from J. Poitevineau

[Partition and AddTree.zip](#)

## PERCEVAL

Ghio A., André C., Teston B., Cavé C., (2003) "PERCEVAL: une station automatisée de tests de PERception et d'EVALuation auditive et visuelle", [TIPA](#), Aix-en-Provence, France, vol. 22, p. 115-133

<http://aune.lpl.univ-aix.fr/~lpldev/perceval/percevaldownload.html>

### 8.3. ADRESSE DES SITES WEB

---

<http://www.lexique.org>

Site réalisé par Boris New et Eric Pallier, hébergé par pe RISC (Relais d'information sur les sciences de la cognition, UMS 332)

<http://mesh.inserm.fr/mesh/index.htm>.

MESH

<http://www.has-santé.fr>

Haute Autorité de Santé

[http://www.ctnerhi.com.fr/accueil\\_ctnerhi3.php](http://www.ctnerhi.com.fr/accueil_ctnerhi3.php)

Centre Technique National d'Etudes et de Recherches sur les Handicaps et les Inadaptations : CTNERHI 236 bis rue de Tolbiac – 75013 Paris

<http://www.bdsp.ehesp.fr/Fulltext/309256/>

École des hautes études en santé publique (EHESP) Banque de données en santé publique (BDSP) Avenue du Professeur Léon Bernard 35043 Rennes

<http://www.who.int/classification/icf>

Organisation mondiale de la santé