



Les états émotionnels et la Prosodie : paradigmes, modèles, paramètres

Geneviève Caelen-Haumont

► To cite this version:

Geneviève Caelen-Haumont. Les états émotionnels et la Prosodie : paradigmes, modèles, paramètres. Nguyen, Noël. Phonologie et phonétique: Forme et substance, Hermès, pp.397-424, 2005. <hal-00142950>

HAL Id: hal-00142950

<https://hal.science/hal-00142950v1>

Submitted on 23 Apr 2007

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

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



HAL Authorization

Chapitre 15

Les états émotionnels et la prosodie : paradigmes, modèles, paramètres

15.1. Introduction

Avec le domaine des émotions, nous sommes au cœur des processus humains à la fois les plus intimes et les plus universels, et de ce fait, les plus variables et les plus complexes. Nos comportements, nos états mentaux, nos activités intellectuelles, nos analyses scientifiques même sont constamment imprégnés par nos états émotionnels, au point que nous sommes confrontés à une alternative : idéaliser le principe d'objectivité en élaguant toute cause ou trace de subjectivité, mais en prenant le risque de se couper de la réalité du vivant, et donc de n'accéder qu'à une vue partielle des choses, ou bien prendre comme telles les différentes dimensions de cette réalité mais renoncer alors aux exigences d'une pure objectivité.

Ce débat a intéressé sinon « passionné » les hommes depuis l'antiquité car ce sont non seulement les processus fondamentaux (raison, émotions, sensations) de notre identité (esprit, âme, corps) qui sont ainsi posés, ainsi que leurs relations, mais aussi les conditions d'accès à la connaissance, et son statut. Au cours des millénaires, l'histoire des idées n'a eu de cesse d'ailleurs de redistribuer les rôles et les dominances de ces composantes. Notre époque a le privilège d'aborder le domaine avec un héritage culturel, des moyens et des outils jamais égalés jusque là, ce qui nous autorise précisément, en ayant les moyens de contrôler notre subjectivité, à mieux l'explorer. Et de fait nous constatons que le champ des émotions s'est considérablement développé ces dernières années, que nombreux sont

désormais les colloques sur l'émotion, que nombreuses sont les disciplines qui les étudient, et tout autant leurs applications avec par exemple les technologies de la parole qui ouvrent un créneau commercial considérable.

Après un rapide rappel des conceptions des philosophes et hommes de sciences qui ont prévalu jusqu'à l'aube du 20^{ème} siècle, nous présenterons les différents paradigmes en cours à l'heure actuelle, puis nous aborderons les principaux champs de recherches en matière de parole affective et de prosodie. Pour les aspects linguistiques de la prosodie, voir Di Cristo, ce volume.

Les analyses essentielles de Descartes [DES 1649] en la matière, exposées dans son ouvrage *Des passions de l'âme*, contiennent les fondements des positions philosophiques qui vont prévaloir pendant plusieurs siècles, jusqu'aux travaux de Freud, à savoir le fameux principe de la dualité corps/esprit, en vertu duquel il existe peu de relations entre ces deux entités. Ceci a en particulier l'effet de cloisonner les émotions en les privant d'une relation au corps. Comme ses contemporains, Descartes propose par ailleurs une liste des *émotions* fondamentales ou *primaires* (*l'admiration, l'amour, la haine, le désir, la joie, la tristesse*) à partir desquelles sont composées les plus complexes. Comme Platon, Descartes n'envisage ces états affectifs que de manière négative, voyant en eux un obstacle au fonctionnement clair et objectif de la raison.

C'est encore cette position qui prévaut chez Darwin [DAR 1872]. Toutefois Darwin a été l'un des initiateurs de l'analyse des expressions vocales de l'émotion, qu'il a comparées avec les cris des animaux dans des réactions adaptatives. Il considère que certaines émotions qui survivent n'ont plus de fonction réelle à notre époque, et il voit ces expressions vocales de l'émotion comme des stigmates de fonctions d'états pré-humains. Rapprochant le bébé et le proto-humain, il développe de nouveaux outils conceptuels, la phylogénèse et l'ontogénèse, mais ce faisant, le parallélisme qu'il établit entre ces deux états, le bébé humain et l'adulte pré-humain, n'est pas valide. À l'heure actuelle, comme les nombreuses études le montrent dans le domaine de l'acquisition et du développement, la prosodie est sans doute du point de vue de la perception comme de la production, l'un des tout premiers systèmes de communication que mettent en œuvre les bébés (voir à ce sujet Wauquier-Gravelines, ce volume).

La vision d'Aristote est au contraire largement positive, et sa pensée est toujours aussi d'actualité. En effet Aristote élabore les premiers fondements d'une dimension cognitive. Bien loin de subir les états émotionnels, l'homme les utilise pour construire son univers personnel de références. Ainsi non seulement il pose que nos sentiments subjectifs reposent sur nos connaissances personnelles et nos croyances, mais il établit aussi une *valence* des émotions, distinguant ainsi parmi les états, ceux qui sont perçus comme plaisants ou agréables, et les désagréables. Il perçoit aussi

l'émotion sous la perspective de l'action et des comportements qu'elle suscite, et finalement il pose que les émotions ressenties dans les situations particulières vont affecter nos jugements et les actions qui en découlent. Près de 25 siècles avant nous, il est ainsi captivant de constater que pour Aristote, les émotions possèdent des déterminants cognitifs et produisent des effets sur le fonctionnement du cerveau, ce qui rejoint nos conceptions les plus récentes dans le domaine !

Il revient à Damasio [DAM 94], neurologue à l'Université de l'Iowa, de poser les limites de ce dualisme cartésien en stigmatisant « l'erreur de Descartes : il a instauré une séparation catégorique entre le corps, fait de matière, doté de dimensions, mû par des mécanismes, d'un côté, et l'esprit, non matériel, sans dimensions et exempt de tout mécanisme, de l'autre ; il a suggéré que la raison et le jugement moral ainsi qu'un bouleversement émotionnel ou une souffrance provoquée par une douleur physique, pouvaient exister indépendamment du corps. Et spécifiquement, il a posé que les opérations de l'esprit les plus délicates n'avaient rien à voir avec l'organisation et le fonctionnement d'un organisme biologique. »

La perspective de Damasio et avec lui tout un courant issu des années 60 (voir ci-dessous), pose en fait le problème non plus sous l'angle réducteur de l'opposition corps/esprit, mais instaure les émotions dans les sciences de la cognition comme des états de première importance. Les études et expériences dans le domaine de la neurophysiologie, de la neuropathologie, l'investigation des aires cérébrales, les derniers développements dans le domaine de l'imagerie cérébrale, de la psychologie, de la linguistique, de la prosodie, des outils informatiques, révèlent l'imbrication extrêmement complexe des dimensions conceptuelles, affectives et corporelles. Isoler une de ces dimensions revient à se méprendre sur le fonctionnement des autres. L'enjeu est alors de faire converger les études de ces diverses sciences ou techniques pour étudier le processus cognitif dans toute son étendue et pour échafauder une véritable théorisation du langage et de la parole.

15.2. Les paradigmes actuels sur les états émotionnels

Pour prendre en compte totalement le phénomène de l'émotion, et dans la continuation des recherches précédentes, il faut l'envisager sous trois angles, c'est-à-dire comme des systèmes de réponses incluant les *réactions physiologiques*, les *réactions comportementales et expressives*, et les *réactions cognitives et expérientielles*. La recherche actuelle insiste sur l'évaluation cognitive de la situation qui détermine à son tour le type d'état émotionnel ressenti et son intensité, et sur l'état de préparation mentale à l'action qui précède de manière inconsciente la réponse comportementale, comme par exemple l'attaque en cas de colère.

Il semble à l'heure actuelle que le consensus ne soit pas atteint parmi les spécialistes, de sorte que trois paradigmes se partagent aujourd'hui le champ d'analyse des états affectifs. Le premier axe oppose les tenants de l'approche universaliste de l'émotion et les tenants de l'approche différentialiste qui mettent en avant le principe culturel. Un autre paradigme infiltre toutefois toutes les tendances : le paradigme cognitiviste. De fait dans le domaine de la parole, c'est ce dernier courant qui rassemble le plus grand nombre de travaux.

15.2.1. *À la suite de Darwin, le paradigme universaliste*

Ce paradigme se constitue autour d'Ekman [EKM 76, 84], psychologue à l'Université de Californie, chef de file d'un courant de recherches bien implanté. Il considère ainsi qu'au cours de l'histoire humaine, les émotions ont évolué pour répondre de manière adaptative à l'évolution des conditions de vie et de survie. Elles se seraient ainsi organisées en patrons de réponse adaptatifs à certains événements. De la sorte certaines émotions de base seraient présentes de manière universelle [EKM 84]. Il en veut pour preuve que le processus d'évaluation cognitive qui précède l'état affectif, du fait qu'il est automatique, échappe par là-même à l'emprise d'un contexte social ou culturel.

Shaver et ses collègues [SHA 92] ont montré dans une étude menée sur le sens des différentes émotions en italien, chinois et anglais américain, qu'il existe des universaux interculturels, déterminés par l'évolution.

Le secteur biologique fournissant sans doute les arguments les plus aisés à ce paradigme, Ekman souligne le caractère adaptatif des émotions par le déclenchement extrêmement rapide de la réponse, donnant à l'individu la capacité de mobiliser sur le champ toutes les ressources de son organisme pour faire face à une situation donnée. Toutefois, malgré l'accumulation des travaux dans son laboratoire, Ekman reste contesté dans deux domaines : celui des profils physiologiques distincts pour les émotions telles que la peur, le dégoût ou la colère, et celui du signal distinctif de l'émotion. Ainsi un signal universel et identique comme le sourire peut traduire des états émotionnels très différents (joie, ironie, fierté, soulagement etc.). Par ailleurs, Ekman et son groupe s'appuient naturellement sur des données physiologiques et comportementales expressives, mais ce faisant, ils ne prennent pas en charge les aspects expérientiels cognitifs. Finalement, à la suite de Darwin, la conception de l'émotion reste éthologique.

15.2.2. *La prééminence du culturel et le constructivisme social*

Psychologue à l'Université du Massachusetts, Averill conteste la prééminence du biologique défendue par Ekman, et s'attache à montrer [AVE 80], avec Rom Harré [HAR 86], mais avec l'appui aussi d'anthropologistes et de philosophes, les limitations de l'approche universaliste. Tout d'abord, il argumente sur le fait que les émotions complexes, encore fort peu étudiées, telles que l'espoir, la fierté, la culpabilité..., nous informeraient davantage en psychologie sur le phénomène de l'émotion que les émotions de base. Ensuite, il souligne avec justesse que dans les différentes cultures, il n'existe pas de correspondances parfaites entre les états émotionnels observables ni entre les mots qui les traduisent d'une langue à une autre. De la sorte, il existe des langues, qui à la différence d'autres, n'ont pas lexicalisé certains états émotionnels, d'où une impossibilité de traiter de manière universaliste les émotions. Reste à vérifier cependant que l'absence d'universalité de l'expression aille effectivement de pair avec une absence d'universalité du ressenti.

Averill montre aussi qu'une émotion vécue d'un point de vue social ou psychologique peut être tout aussi vitale. Ainsi, psychologiquement, les émotions les plus fortes rapportées par les individus concernent le plus souvent les grandes épreuves rencontrées qui les ont soit brisés, soit au contraire renforcés.

En fait pour les constructivistes sociaux, selon Cornelius [COR 00], la manière dont les personnes expriment leurs émotions en fonction de leur sexe ou de leur groupe social, révèle en fait la façon dont les émotions sont construites dans une culture donnée pour servir des objectifs sociaux déterminés. Même si le processus de l'évaluation peut être une adaptation biologique, le culturel détermine le contenu des évaluations d'où procèdent les émotions. Ainsi pour Spellman [SPE 89], une part de la fonction sociale des émotions sert à réguler le comportement : par exemple la colère régule les relations interindividuelles en établissant une frontière entre ce qui est permis et ce qui ne l'est pas. Si la colère n'est pas par exemple une expression autorisée dans certaines cultures pour les femmes, dans cette perspective, on lira ce comportement comme le symptôme de leur assujettissement.

15.2.3. *Le paradigme cognitiviste*

Ce paradigme qui trouve ses origines les plus lointaines chez les philosophes grecs, en particulier chez Aristote, est actuellement dominant pour ce qui concerne l'analyse des émotions. C'est sans doute l'ouvrage d'Arnold [ARN 60] qui, en matière d'émotions, ouvre la voie à l'analyse cognitive. Le concept central de ce courant de recherches est que *la pensée et l'émotion sont inséparables*. Plus précisément, selon Arnold, toutes les émotions mettent en œuvre une *évaluation* des événements en vertu de laquelle ils nous apparaissent comme positifs ou négatifs.

Pour les cognitivistes, cette notion d'évaluation est fondamentale. En effet à chaque type d'émotion, correspond un modèle différent d'évaluation. Comme le décrit Cornelius [COR 00], ces modèles représentent le lien entre les caractéristiques de la personne, son histoire, ses apprentissages, son tempérament, sa personnalité, son état physiologique et celles de la situation présente où il se trouve.

Dans la mesure où le processus d'évaluation informe l'organisme de la présence dans la situation de caractéristiques particulières, et le dispose à y réagir, Frijda [FRI 86] pose que les émotions sont des *tendances à l'action*. Selon Arnold [ARN 60], ces évaluations sont « directes, immédiates, non réfléchies, non intellectuelles, automatiques. »

Les chercheurs de tradition cognitiviste ont dépensé beaucoup d'énergie à décrire les dimensions primitives des évaluations qui sous-tendent les émotions et les différencient, telles que le côté agréable ou désagréable, l'effort, le contrôle, la responsabilité. Par exemple pour Smith et Ellsworth [SMI 85], si colère et culpabilité mettent toutes deux en jeu une évaluation déplaisante de la situation, elles s'opposent dans la mesure où la première met en cause la responsabilité d'une autre personne, la production d'un effort important, et la deuxième, la responsabilité propre de la personne en vertu de la culpabilité ressentie, cette sensation intérieure ne réclamant que peu d'effort ou d'énergie.

Scherer [SCH 87] s'est consacré à l'étude des processus de l'évaluation dans ses différentes composantes, montrant la présence dans toute émotion d'une phase invariante au cours de laquelle des contrôles (*stimulus evaluation checks*) sélectionnent certains types d'informations concernant les événements générateurs d'émotions, et en particulier ceux qui sont porteurs de nouveauté et de plaisir.

D'autres théoriciens comme Levenson *et al.* [LEV 90], considèrent les émotions comme des processus de programmes affectifs qui activent différents sous-systèmes de réponses, à savoir les sous-systèmes expressifs, moteurs et expérientiels. L'activation de l'un d'entre eux entraîne celle des autres. Levenson [LEV 92] montre que non seulement la mémoire peut être un stimulus engendrant l'émotion, mais aussi d'autres systèmes jusque-là envisagés comme des réponses (par exemple les expressions faciales, les vocalisations...). Et de fait n'importe quel composant de l'émotion pourrait jouer ces rôles symétriques. Ainsi, selon Cornelius [COR 00], la prosodie peut être considérée à la fois comme un de ces systèmes activables par un des programmes affectifs, et symétriquement comme un programme pouvant activer les autres systèmes affectifs.

15.3. Précisions, définitions, repères

Comme le précise justement Cowie [COW 00], il y a lieu d'abord de repérer deux emplois du terme *émotion*, emplois qui coïncident avec les formes du mot au singulier et au pluriel. Ainsi au pluriel, le terme réfère aux émotions pures, avec leurs caractéristiques propres (*ensemble de traits*) et leurs frontières, telles que la peur, la joie, la colère etc. Au singulier, le terme renvoie à l'attribut de certains états psychologiques : par exemple on entend dire souvent que telle personne a parlé avec émotion. Il désigne donc un *état émotionnel*. Enfin pour rendre compte des développements actuels de la recherche, il convient d'ajouter l'expression *systèmes émotionnels*, pour une application biologique.

Depuis Descartes, il est d'usage par ailleurs de distinguer des autres les *émotions de base*, généralement au nombre de six (peur, colère, joie, tristesse, surprise, dégoût), auxquelles on ajoute parfois l'amour et le mépris, et deux déclinaisons de la colère (« chaude » et « froide »). Les états émotionnels, en dehors de ces émotions primaires, sont très diversifiés, et Plutchik [PLU 80] par exemple en a répertorié jusqu'à 142. Ces états émotionnels correspondent à ce que l'on désigne parfois aussi comme *émotions secondaires*. Ces états émotionnels sont complexes, dans la mesure où ils combinent souvent plusieurs caractéristiques des émotions de base.

Dans le domaine de l'affectivité, un concept voisin de celui d'émotion, est celui d'*attitude*, fréquemment utilisé dans les études linguistiques. Il est très difficile de distinguer par des traits définitoires les deux notions. Ces attitudes sont encore plus nombreuses que les états émotionnels, et O'Connor et Arnold [O'C 73] par exemple, en utilisent près de 300, parmi lesquelles on peut citer par exemple : « abrupt, accusateur, affable, affecté, approbateur, arrogant », etc.

Wichman [WIC 00] propose de manière très intéressante une analyse comparative du concept « attitude » en fonction des divers domaines de recherches (psychologie sociale, linguistique, pragmatique) et montre que les acceptions sont parfois divergentes. La première tentative de définition du terme est proposée par Couper-Kuhlen [COU 86], qui oppose ainsi l'émotion comme état du locuteur, à l'attitude comme type de comportement.

Mais ceci laisse de côté les attitudes cognitives répertoriées par les psychologues sociaux qui sont fonction d'opinions, de croyances ou de connaissances. Ces dernières sont appelées des *attitudes propositionnelles*, dans la mesure où elles s'expriment dans et relativement à des propositions : par exemple dans la phrase « je suis impressionnée par tes propos », « impressionnée » rend compte d'une émotion particulière telle que l'admiration exprimée vis-à-vis d'un énoncé.

Dans le domaine de l'intonation, Wichman propose de réserver le terme d'*intonation expressive* aux énoncés porteurs d'indices prosodiques de l'émotion et aux énoncés exprimant une attitude propositionnelle, issue de croyances, connaissances ou opinions, et le terme d'*intonation attitudinale* aux énoncés qui dénotent un comportement de la part du locuteur, intentionnel et/ou inféré par son auditeur dans une situation donnée.

Par ailleurs, certaines méthodes et concepts-clé traversent tout le domaine. Pour analyser plus objectivement les émotions perçues, les psychologues travaillent sur la définition de caractéristiques pourvues d'une graduation, c'est-à-dire sur des dimensions. Ainsi avec « *Feeltrace* », Cowie et son équipe [COW 00, 00b] proposent une représentation sous la forme d'un espace circulaire où les dimensions de l'*activation* et de l'*évaluation* sont représentées sur 2 axes orthogonaux, dimensions telles que les perçoivent les observateurs. Ce dispositif est complété par une étude de l'évolution des paramètres dans le temps.

L'*activation* mesure la dynamique de l'état émotionnel, c'est-à-dire le degré d'*excitation*. L'*évaluation* mesure la *valence*, c'est-à-dire le sentiment positif ou négatif associé à l'état émotionnel vis-à-vis des personnes, faits ou choses. Au centre de cet espace circulaire se situe le point de neutralité émotionnelle, à la périphérie, le degré d'émotion la plus intense.

Le degré d'activation permet d'évaluer la force en fonction de laquelle une personne est disposée à entreprendre telle action plutôt que telle autre. Cette dimension est à associer en particulier aux recherches de Frijda pour qui, comme on l'a vu plus haut, les états émotionnels sont vus comme des tendances à l'action. Elle trouve bien entendu aussi des corrélats dans le domaine de la parole. Des études ont été menées en ce domaine pour déterminer si les paramètres vocaux permettent de graduer les états d'excitation des émotions.

Dans le domaine de la parole, une activation positive (joie, peur, colère, et à moindre titre, surprise, excitation, perplexité) contribue à l'augmentation du F0 moyen et de son amplitude, et à celle de la tension de la voix, alors qu'une activation négative (tristesse, chagrin, ennui) est inversement associée à leur diminution. Pour une présentation plus détaillée, voir ci-dessous paragraphe 15.4.2.2.

15.4. La parole affective

Une fois posé le cadre conceptuel global des recherches qui traitent de l'affectivité, nous pouvons maintenant aborder le développement des analyses dans le domaine de la parole, qui bénéficient bien entendu des apports des sciences connexes.

15.4.1. *À la recherche des caractéristiques prosodiques objectivables des émotions*

Contrairement aux sciences de la nature, en psychologie et dans les sciences humaines en général, il est nécessaire de construire son objet de recherche. Pour assurer un maximum d'objectivité aux analyses, il convient alors prioritairement de chercher les caractéristiques qui précisément sont objectivables.

Léon [LEO 71] propose ainsi de recenser dans le domaine de la « phonostylistique » (dénomination qui n'a pas été retenue par la suite par la communauté scientifique) au niveau de la substance et de la *forme* des faisceaux de traits définitoires des émotions, qu'il appelle des « *phonostylèmes* », et qui fonctionnent à l'image des « phonèmes » en phonologie. Parmi ces phonostylèmes, figurent la fréquence fondamentale moyenne (ou F0 moyen) du locuteur dans l'énoncé, les écarts d'amplitude (écart appelé souvent par la suite à la suite des anglais, le « *range* ») par rapport à ce F0 moyen, les types de contours de F0, la durée, et l'intensité. Aucun indice à lui seul n'est suffisant pour identifier une émotion, un contour par exemple pouvant assumer plusieurs fonctions linguistiques.

Plus récemment, Roach [ROA 00] souligne le problème fondamental de l'objectivité à propos de l'étiquetage des phénomènes : tel trait par exemple a-t-il des corrélats physiques mesurables objectivement, ou n'est-il qu'une étiquette subjective issue de l'intuition du chercheur ? Pour pallier cette difficulté, s'inspirant de Crystal et Quirk [CRY 64] et de Laver [LAV 68], il propose un étiquetage prosodique et paralinguistique assez complet (36 étiquettes). Le codage prosodique intègre quatre dimensions de F0 (hauteur et étendue), quatre également pour l'intensité (niveau et évolution), 6 pour le tempo (rapidité, évolution, qualité). Depuis l'origine des travaux, beaucoup de systèmes ont été mis au point pour rendre compte des particularités, des événements vocaux et divers bruits qui accompagnent l'énonciation. Ainsi le codage paralinguistique intègre neuf caractéristiques pour étiqueter la qualité de la voix, cinq pour les réflexes qui affectent la fluidité de la parole (raclement de gorge, reniflement...) et qui peuvent être les indices d'émotions, et trois pour des expressions d'états émotionnels (rire, pleurs, voix tremblante).

Une autre manière de consigner les faits avec plus d'objectivité est aussi de recourir à une standardisation des notations prosodiques. Pour ce point important mais qui sort de l'objectif de ce chapitre, voir Durand, ce volume, chapitre 2.

15.4.2. *La prosodie des états émotionnels*

Nous envisagerons l'analyse sous deux aspects, celui du cerveau, et celui des analyses (modèles et résultats) de la parole affective.

15.4.2.1. *Prosodie émotionnelle et cerveau*

On a longtemps privilégié dans les études le rôle de l'hémisphère droit du cerveau dans le traitement des émotions et de ses marqueurs prosodiques. Avec le développement des recherches en imagerie cérébrale, en neuropathologie, des progrès considérables ont été accomplis, permettant d'affiner sensiblement les résultats. Comme le souligne Pell [PEL 02], la grande difficulté est de définir et d'isoler, là où c'est possible, les composants physiologiques, comportementaux et cognitifs de la prosodie émotionnelle. Au niveau du contrôle cognitif, il reste aussi beaucoup à faire pour déterminer la structure componentielle de la prosodie affective, et donc beaucoup à connaître sur le degré d'indépendance des traitements ou processus cognitifs vis-à-vis des différentes aires cérébrales, aires que l'on pressent organisées en réseaux sur le plan fonctionnel.

À l'issue de nombreux travaux réalisés depuis 25 ans sur des populations de patients souffrant de lésions cérébrales par accident ou maladie (notamment de Parkinson), confortés par les nouvelles technologies d'exploration cérébrale, selon Pell [PEL 02], le rôle de différents composants du cerveau est désormais mieux connu. Ainsi l'*hémisphère gauche* est davantage spécialisé dans le *traitement linguistique* en général et le *traitement linguistique de la prosodie* en particulier lorsqu'il s'agit d'en dériver une signification linguistique (et non pas de traiter les traits émotionnels des stimuli). Les deux hémisphères, dans leurs zones corticales et sous-corticales, sont impliqués dans le processus de régulation cognitive de la prosodie, incluant la prosodie émotionnelle. Il est également établi que les *régions préfrontales inférieures des deux hémisphères* se spécialisent aussi dans le *traitement spécifique de la prosodie affective*, notamment au niveau de l'intégration et l'activation des connaissances liées à l'affectivité.

Le rôle de l'*hémisphère droit*, dont la fonction des traitements linguistiques est beaucoup plus limitée, serait ainsi d'héberger un « *lexique* » *affectif* non verbal (prosodie, expressions faciales, gestes) qui fournirait un savoir représentationnel sur la manière dont l'émotion est communiquée, y compris par les corrélats prosodiques. Les propriétés saillantes de tels signaux initialement activés et traités par un réseau réparti dans les deux hémisphères, seraient ainsi comparées au savoir représentationnel de l'hémisphère droit, ce qui permettrait ensuite d'instancier une inférence sur la signification sociale de tels signaux.

Selon Pell [PEL 02], il semble bien attesté également que l'*hémisphère droit* aurait plus de capacité à traiter les *indices de F0* que les aspects temporels, sans doute parce qu'ils sont, comme les études le soulignent bien, le *véhicule des intentions liées aux émotions*. Il semble tout à fait approprié de considérer ainsi que l'hémisphère droit aurait un rôle privilégié pour le traitement de la prosodie émotionnelle dans ses aspects de comportement social et pragmatique.

Par ailleurs, Lieberman [LIE 00], faisant le point sur les travaux en cours, a montré l'importance des ganglions basaux, et tout particulièrement le caudate, pour l'apprentissage implicite et le décodage non conscient et intuitif de certains événements, leur rôle prépondérant dans le traitement des associations temporelles (plus que conceptuelles) des événements, et l'assignation de valeurs prédictives à la signification de comportements. Selon Pell [PEL 02], il y a tout lieu de penser que *la prosodie*, qui nécessite un traitement temporel de ces indices, et qui est vraisemblablement un système comportemental, bénéficie des *propriétés computationnelles des ganglions basaux*.

Un autre type d'études (menées par Schirmer et Kotz, [SCH 02] utilisant les techniques de l'électroencéphalogramme (électrodes disposées sur les quatre régions antérieures et postérieures, gauches et droites) a montré que les traitements des indices prosodiques émotionnels varient en fonction du sexe des auditeurs. Cette étude consistait à faire entendre aux sujets des mots de valence différente (sens positif, neutre, négatif) associés à une prosodie congruente ou pas. Dans une tâche, les sujets jugeaient de la prosodie en ignorant le sens des mots, dans une autre, ils jugeaient du sens émotionnel des mots en ignorant la prosodie.

Les résultats ont montré tout d'abord que chez les deux sexes, les hémisphères antérieurs traitent la prosodie émotionnelle de manière indépendante. Ensuite qu'il existait pour le groupe des femmes seulement, une interaction entre la prosodie émotionnelle d'un mot et le sens de ce dernier. Le fait que la prosodie a un impact plus fort sur le jugement de la valence du mot que l'inverse, suggère que chez les femmes, c'est la prosodie plutôt que le sens des mots qui gouverne le traitement des émotions dans la parole. En contraste avec ce qui se passe chez les femmes, chez les hommes, des traitements indépendants entre la prosodie et le sens des mots sont mis en œuvre. Dans d'autres expérimentations portant sur des tâches comportementales, à égalité des performances, les femmes produisaient des jugements émotionnels plus rapides. Ces résultats tendent à montrer que les femmes intègrent l'information émotionnelle de la prosodie et des mots de manière plus précoce et plus automatique que les hommes.

15.4.2.2. *Prosodie et états émotionnels*

Les études sur les corrélats prosodiques des états émotionnels sont très anciennes. Dans cette voie les premières recherches sont sans doute celles de Fairbanks et Pronovost [FAI 39], et de Fairbanks et Hoaglin [FAI 41].

	Colère	Joie	Tristesse	Peur	Dégoût
Débit	légèrement plus rapide	plus rapide ou plus lent	légèrement plus lent	beaucoup plus rapide	vraiment beaucoup plus lent
Moyenne de F0	vraiment beaucoup plus haut	beaucoup plus haut	légèrement plus bas	vraiment beaucoup plus haut	vraiment beaucoup plus bas
Amplitude de F0	beaucoup plus large	beaucoup plus large	légèrement plus restreinte	beaucoup plus large	légèrement plus restreinte
Intensité	plus forte	plus forte	plus faible	normale	plus faible
Qualité de la voix	soufflée, voix de poitrine	soufflée, explosive	résonnante	voisement irrégulier	grognement, voix de poitrine
Variations de F0	abruptes, sur les syllabes accentuées	régulières, inflexions montantes	inflexions descendantes	normales	grandes, inflexions terminales descendantes
Articulation	tendue	normale	relâchée	précise	normale

Tableau 15.1. *Résumé des effets des émotions sur la parole traduit de l'anglais d'après Murray & Arnott [MUR 93]. Les effets décrits sont ceux les plus communément associés avec les émotions citées ci-dessus, et sont relatifs à la parole neutre. F0 est l'abréviation de « fréquence fondamentale ».*

Par la suite, de nombreux travaux ont vu le jour, parmi lesquels, entre autres, ceux de Williams and Stevens [WIL 72], Bolinger [BOL 45, 46, 65, 72], Crystal [CRY 75]. Spécifiquement pour le français, on peut citer parmi les premiers travaux, comme on l'a vu ci-dessus, Léon [LEO 70, 71, 76], Faure [FAU 70, 73], Fónagy *et al.* [FON 73], Fónagy et Sap [FON 97], Fónagy [FON 82a, 82b]. Ces études se caractérisent par la recherche de profils acoustiques distinctifs (valeurs de F0, formes et patrons mélodiques, durées et débits).

Depuis cette époque, beaucoup d'études se sont évertuées à dresser les profils prosodiques des émotions pour tenter de les discriminer. Les caractéristiques prosodiques relevées par les auteurs sont concordantes d'un auteur à un autre, et globalement d'une langue à une autre, au moins en ce qui concerne les émotions primaires. Il est donc possible d'en proposer un résumé. Le tableau 15.1 ci-dessus présente une description des principales caractéristiques prosodiques de cinq émotions primaires pour l'anglais, et qui sont applicables à beaucoup d'autres langues. Ce tableau a été dressé par Murray and Arnott [MUR 93]. F0 est l'abréviation de « fréquence fondamentale ».

En guise d'illustration, nous proposons ci-dessous plusieurs extraits de parole spontanée, montrant le premier, l'émotion primaire de la joie, le second, deux attitudes. Le premier ci-dessous (voir figure 15.1) : « *[cela m'a marquée petite*

(rires)] *j'me rappelle le grenier (rires)...* », est énoncé par une locutrice relatant avec une émotion de *joie* un souvenir d'enfance, en l'occurrence l'expérience d'un grenier à gouttières. On peut ainsi observer sur cet extrait :

- la portion d'énoncé sélectionnée dans la fenêtre du logiciel Praat [BOE 04] « *j'me rappelle le grenier* » qui est prononcée sous l'emprise de la joie, portion ponctuée par les rires de la locutrice,

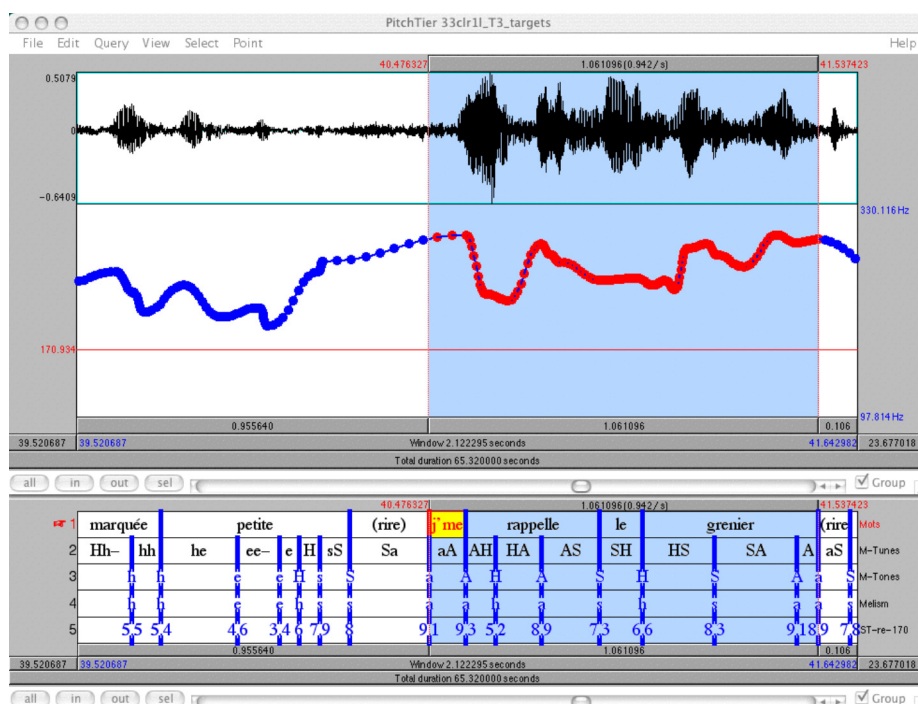


Figure 15.1. Extrait d'un corpus de parole spontané sous Praat [BOE 01] rappelant un souvenir d'enfance exprimé avec joie: « [cela m'a marquée petite (rires)] je me rappelle le grenier (rires) ». De haut en bas, le signal de parole, la courbe de F0 représentée à l'aide du logiciel MOMEL [HIR 93], avec la moyenne de la F0 de la locutrice (170 Hz) représentée sous la forme d'une ligne rouge horizontale, le texte de l'énoncé, le codage des séquences tonales relatives aux unités linguistiques, les cibles tonales [CAE 04] codant les valeurs de F0 exprimées en demi-tons sur la ligne du dessous.

- dans cet extrait, conformément aux caractéristiques décrites ci-dessus dans le tableau 15.1, l'intensité de cette séquence est plus forte que la portion précédente (comparer à ce sujet, les hauteurs respectives du signal de parole de la séquence précédente « *cela m'a marquée petite* » et de la séquence émotionnelle) ;

- la hauteur de F0 (voir la courbe de la fréquence fondamentale) qui monte vers les régions aiguës du registre de la locutrice, bien au-delà de sa valeur de F0 moyenne (170 Hz). En effet sur les neuf niveaux retenus (par ordre décroissant, *Aigu*, *Supra*, *Haut*, *Elevé*, *Moyen*, *Centré*, *Bas*, *Infra*, *Grave*) pour étudier avec précision les motifs mélodiques de la voie affective [CAE 04], on note que les valeurs de F0 appartiennent aux niveaux les plus élevés soit *A*, *S* et *H* (en majuscules ou minuscules).

Le deuxième exemple d'une autre locutrice propose une succession de deux attitudes. On peut ainsi voir sur la figure 2 ci-dessous :

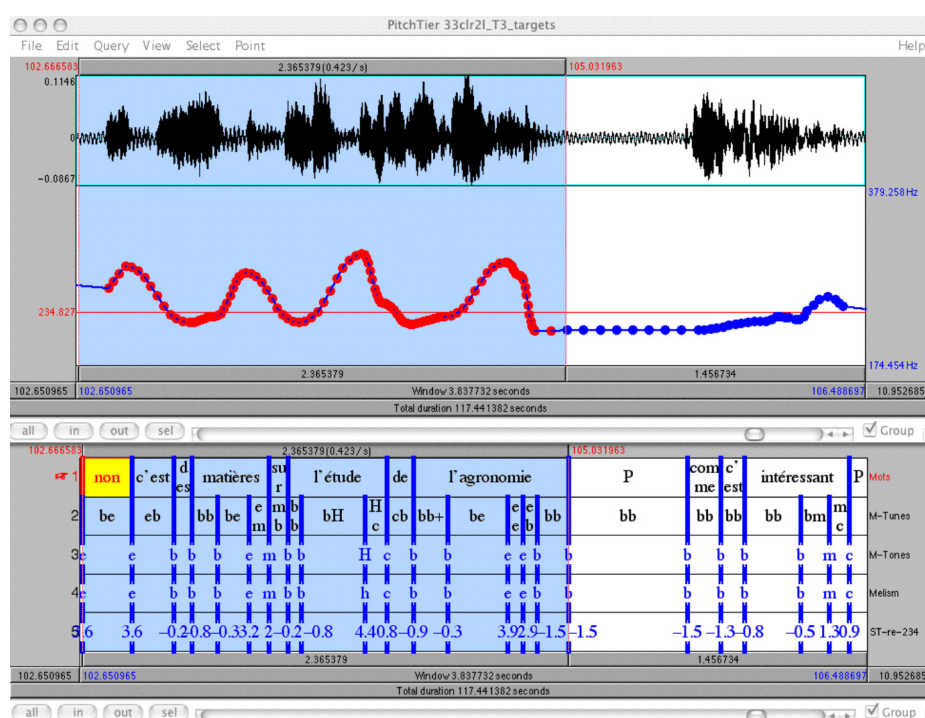


Figure 15.2. Extrait d'un corpus de parole spontané sous Praat [BOE 01, HIR 93] prononcé sous l'émotion de l'agacement contenu puis de l'ironie : « Non, c'est des matières sur l'étude de l'agronomie (pause) comme c'est intéressant... » De haut en bas, le signal de parole, la courbe de F0, le texte de l'énoncé, le codage des séquences tonales relatives aux unités linguistiques, les cibles tonales codant les valeurs de F0 exprimées en demi-tons sur la ligne du dessous.

- la portion d'énoncé sélectionnée dans la fenêtre du logiciel Praat « *non, c'est des matières sur l'étude de l'agronomie (pause) comme c'est intéressant...* ». Dans

cet extrait, l'énoncé est exprimé par la jeune fille à l'encontre de sa famille (famille qui l'a inscrite à la rentrée suivante dans un lycée agricole, alors qu'elle souhaite faire des études commerciales). La partie avant la pause est prononcée avec l'attitude d'un « *agacement contrôlé* », la deuxième partie sous le signe de l'*ironie*,

- on constate que l'ensemble de cet énoncé ne se situe plus dans le registre élevé, mais de part et d'autre de la valeur de F0 moyenne (pour cette locutrice 234 Hz),

- la partie avant la pause, exprimant un agacement à l'encontre de sa mère qui tente de lui présenter sa future scolarité sous un jour agréable, se compose d'une succession de pentes alternées assez rapides qui martèlent chacun des mots lexicaux. Les niveaux sont compris entre *b* et *h*, la grande majorité entre *b* et *c*,

- la partie après la pause s'oppose par sa forme à la précédente. Elle est dite avec *ironie* dans un registre proche mais avec une amplitude de F0 plus restreinte : entre *b* et *m*, avec un plateau initial niveau *b*, ce qui lui confère un contour assez plat et monotone, contrastant vivement avec le contenu de l'énoncé.

15.4.2.2.1. L'approche linguistique / pragmatique de la prosodie émotionnelle

À la fois linguiste des langues et de la parole, auteur de près de 350 écrits, Bolinger (1907-1992) s'est intéressé dès 1945 aux systèmes prosodiques de différentes langues. Dès 1972 [BOL 72], il est un des premiers à lier de manière essentielle l'intonation aux aspects sémantiques et au contexte du discours, montrant que la prééminence prosodique signalant l'information (nouvelle ou importante) est liée à l'accentuation. Pour lui les fonctions grammaticales de la prosodie sont indirectes, l'intonation ne définissant pas les catégories grammaticales, et réciproquement. Malgré la vague conceptuelle de l'explication syntaxique qui est alors dominante, et avant le nouvel essor au niveau international des considérations sémantiques puis pragmatiques pour la prosodie, une perspective proche à cette époque, est tenue en France par Faure [FAU 70, 73], puis par Caelen-Haumont [CAE 78/81].

Concernant les émotions, Bolinger [BOL 86] montre que la prosodie renforce les intentions sous-jacentes du locuteur, à savoir les croyances et le désir d'attention. Il distingue pour l'anglais, les accents de pouvoir qui donnent de la force à l'assertion (et qui sont positionnés à la frontière droite du mot), par le positionnement d'un point culminant, et les accents d'intérêt. Une manière de réduire l'accent de pouvoir en anglais est de faire glisser l'accent vers la frontière de gauche du mot. Bolinger souligne aussi, en dehors de la question de l'accent, que la mélodie fournit pour l'expression des sentiments, les indices de l'excitation (mélodie haute) vs. du relâchement (mélodie basse), avec la fin de la tension.

Ces conceptions décrivent en fait plus la prosodie subjective que la prosodie des émotions de base, ce que Caelen-Haumont et Bel décrivent pour le français comme

« l'émotion ordinaire » [CAE 00], et que l'on pourrait ici dénommer plus à propos comme l'émotion-racine, celle qui est à la source de l'être subjectif, de ses croyances, de ses valeurs, de ses ressentis, de ses expressions, émotion-racine par nature indifférenciée, et qui active en second lieu les émotions bien différenciées, les émotions de base et les dérivées, de même que les attitudes, avec leur cohorte d'indices prosodiques et vocaux.

Parmi les tout premiers essais d'interprétation des indices prosodiques liés à l'expression des émotions et attitudes, Léon [LEO 70], déjà mentionné précédemment, associe pour le français perceptivement un registre élevé à « la gaieté, timidité, légèreté », un registre bas à « la tristesse, assurance, gravité », et l'augmentation de l'amplitude tonale à la joie, la colère, ainsi que sa réduction à la tristesse. Le contour inversement n'est pas considéré à lui tout seul comme discriminant, tandis que l'intensité renforce le sentiment exprimé.

Il semble qu'une caractéristique des nombreux travaux dans le domaine est d'avoir associé les émotions à la conjonction paramètres de F0 / types ou modalités de phrase. Ainsi pour O'Connor et Arnold [O'C 73], les « *wh-questions* » avec « chute haute » sont qualifiées de « vives, affairées, attentionnées, non inamicales, vivantes, ou intéressées ». Pour Halliday [HAL 94], les « *wh-questions* » avec un ton montant signalent une continuité, tandis que les « questions oui-non » avec un ton descendant sont péremptoires. Une déclarative finissant par une montée dénote « un défi, de l'agressivité, une défensive ou de l'indignation. »

Comme nous le verrons dans le paragraphe suivant, Scherer *et al.* [SCH 84b], dans le modèle « configuration », reprend cette idée de lier l'émotion à la relation contour / type de phrase. Ainsi dans l'étude des questions, le fait d'inverser les patrons standards de l'anglais (chute finale pour les *wh-questions*, montée finale pour les questions oui-non) produit une perception plus négative. Comme le souligne Wichman [WIC 02], ces résultats montrent que les choix marqués génèrent des attitudes, en particulier négatives. On peut aussi avancer l'idée qu'une autre source de sens affectif peut être générée dans les associations conventionnelles contour / acte de parole qui sont rompues (intentionnellement ou pas). Dans ces conditions, il serait bien d'établir le patron des choix non marqués pour tous les types de phrases pour comparaison. Mais la tâche n'est pas simple car elle possède des déterminants qui ne sont pas uniquement linguistiques.

Pour le français, Fónagy et Bérard [FON 73], Fónagy *et al.* [FON 82c], ont étudié en détail les différents clichés mélodiques du français parisien, clichés qui sont dépendants du style de parole. Fónagy [FON 82a] [FON 82b], spécialise ses études dans l'analyse des styles de parole (par exemple lecture de contes de fée, informations journalistiques...), des attitudes, et des émotions, détaillant avec précision contours et cibles mélodiques.

Dans une étude pragmatique de 88 contours prosodiques de phrases anglaises se terminant par une requête (« *please* »), se réalisant la plupart du temps par une interrogation avec une finale soit montante, soit descendante, Wichman [WIC 02] montre que les requêtes avec finale haute sont utilisées dans une relation privée (amis, collègues), alors que celles avec finale basse le sont en situation publique (statut social inégal). Ces résultats sont en congruence avec le Code de Fréquence proposé par Ohala [OHA 83], attribuant pour l'anglais un effet de soumission au registre élevé, et un effet de pouvoir au registre grave. De fait Wichman conclut qu'un registre élevé dans une situation informelle laisse à l'interlocuteur la possibilité d'achever l'échange, tandis que dans une situation formelle et plus impersonnelle, un contour à finale descendante est la propriété de celui qui possède l'autorité, qui la manifeste sous la forme d'une invitation (qu'il n'est pas bon de refuser), ou d'une directive.

Concernant le français, dans la continuité du Code de Fréquence défini comme on l'a vu ci-dessus par Ohala, Piot [PIO 01] se propose de tester sur les questions à contour montant l'hypothèse que la montée de F0 serait « plus grande lorsque le locuteur est plus désespéré, c'est-à-dire dans ce cas plus ignorant ». Utilisant les ressources de la synthèse de la parole pour manipuler plus facilement les items du test et leurs combinaisons acoustiques croisées (F0, intensité, débit), deux expériences perceptives ont été menées, la première sur la question totale, pour l'évaluation 1° de « l'ignorance exprimée », 2° de « l'envie de savoir exprimée », la deuxième sur l'assertion à finale haute pour évaluer 1° « l'ignorance d'autrui exprimée », 2° « l'envie d'informer exprimée ».

De ces deux expérimentations, il ressort tout d'abord que la hauteur du pic tonal est déterminante pour exprimer l'ignorance du locuteur *vs.* de l'allocutaire, et l'envie de savoir *vs.* d'informer. Viennent ensuite l'importance de l'amplitude et de la finale des contours pour la perception des attitudes, ainsi que l'existence d'une gradience pour la relation hauteur tonale / nouveauté. Un plus grand débit renforce la perception d'une plus grande envie de savoir, mais pas l'envie d'informer, et exerce une influence négative sur l'expression de l'ignorance de l'allocutaire. Enfin l'intensité finale a un effet sur les assertions mais pas sur les questions.

Dans la continuité des travaux de Scherer sur la covariance et la configuration (voir plus loin), et leurs implications, Ladd *et al.* [LAD 85] entreprennent trois expérimentations du même ordre. Les sujets devaient juger de l'émotion véhiculée par les énoncés où l'amplitude de F0, le type de contour, et la qualité de la voix étaient systématiquement modifiés. Une tâche de classement était proposée pour les états liés à une excitation, et une autre pour les attitudes cognitivement liées. L'objectif de la dernière expérimentation était de tester si les variations dans l'amplitude de F0 avaient des effets continus ou catégoriels sur la perception des émotions.

Les résultats ont montré d'une part que l'amplitude de F0 et la qualité de la voix avaient une forte incidence à la fois sur l'inférence que fait le sujet sur l'état d'excitation du locuteur, et sur l'inférence des attitudes cognitivement liées. Il n'a pas pu être montré que le type de contour affectait massivement le classement des attitudes, mais au contraire plutôt le classement des émotions. Comme on s'y attendait, les variations de l'amplitude de F0 ont causé des effets plus continus que catégoriels dans les jugements relatifs à l'émotion du locuteur. Une conclusion importante a été tirée, à savoir que les trois indices prosodiques, amplitude de F0, type de contour et qualité de la voix fonctionnent indépendamment les uns des autres.

15.4.2.2.2. L'approche par les modèles

Les indices prosodiques des dimensions de la prosodie : valence, activation, pouvoir

Ayant constitué, dans le cadre de l'allemand, une base de données de dialogues spontanés recueillis dans une situation de construction de Lego, Kehrein [KEH 02] a sélectionné sur la base de consensus perceptuels, les meilleurs candidats parmi les énoncés porteurs d'émotion. Après transcription orthographique, sous Praat, les paramètres du F0 moyen, début / fin de F0, amplitude de F0, F0 maximum / minimum, nombre de syllabes par seconde ont été extraits. L'objectif était de tester la relation entre ces divers paramètres et les dimensions de la prosodie : valence, activation, pouvoir, auxquels a été ajouté le caractère inattendu.

Il apparaît très clairement que l'information inattendue est liée à l'élévation du maximum de F0 : plus F0 est haut et plus la force du caractère inattendu est importante. Ceci confirme les résultats obtenus par Caelen-Haumont [CAE 91] pour le français avec un modèle des informations inattendues. Dans l'étude de Kehrein, l'émotion correspondante est celle d'un locuteur « alarmé, surpris, horrifié ».

Concernant la dimension de l'activation, c'est le tempo qui est le facteur dominant : plus le rythme est rapide, plus le locuteur est perçu comme excité, et inversement. C'est le cas lorsque les locuteurs sont perçus comme « excités, agités, incertains, anxieux ou en colère ». Pour le rythme lent, les locuteurs sont perçus comme « calmes, relaxés, satisfaits ou encore démotivés ».

Pour la dimension du pouvoir (ou dominance), la force est liée à une augmentation de la prééminence, ce qui correspond à une augmentation de l'amplitude de F0 et de l'intensité, et inversement pour la qualité opposée. Les locuteurs sont alors perçus comme « énergiques, sûrs, en colère et ennuyés », mais aussi « ravis ». Dans le cas contraire, ils sont perçus comme « peu sûrs », « perplexes, s'excusant, résignés, frustrés, déçus ».

Concernant la valence positive, les paramètres sont un étirement temporel de la syllabe, un contour avec un F0 maximum, et un contour alterné faible / fort pour l'intensité. Les locuteurs sont alors qualifiés de « ravis », bien que d'autres études mentionnent aussi pour une telle configuration par exemple le plaisir gastronomique.

Les codes biologiques et autres codes

Ohala [OHA 83] a proposé d'instancier un *Code biologique* de la fréquence fondamentale (*Frequency Code*) fondé sur le fait que les larynx de petite taille ont tendance à produire des notes plus élevées que ceux de plus grande taille. Il en découle que les fréquences élevées sonnent comme vulnérables et soumises, les fréquences basses comme protectrices et dominantes. Des expérimentations ont montré que pour le code fréquentiel, le registre fréquentiel (F0 moyen) et l'étendue de F0 étaient pertinents. Quand l'étendue de F0 augmente, le degré du caractère amical du message augmente, mais le degré de confiance diminue. Ceci varie cependant en fonction de la langue, et cette perception, à étendue de F0 égale, est reçue comme plus amicale et générant plus de confiance dans les stimuli de langue britannique que leurs correspondants hollandais. L'explication en est que les différences linguistiques des étendues de F0 induisent des effets non uniformes sur la perception des significations signalées par de telles variations.

Chen *et al.* [CHE 02] proposent à leur tour deux autres codes, le *Code de Production* et le *Code de l'Effort*. Ce dernier est fondé sur le fait qu'un effort articulatoire plus intense a tendance à créer des réalisations phonétiques plus élaborées et plus explicites. Les codes biologiques se prêtent à deux types d'interprétations (information et affectivité), et dans le cas du Code de l'Effort, l'emphase répond à une information sur l'importance subjective que le locuteur accorde à son message, tandis qu'une émotion comme la surprise répond à celle de l'affectivité. Dans les deux cas, le locuteur dépense de l'énergie pour signaler ces fonctions. Trois indices de F0 manifestent cet effort et se révèlent pertinents pour l'interprétation des significations : la hauteur des pics de F0, le retard dans l'alignement des pics, et le registre de F0.

Concernant l'emphase, les résultats interlangues montrent une corrélation positive avec la hauteur du pic de F0, mais à hauteurs identiques, les stimuli hollandais sont perçus comme plus emphatiques que les stimuli de l'anglais. Ceci trouve sans doute son explication dans le fait que l'étendue de F0 en hollandais est plus faible qu'en anglais. Concernant le registre du *pitch*, une corrélation négative est cette fois trouvée entre les deux langues, le degré d'emphase croissant avec un registre du *pitch* plus aigu pour le hollandais, avec un registre du *pitch* plus grave pour l'anglais. Une explication possible est que la hauteur du *pitch* pourrait être utilisée pour un autre type de signification.

Concernant la surprise, dans les deux langues, on trouve une corrélation positive d'une part entre la hauteur du pic et le degré de surprise, d'autre part entre ce dernier et la production tardive du pic de F0. Dans les deux langues en effet, retarder l'occurrence d'un pic de F0 est perçu avec la même fonction qu'un pic plus haut (non retardé). Dans les deux cas, à hauteurs de pic identiques et à délais identiques de production du pic, les stimuli hollandais sont perçus comme générant plus de surprise. Ceci tient au fait déjà constaté de l'usage d'une étendue de F0 plus restreinte dans le cas du hollandais.

Les deux significations sont dérivées du Code de l'Effort qui stipule que les excursions les plus larges et les hauteurs les plus aiguës correspondent aux significations associées à la dépense de l'effort, telles que la signifiante du message (emphase, insistance) ou l'état d'agitation du locuteur (surprise). Un usage différent des indices de F0 / signification apparaît dans les deux langues, les anglais, relativement au Code de Fréquence, réservant le registre haut comme un signal de soumission et de disposition amicale (*vs.* protection et autorité), les hollandais, relativement au Code de l'Effort, l'utilisant pour l'insistance et l'emphase dans la mesure où la hauteur de F0 reste un composant de l'étendue du *pitch*.

Les modèles de Scherer

Dès le début des années 1970, Scherer monte un groupe de recherches spécialisé dans le domaine des paramètres vocaux comme indicateurs potentiels des états du locuteur tels que motivation, émotion et stress. Menés en collaboration avec un groupe dirigé par Ekman, les résultats d'expérimentations [EKM 76] montrent que sous le stress, le niveau moyen de F0 augmente. Toutefois, les travaux ultérieurs de Scherer ont montré que les organismes répondaient de manière non uniforme aux différents types de stress.

Il se trouvait que les modèles transactionnels défendaient cette thèse en argumentant que le type d'évaluation et le potentiel de confrontation de l'individu co-déterminent le type de réponse au stress. Le cadre du travail de Scherer était ainsi dessiné : focaliser les recherches « sur les mécanismes d'évaluation de l'individu en essayant de comprendre les réponses physiologiques et motrices aux événements stressants » [SCH 86, 87].

Un temps de latence existant manifestement entre l'évaluation d'un événement et l'exécution d'une réponse comportementale, ce temps pouvait être l'indice d'une activation de sous-systèmes de l'organisme. Scherer [SCH 84a] propose alors un modèle liant les composants de l'émotion aux différents sous-systèmes exerçant chacun une fonction spécifique, suggérant ainsi que les états émotionnels devaient être étudiés dans les termes d'un « *component process model* ». L'originalité des travaux se situent plus particulièrement dans la définition d'une série de contrôle

dans l'évaluation du stimulus. Cette série de contrôles concerne, dans l'ordre, la dimension de la nouveauté, de l'agrément (situation agréable ou pas), de la proximité dans la satisfaction des buts ou des besoins, du potentiel de maîtrise de soi, de la compatibilité aux normes personnelles et sociales. Scherer *et al.* [SCH 03] fournissent une description différenciée des états émotionnels en fonction de ces divers contrôles, sous la forme d'une matrice de traits distinctifs, permettant de prédire les indices vocaux correspondants. Du fait que ces contrôles varient grandement d'un individu à un autre, même pour un stimulus identique, le modèle prédit que les réponses des individus au stress seront très différentes, qu'il s'agisse des sentiments, de la réponse physiologique ou de l'expression motrice. Scherer [SCH 86] propose différents tableaux prédisant quels types de réponses sont attendus de l'organisme en fonction de ces domaines.

Par ailleurs, Scherer *et al.* [SCH 84b] proposent de distinguer pour l'intonation deux approches, l'approche *covariance*, et l'approche *configuration*. La première stipule que l'information sur l'émotion et l'information sur le contenu linguistique au sens strict fonctionnent de manière indépendante, ce qui entraîne la conséquence que les traitements linguistiques et paralinguistiques puissent s'effectuer en parallèle. Les indices de F0 associés à cette approche sont le F0 global et l'amplitude de F0. L'approche configuration pose que le type de contour de F0 est un élément linguistique. De ce fait on distingue entre les variations linguistiques de F0 et les variations paralinguistiques. Cela sous-entend que l'information liée au contour et à l'émotion n'est valide que si elle est traitée en interaction avec les traits linguistiques de l'énoncé. L'indice de F0 correspondant est le type de contour. Au terme d'expérimentations perceptuelles, il apparaît que les deux approches sont pertinentes, la covariance semblant plus adéquate pour décrire la parole affectée par les facteurs biologiques, la configuration, pour la parole affectée par les conventions linguistiques et socioculturelles.

Scherer [SCH 89] par ailleurs propose de distinguer deux types d'effets au niveau des expressions de la parole émotionnelle, les *effets push* produits par des « *déterminants internes au sujet* » causant des modifications physiologiques en réponse à une activation émotionnelle, et les *effets pull* produits par des « *déterminants externes au sujet* » en relation avec le contexte environnemental social où se trouve le locuteur. Toutefois, selon Scherer, même s'il est parfois possible d'évaluer la charge respective de ces effets pour un locuteur dans une situation donnée, les effets se combinent dans les indices acoustiques de la voix sans qu'on puisse les dissocier.

Les recherches de l'équipe Scherer semblent maintenant s'orienter d'une part vers l'analyse des processus sous-jacents à l'encodage et décodage des émotions, et d'autre part grâce à l'analyse des cartes topographiques de potentiels évoqués, vers

le traitement du discours temporel dans les processus cognitifs des composantes émotionnelles.

Caelen-Haumont et Bel [CAE 00, 02, 04] envisagent dans la prosodie deux strates, l'une qui serait la partie linguistique et conventionnelle et qui correspondrait à l'intonation (relation avec le structure linguistique), et l'autre qui serait la partie exprimant l'affectivité (« l'émotion-racine » ou émotions discrètes) et qui serait exprimée par le mélisme, d'expression locale, souvent mono-lexicale, caractérisée d'une part par une amplitude de F0 très importante, ce qui implique des valeurs maximales de F0, et d'autre part souvent liée à une rupture mélodique ou prosodique de la trame linguistique, soit en dehors des frontières de syntagmes [CAE 78/81]. Ce double traitement prosodique (intonation / mélisme) pourrait ainsi être la marque prosodique de processus cognitifs que les études établissent.

Par exemple, au niveau des aires cérébrales tout d'abord, selon Pell [PEL 02] il semble attesté d'une part que l'hémisphère gauche soit plutôt spécialisé dans le traitement de la linguistique et de la prosodie linguistique (niveau représentationnel), alors que l'hémisphère droit aurait plus d'aptitude à traiter la prosodie émotionnelle, avec les indices de F0, véhicules attestés des intentions liées aux émotions.

Par ailleurs, les modèles configuration et covariance de Scherer *et al.* [SCH 84b], qui interviennent dans le traitement de la prosodie, valident totalement ces conceptions, le modèle de covariance, rendant compte du traitement des effets à la fois émotionnels et linguistiques, le modèle de configuration, le traitement purement linguistique. Cette dichotomie des fonctions se retrouve également dans le modèle de Scherer [SCH 89], « *pull* » et « *push* », deux processus qui rendent compte des expressions de la parole émotionnelle sous contrôle cognitif.

Ainsi la distinction intonation / mélisme, établie à partir de l'observation des données, pourrait servir d'ancrage factuel aux investigations biologiques et aux modèles cognitifs.

Le modèle unifié de Zei-Pollermann

Une autre validation concernant cette fois « l'émotion-racine » est apportée également par Zei-Pollermann [ZEI 02]. En effet s'appuyant d'une part sur les conceptions de Piaget [PIA 70] et de Prieto [PRI 75], et d'autre part sur les dimensions reconnues généralement à la prosodie (valence, pouvoir, activation), l'auteur propose un modèle unifié cognition / émotion, argumentant que les processus cognitifs incluent nécessairement ces dimensions, puisque les connaissances sur le monde sont construites en fonction de la signification que l'individu accorde à ses intérêts personnels et ses comportements. Dans ces conditions le processus d'identification du stimulus et son évaluation inclut

l'attribution de 3 étiquettes en tant que trois dimensions de l'espace cognitif-affectif. Ce modèle suggère donc que les processus cognitifs impliqués dans l'interaction avec l'environnement incluent nécessairement ces dimensions, dimensions jusque-là réservées à la caractérisation des états émotionnels.

Cette conception est intéressante à plusieurs égards. Elle permet d'abord de classifier les réponses de l'organisme (émotionnelles *vs.* non-émotionnelles) en fonction d'un seuil, un état non émotionnel se caractérisant par une variation habituelle de valeurs liées aux normes individuelles, l'état émotionnel étant atteint avec le dépassement du seuil pour au moins l'une des trois dimensions. Ensuite elle instaure une continuité entre les états et processus émotionnels *vs.* non émotionnels. Enfin elle rend compte du fait que tout état d'un organisme n'est jamais affectivement neutre.

15.5. Conclusion

Nous avons réalisé une large revue des études dans le domaine du traitement des états émotionnels et de la parole, cherchant à replacer les analyses dans leurs modèles, les modèles dans leurs concepts, les concepts dans leurs paradigmes, et les paradigmes dans leur historicité. Au terme de cette étude, on a le sentiment que les paramètres prosodiques de l'émotion, malgré l'immense variabilité intralangue et interlangue, semblent converger à travers les langues, vers une récurrence de leurs traits, pour ne pas dire une invariance. Parmi ceux-ci, on retrouve généralement les variations extrêmes des valeurs de F0 (maximum, minimum), l'amplitude forte *vs.* faible de F0, les variations de débit. Ce qui peut varier en fait d'une langue à une autre, c'est l'usage des quantités attaché à ces indices, ainsi que les patrons.

Cette convergence concerne sans doute moins les états émotionnels spécifiques, car certains d'entre eux résistent encore à une discrimination. Ceci est sans doute explicable par l'insuffisance de nos études, mais aussi par une certaine indifférenciation naturelle de l'usage des indices prosodiques (F0, valeurs et contours, intensité, tempo).

Ainsi ces récurrences de traits à travers les langues concerneraient davantage des *types* d'émotions, qui se partageraient une grande partie de leurs traits prosodiques. L'intérêt d'une description des états émotionnels par des dimensions telles que la valence, l'activation, la dominance, est sans doute d'offrir une grille d'analyse contrastive aux états émotionnels, et à leurs paramètres associés, et ainsi de constituer des catégories d'états émotionnels sur des critères de plus grande proximité de traits ou de configurations semblables des valeurs de ces dimensions. On pourrait poser l'hypothèse que ce qui pourrait être l'objet d'une tendance à l'invariance prosodique dans un ensemble de langues, ce serait non pas les émotions particulières, mais ces catégories d'émotions que les dimensions permettent de discriminer grâce à la proximité de leurs configurations de traits.

15.6. Bibliographie

- [ARN 60] ARNOLD M., « Emotion and personality », vol. 1, *Psychological aspects*, New York, Columbia University Press, 1960.
- [AVE 80] AVERILL J., « A constructivist view of emotion », PLUTCHIK R., KELLERMAN H., eds., *Emotion: Theory, Research and Experience*, vol. 1, New York, Academic Press, 1980.
- [BOE 01] BOERSMA P., « Praat, a system for doing phonetics by computer », *Glott International*, 5, 9-10, p. 341-34., <http://www.praat.org>, 2001.
- [BOL 45] BOLINGER D., « Spanish intonation », Review of Tomás Navarro, *Manual de entonación española*, *American Speech*, 20, p. 128-130, 1945.
- [BOL 46] BOLINGER D., « The intonation of quoted questions », *Quarterly Journal of Speech*, 32, p. 197-202, 1946.
- [BOL 65] BOLINGER D., *Forms of English: Accent, Morpheme, Order*, ABE I., KANEKIYO T., eds., Cambridge, Harvard University Press, Tokyo, Hokuou, 1965.
- [BOL 72] BOLINGER D., *Intonation*, Harmondsworth, Penguin, 1972.
- [BOL 86] BOLINGER D., « Intonation and emotion », *Quaderni di Semantica*, 7, p. 13-21, 1986.
- [CAE 78/81] CAELEN-HAUMONT G., *Structures prosodiques de la phrase énonciative simple et étendue*, Thèse de doctorat de 3ème cycle, Université du Mirail, Toulouse, & Hamburger Phonetische Beitrage, Band 34, Hamburg Buske, 1978, 1981.
- [CAE 91] CAELEN-HAUMONT G., *Stratégies des locuteurs en réponse à des consignes de lecture d'un texte : analyse des interactions entre modèles syntaxiques, sémantiques, pragmatiques et paramètres prosodiques*, Thèse de doctorat d'état, Aix-en-Provence, 1991.
- [CAE 02] CAELEN-HAUMONT G., « Towards naturalness or the challenge of subjectivity », in KELLER E., BAILLY G., MONAGHAN A., TERKEN J., HUCKVALE M., eds., *Improvements in Speech Synthesis*, John Wiley, London, p. 353-362, 2002.
- [CAE 00] CAELEN-HAUMONT G., BEL B., « Le caractère spontané dans la parole et le chant improvisés : de la structure intonative au mélisme », *Revue Parole*, p. 251-302, 2000.
- [CAE 04] CAELEN-HAUMONT G., AURAN C., « The phonology of melodic prominence: the structure of melisms », *Proceedings of Speech Prosody 2004*, Nara, Japon, p. 143-146, 2004.
- [CHE 02] CHEN A., GUSSENHOVEN C., RIETVELD A., « Language-specific uses of the effort code », *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, <http://www.lpl.univ-aix.fr/sp2002/>, 2002.
- [COR 00] CORNELIUS R.R., « Theoretical approaches to emotion », *Proceedings of the ISCA workshop on speech and emotion*, Belfast, Proceedings on line www.qbc.ac.uk/en/isca/proceedings, 2000.

- [COW 00] COWIE R., « Describing the emotional states expressed in speech », *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Proceedings on line, <http://www.qbc.ac.uk/en/isca/proceedings>, 2000.
- [COW 00b] COWIE R., DOUGLAS-COWIE E., SAVVIDOU S., MCMAHON E., SAWEY M., SCHRÖDER M., « “Feeltrace” : an instrument for recording perceived emotion in real time », *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Proceedings on line, <http://www.qbc.ac.uk/en/isca/proceedings>, 2000.
- [COU 86] COUPER-KUHLEN E., *English Prosody*, London, Edward Arnold, 1986.
- [CRY 75] CRYSTAL D., *The English Tone of Voice*, London, Edward Arnold, 1975.
- [CRY 64] CRYSTAL D., QUIRK R., *Systems of Prosodic and Paralinguistic Features in English*, Mouton, 1964.
- [DAM 94] DAMASIO A. R., *L'Erreur de Descartes*, Paris, Odile Jacob, 1994.
- [DAR 1872] DARWIN C., *The Expression of the Emotions in Man and Animals / L'Expression des émotions chez l'homme et les animaux*, Paris, Rivages, 1872 / 2001.
- [DES 1649] DESCARTES R., *Traité des passions de l'âme*, Paris, Garnier Flammarion, 1649 / 1996.
- [EKM 76] EKMAN P., FRIESEN W. V., SCHERER K.R., « Body movements and voice pitch in deceptive interaction », *Semiotica*, 16, p. 23-27, 1976.
- [EKM 84] EKMAN P., « Expression and the nature of emotions », SCHERER K., EKMAN P., eds., *Approaches to Emotions*, Hillsdale, Erlbaum, 1984.
- [FAI 41] FAIRBANKS G., HOAGLIN L.W., « An experimental study of the duration characteristics of the voice during the expression of emotion », *Speech Monographs*, 8, p. 85-90, 1941.
- [FAI 39] FAIRBANKS G., PRONOVOST W., *Speech Monographs*, 6, p. 87-104, 1939.
- [FAU 70] FAURE G., « Contribution à l'étude du statut phonologique des structures prosodématiques », *Studia Phonetica*, 3, p. 93-107, 1970.
- [FAU 73] FAURE G., « Tendances et perspectives de la recherche intonologique », *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, p. 5-29, 1973.
- [FON 82a] FONAGY I., « Variations et normes prosodiques », *Folia Linguistica*, 161/4, The Hague, Mouton de Gruyter, p. 17-39, 1982.
- [FON 73] FONAGY I., BERARD E., « Questions totales et implicatives », *Studia Phonetica*, 8, p. 53-98, 1973.
- [FON 82b] FONAGY I., *Vive voix. Essais de psycho-phonétique*, Paris, Payot, 1982.
- [FON 82c] FONAGY I., BERARD E., FONAGY J., « Les clichés mélodiques », *Folia Linguistica*, 161/4, The Hague, Mouton de Gruyter, p. 153-185, 1982.
- [FON 97] FONAGY I., SAP J., « Traits prosodiques distinctifs de certaines attitudes intellectuelles et émotives », *Actes des 8^e Journées d'Études sur la Parole*, Aix-en-Provence, p. 237-246, 1997.

- [FRI 86] FRIJDA N.H., *The Emotions*, Cambridge University Press, Cambridge, GB, 1986.
- [HAL 94] HALLIDAY M.A.K., *An Introduction to Functional Grammar*, 2ème édition, London, Edward Arnold, 1994.
- [HAR 86] HARRÉ R., *The Social Construction of Emotions*, Oxford, Basil Blackwell, 1986.
- [HIR 93] HIRST D., ESPESSER R., « Automatic labelling of fundamental frequency using a quadratic spline function », *Travaux de l'Institut de Phonétique d'Aix*, 15, p. 71-85, 1993.
- [KEH 02] KEHREIN R., « The prosody of authentic emotions », *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, <http://www.lpl.univ-aix.fr/sp2002/>, 2002.
- [LAD 85] LADD D.R., SILVERMAN K.E.A., TOLKMITT F., BERGMAN G., SCHERER K.R., « Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker effect », *J. Acoust. Soc. Am.*, 78, p. 435-444, 1985.
- [LAV 68] LAVER J., « Voice quality and indexical information », *British Journal of Disorders of Communication*, 3, p. 43-54, 1968.
- [LEO 70] LEON P.R., « Systématique des fonctions expressives de l'intonation, Analyse des faits prosodiques », *Studia Phonetica*, 3, p. 56-71, 1970.
- [LEO 71] LEON P.R., « Essais de phonostylistique », *Studia Phonetica*, 4, 1971.
- [LEO 76] LEON P.R., « De l'analyse psychologique à la catégorisation auditive et acoustique des émotions dans la parole », *Journal de Psychologie*, 3-4, p. 305-324, 1976.
- [LEV 90] LEVENSON R.W., EKMAN P., FRIESEN W.V., « Voluntary facial action generates emotion-specific autonomic nervous system activity », *Psychophysiology*, 27, p. 363-384, 1990.
- [LEV 92] LEVENSON R.W., « Autonomic nervous system differences among emotions », *American Psychological Society*, 3, 1, p. 23-27, 1992.
- [LIE 00] LIEBERMAN M., « Intuition: a social cognitive neuroscience approach », *Psychological Bulletin*, 126 1, p. 109-137, 2000.
- [MUR 93] MURRAY L.R., ARNOTT J.L., « Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion », *J. Acoust. Soc. Am.* 93(2), p. 1097-1108, 1993.
- [O'C 73] O'CONNOR J.D., ARNOLD G., *Intonation of Colloquial English*, London, Longman, 1973.
- [OHA 83] OHALA J. J., « Cross-language use of pitch: an ethological view », *Phonetica*, 40, p. 1-18, 1983.
- [PEL 02] PELL M. D., « Surveying emotional prosody in the brain », *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, <http://www.lpl.univ-aix.fr/sp2002/>, 2002.
- [PIA 70] PIAGET J., *Epistémologie génétique*, Paris, PUF, 1970.

- [PIO 01] PIOT O., « Ignorance, empathie et motivation : une évaluation de leurs expressions dans la prosodie du français », *Actes des Journées Prosodie*, Grenoble, p. 139-143, 2001.
- [PLU 80] PLUTCHIK R., *Emotion: A Psychoevolutionary Synthesis*, New York, Harper and Row, 1980.
- [PRI 75] PRIETO L. J., *Pertinence et pratique*, Paris, éditions de Minuit, 1975.
- [ROA 00] ROACH P., « Techniques for the phonetic description of emotional speech », *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Proceedings on line, <http://www.qbc.ac.uk/en/isca/proceedings>, 2000.
- [SHA 92] SHAVER P.R., WU S., SCHWARTZ J.-C., « Cross-cultural similarities and differences in emotion and its representation: A prototype approach », CLARK M.S., ed., *Emotion*, Newbury Park, Sage, 1992.
- [SCH 84a] SCHERER K.R., « On the nature and function of emotion: A component process approach », SCHERER K., EKMAN P., eds., *Handbook of Methods in Nonverbal Behavior Research*, Hillsdale, New Jersey, Lawrence Erlbaum, p. 293-318, 1984.
- [SCH 86] SCHERER K.R., « Voice, stress and emotion », APPLEBY M.H., TRUMBULL R., eds., *Dynamics of Stress*, New York, Plenum Press, p. 157-179, 1986.
- [SCH 87] SCHERER K.R., « Toward a dynamic theory of emotion: The component process model of affective states », *Geneva Studies in Emotion and Communication*, I, p. 1-98, 1987.
- [SCH 89] SCHERER K.R., « Vocal correlates of emotion », WAGNER H., MANSTEAD T., eds., *Handbook of Psychophysiology: Emotion and Social Behavior*, London, Wiley, 1989.
- [SCH 84b] SCHERER K.R., LADD D.R., SILVERMAN K.E.A., « Vocal cues to speaker affect: Testing two models », *Journal of the Acoustical Society of America*, 76 5, p. 1346-1356, 1984.
- [SCH 03] SCHERER K.R., BÄNZIGER T., GRANDJEAN D., « L'étude de l'expression vocale des émotions : mise en évidence de la dynamique des processus affectifs », COLLETTA J.M., TCHERKASSOF A., eds., *Les Emotions : Cognition, langage et développements*, Bruxelles, Mardaga, p. 39-58, 2003.
- [SCH 02] SCHIRMER A., KOTZ S.A., « Sex differentiates the STRRP-effect in emotional speech: ERP evidence », *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, <http://www.lpl.univ-aix.fr/sp2002/>, 2002.
- [SMI 85] SMITH C.A., ELLSWORTH P.C., « Patterns of cognitive appraisal in emotion », *Journal of Personality and Social Psychology*, 48, p. 813-838, 1985.
- [SPE 89] SPELLMAN E.V., « Anger and insubordination », GARRY A., PEARSAIL M., eds., *Women, Knowledge, and Reality: Explorations in Feminist Philosophy*, Boston, Unwin Hyman, 1989.
- [WIC 00] WICHMAN A., « The attitudinal effects of prosody, and how they relate to emotion », *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, Proceedings on line, <http://www.qbc.ac.uk/en/isca/proceedings>, 2000.

- [WIC 02] WICHMAN A., « Attitudinal intonation and the inferential Process », *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, <http://www.lpl.univ-aix.fr/sp2002/>, 2002.
- [WIL 72] WILLIAMS C.E., STEVENS K.N., « Emotions and speech : some acoustical correlates, » *J. Acoust. Soc. Am.*, 52, 4/2, p. 1238-1250, 1972.
- [ZEI 02] ZEI-POLLERMANN B., « A place for prosody in a unified model of cognition and emotion », *Proceedings of the 1st International Conference on Speech Prosody*, SP 2002, Aix-en-Provence, Proceedings on line, <http://www.lpl.univ-aix.fr/sp2002/>, 2002.