UNIVERSITE PARIS III - SORBONNE NOUVELLE

Institut de Linguistique et Phonétique Générales et Appliquées

L'expression et la perception de l'émotion extraite de la parole spontanée : évidences du coréen et de l'anglais

THESE

pour le doctorat (en 2000)

Présentée par Soo-Jin CHUNG

Sous la direction du Professeur :

Jacqueline Vaissière

Université de la Sorbonne Nouvelle (Paris III)

Remerciements

Mes plus vifs remerciments vont à Mme. Jacqueline VAISSIERE, professeur, responsable du DEA de Phonétique aux multi-universités de Paris III-V-VII, qui a dirigé mon travail avec les discussions précieuses et les suggestions brillantes. Son esprit scientifique et son encouragement constructif m'ont toujours guidée malgré la distance tout au long de la préparation de cette thèse. Elle m'a initiée à la science phonétique et m'a appris la passion de la recherche. Son enthousiasme envers les étudiants demeura en moi en tant que modèle d'enseignant.

Je remercie également M. Philip LIEBERMAN, professeur du département des sciences linguistique et cognitive à l'université de Brown, qui m'a invitée dans son département. Nos discussions se sont révélées enrichissantes et fructueuses. Le séjour à Brown m'a permis de travailler avec une grande facilité expérimentale dans un excellent environnement académique. Son aide aussi bien scientifique que pédagogique était indispensable pour la continuation de mes recherches aux Etats-Unis.

J'exprime ma reconnaissance à M. Klaus SCHERER, professeur de FPSE à l'université de Genève professeur, dont je me suis largement inspirée dans la recherche de l'émotion vocale. J'étais bien motivée par son intérêt pour ma recherche lors de notre rencontre aux JEP en 1998. Quel honneur qu'il ait bien voulu accepter d'être un des prérapporteurs de ma thèse.

Ma reconnaissance s'adresse également à M. Ivan FONAGY, ancien professeur de l'université de Paris III, d'avoir bien voulu accepter de faire partie de mon jury. J'ai beaucoup apprécié ses commentaires sur mon travail dès le premier contact. Quel honneur qu'il est assisté pour moi à la présentation de cette thèse.

Je tiens à remercier M. Jean-Yves DOMMERGUES, professeur de l'université de Paris VII, pour ses commmentaires précieux concernant mes analyses statistiques. Grâce à ces analyses, j'ai pu relever des facteurs significatifs dans l'interprétation des résultats de mes expériences.

Je voudrais aussi exprimer ma reconnaissance à Mme. Mary-Annick MOREL, professeur de l'université de Paris III, d'avoir bien voulu accepter de faire partie de mon jury. Son travail m'a beaucoup inspirée dans la recherche de la parole spontanée.

Je remercie M. Bernard GAUTHERON, ingénieur de recherches à l'université de Paris III, et M. John MERTUS, ingénieur de recherche à l'université de Brown, pour m'avoir aidé à manipuler des instruments expérimentaux. Grâce à eux, j'ai pu examiner mes données d'analyse de façon systématique avec des mesures quantitatives.

Je tiens à exprimer ma profonde gratitude à Mme. Mijin LEE et à M. Yongil LEE, pour leur prière et leur conseil pratique. Que Charles, Caroline et Deborah, trouvent ici aussi ma reconnaissance pour leur prière aux moments difficiles où je sentais perdue. Je dois beaucoup à mes amis de Paris et de Boston dont la liste serait trop longue à énumérer. Merci à tous ceux qui m'ont aidée scientifiquement, méthodologiquement ou moralement pour l'avancement de ce travail.

Comment remercier mon mari, mon amour, Sangik OH! Il sait comment m'aimer et comment me soutenir. Que mon mari sache combien j'ai été touchée par son amour et son aide. Un grand merci à ma soeur et mon frère, Sooyun et Jaehoon, et à mes belles-soeurs et mon beau-frère, Sangeun, Hyewon et Jaeik, pour leur soutien amical.

Je voudrais aussi exprimer ma reconnaissance à mes beaux-parents, Jungil OH et Soonock KWON, pour leur encouragement et leur affection.

Enfin et surtout, toute ma gratitude va à mes parents, Injik CHUNG et Minja KWON, pour leur soutien moral et matériel. Leur prière m'a toujours encouragée à poursuivre mes études. Sans eux, cette étude n'aurait pas vu le jour.

Je dédie ce mémoire à mes parents.

Paris, juillet 2000

« *Je t'instruirai et te montrerai la voie que tu dois suivre.*
Je te conseillerai, j'aurai le regard sur toi. ». –Psaum 32 :8. - Merci Seigneur.

Table de matière

I.

Chapitre I

Introduction Générale

"*La **vive voix** s'oppose en français, comme en d'autres langues, à la **lettre morte**.*" - Fónagy (1983, p9)

I.1. Objectif

La PAROLE, le propre de l'homme, est un moyen de communication extrêmement subtil et riche. Elle véhicule non seulement le message informatif référentiel mais aussi des informations sur la personnalité et l'état émotionnel du locuteur. L'EMOTION est une réponse motivationnelle et adaptative d'un organisme à l'environnement social. Elle fait partie de la vie quotidienne chez l'homme et sa forme primitive se trouve aussi chez l'animal. Malgré la longue histoire de l'étude de la parole et de celle de l'émotion, relativement peu de travaux ont été consacrés à l'analyse de la PAROLE EMOTIONNELLE (*émotion vocale*). Ce fait peut être attribué au formalisme des sciences modernes et à la difficulté méthodologique de l'étude de l'émotion vocale. En linguistique formelle, les chercheurs se sont préoccupés de décrire la régularité binaire de la langue plutôt que la variabilité pluri-fonctionnelle de la parole[1], et ils considèrent la variation vocale due à l'émotion comme une variable aléatoire, non-systématique, qui ne mérite pas d'analyse scientifique. Dans la psychologie de l'émotion, les chercheurs se sont surtout intéressés à démontrer la nature de l'émotion (soit pré-cognitive, soit post-cognitive) plutôt qu'à décrire l'expression de l'émotion vocale, faciale ou corporelle. De plus, la plupart des études psychologiques de l'émotion sont basées sur des données d'émotion faciale plutôt que celles d'émotion vocale, parce que les données de ces dernières sont plus difficiles à acquérir que celles des premières. Le formalisme se trouve aussi dans le domaine de la technologie vocale : la plupart des systèmes de la synthèse et de la reconnaissance vocale ont été développés à partir d'une modélisation de la structure syntaxique d'une langue donnée sans considération de la variation stylistique et émotionnelle (Scherer, 1998, p249).

[1] Dans le « *Cours de linguistique générale* », Saussure (1916) distingue entre *langue* et *parole*. La *langue* réfère au système abtrait de la connaissance linguistique partagée entre les membres dans une communauté donnée, tandis que la *parole* réfère à la réalisation individuelle concrète de cette connaissance dans un contexte d'énonciation particulière.

Récemment, les chercheurs réalisent de plus en plus l'importance de la connaissance sur la parole émotionnelle pour le développement des recherches dans le domaine pratique et dans le domaine théorique. Le mécanisme global de la communication parlée peut être mieux compris par la connaissance de l'influence de l'émotion sur la production et la perception de la parole (voir Fónagy, 1983a ; Léon, 1993). La nature de l'émotion peut être mieux comprise par l'étude des rapports entre les expressions vocale, faciale et corporelle de l'émotion. Dans la technologie de la parole, l'ajout des traits personnels et émotionnels est essentiel pour augmenter le caractère naturel de la parole synthétique. La reconnaissance automatique de la parole pourra largement bénéficier du développement du système qui peut reconnaître la parole de différents locuteurs dans leurs différents états émotionnels.

La présente thèse est une étude de la parole émotionnelle sur base des données acquises à partir de situations réelles. Elle vise à démontrer comment l'émotion du locuteur est exprimée dans sa parole naturelle en termes d'indices acoustiques et comment l'auditeur la perçoit dans différentes conditions d'audition. Les paramètres étudiés sont la culture de l'auditeur, la taille des stimuli vocaux présentés à l'auditeur, la position de la partie initiale, médiane ou finale dans l'énoncé et le contour de Fo et la durée des stimuli vocaux. Les données d'analyses consistent en des extraits de discours spontanés, exprimant la joie et la tristesse (voix larmoyante) de locutrices Coréenne et Américaines. Les émotions analysées dans ce travail sont considérées comme les vraies émotions *vécues* par le sujet parlant, par opposition des émotions *stylisées* imitées par un acteur. Le choix d'émotions réellement vécues (et non stylisées) repose sur la considération suivante. Etant donné que l'enregistrement des émotions vocales à partir de situations naturelles cause des problèmes sur le plan méthodologique et moral, la plupart des études se basent sur les données d'émotions exprimées par un acteur suivant des instructions de l'expérimentateur sur la façon de s'exprimer. Bien que ces émotions théâtrales soient censées être représentatives de celles qu'on éprouve dans la vie quotidienne, les deux sortes d'émotion sont pourtant différentes aux niveaux sentimental, motivationnel et expressif (Ekman *et al*., 1972, p35-38 cité par Scherer, 1979, p512). L'expression de l'émotion par l'acteur prend souvent une forme exagérée et implique un certain style théâtral, tandis que l'expression de l'émotion par des gens normaux dans l'interaction sociale quotidienne prend plutôt une forme discrète et suit des règles d'exposition régulatrices. Vu la différence entre l'émotion vécue et l'émotion stylisée et la rareté des études de l'émotion vécue, nous avons décidé

d'étudier les données d'expressions émotionnelles naturelles, extraites d'entretiens improvisés. Bien que ces entretiens aient été diffusés à travers la télévision, les émotions qui y sont exprimées sont considérées comme authentiques par rapport à ce que la personne ressentait. Le but des entretiens était d'aider l'invité à trouver des solutions à son problème personnel à travers une discussion publique télévisée. La personne présentait son problème en tant que tel, et son état émotionnel était exprimé de façon naturelle au fur et à mesure du discours raconté. Ce genre d'entretien est différent d'un autre genre d'entretien télévisé ('*talk-show*' en anglais) dont le but est d'amuser le public avec une histoire racontée par un invité.

Cette dissertation consiste en une série d'analyses acoustiques sur les énoncés émotionnels et neutres et en une série d'expériences perceptives sur l'identification de l'émotion par des auditeurs coréens, américains et français. Les problématiques des analyses sont liées les une aux autres autour de cinq questions principales : (1) Comment l'émotion du locuteur est-elle exprimée dans son discours spontané au niveau prosodique et au niveau lexical ? ; (2) Quels sont les meilleurs indices acoustiques pour repérer l'excitation émotionnelle de la joie et celle de la tristesse (*détresse*) ? ; (3) Comment la modification des indices prosodiques influence-t-elle la perception de l'émotion positive (comme la joie) ou de l'émotion négative (comme la tristesse) ? ; (4) Est-ce que l'émotion est exprimée de façon uniforme au cours de l'énoncé et est-elle toujours bien identifiée dans toute partie (initiale, médiane ou finale) de l'énoncé ? Ou bien, l'émotion est-elle exprimée différemment en fonction des différentes parties de l'énoncé et est-elle mieux perçue dans certaines parties que dans d'autres ? ; (5) Dans quelle mesure la perception des émotions primaires (joie et tristesse) est-elle universelle parmi les auditeurs provenant de différentes cultures ? La démarche de nos analyses acoustiques et perceptives sur les données d'expressions émotionnelles sera expliquée dans la section suivante.

I.2. Plan de la thèse

La présente dissertation se compose de sept chapitres : l'introduction générale, la revue des études précédentes, la méthodologie conceptuelle de la recherche sur l'émotion exprimée dans la parole spontanée, l'étude principale faite sur un corpus coréen, l'étude comparative faite avec un corpus anglais, une mise à l'épreuve des résultats à partir de stimuli synthétiques, la suggestion des applications de nos résultats dans le domaine de la technologie vocale et la conclusion générale. Chaque chapitre commence par une introduction de la problématique de l'analyse et finit par une discussion conclusive et suggestive.

Après avoir présenté l'objectif et le plan de cette thèse dans ce chapitre d'introduction générale, nous allons réviser dans le chapitre II des études précédentes sur l'émotion vocale dans divers domaines comme la psychologie, la linguistique (phonétique), la sémiotique, la philosophie et l'ingénierie. Cette révision donnera un aperçu de ce qui a été découvert et de ce qui doit être étudié dans la recherche de l'émotion vocale, en particulier en ce qui concerne l'expression et la perception de l'émotion à travers la voix. Nous allons aussi définir différents termes émotionnels, dont la définition varie largement d'une analyse à l'autre.

Dans le chapitre III, nous allons introduire une distinction entre l'*émotion vécue* et l'*émotion stylisée*, qui est similaire à la distinction faite par Fagyal (1995) entre la *parole spontanée* et la *parole lue*. Nous basant sur cette distinction, nous allons expliquer le choix de notre méthodologie pour les analyses du présent travail. La spontanéité des données phonétiques est inversement liée à la possibilité du contrôle expérimental. Les données authentiques spontanées impliquent une non-intervention de l'expérimentateur (le contrôle expérimental), tandis que l'expérimentation scientifique rigoureuse exige un haut degré de contrôle lors de la structuration des données. Cette situation contradictoire sera expliquée en termes de l'échelle de l'authenticité des données et du contrôle expérimental. A la fin, nous allons proposer une approche typologique fonctionnelle pour la construction scientifique des données de nature spontanée, en tant que solution des problèmes méthodologiques de l'étude de l'émotion vécue.

Dans le chapitre IV, nous allons présenter l'étude principale du corpus coréen. Ce chapitre consiste en cinq sous-chapitres majeurs. Dans la partie IV.2, nous allons rapporter la procédure de l'acquisition des données et les facteurs concernés du point de vue typologique fonctionnel. Le profil de la locutrice Coréenne et le protocole de la segmentation des énoncés dans le corpus coréen seront aussi présentés dans cette partie. La partie IV.3 concerne l'analyse descriptive du corpus coréen dont la question problématique est la suivante : comment l'émotion de notre locutrice est-elle exprimée dans son discours spontané aux niveaux prosodique et lexical ? Le degré d'intensité (*valeur d'activation*) et le degré de positivité (*valeur de valence*) de l'émotion de chaque énoncé seront décrits d'une manière objective avec recours au jugement de dix auditeurs coréens à propos des stimuli vocaux et des stimuli écrits dans deux tests de perception. Dans IV.4, nous présenterons nos mesures acoustiques des énoncés du corpus coréen : Fo moyen, Fo maximum, Fo minimum, moyenne des 20% des valeurs les plus basses de Fo ('*Fo Moy Bas*'), plage de Fo, perturbation de Fo ('*jitter*'), perturbation d'intensité ('*shimmer*') et débit de parole. Les valeurs acoustiques seront comparées entre les énoncés neutres et les énoncés émotionnels (de la joie ou de la tristesse) afin de connaître la modification des traits prosodiques par l'excitation émotionnelle. Ces valeurs acoustiques seront aussi comparées aux valeurs perceptuelles (*valeur d'activation* et *valeur de valence*) dans la partie IV.5, ce qui montrera l'influence de la variation des traits prosodiques sur la perception de l'émotion. La partie IV.5 contient aussi une expérience perceptive sur la perception de l'émotion par les auditeurs de différentes cultures (Coréens, Français et Américains). La similarité et la différence de la perception de l'émotion parmi ces trois groupes d'auditeurs seront interprétées en terme de l'universalité et de la spécificité culturelle de l'émotion. Dans la partie IV.6, nous allons présenter une analyse sur la communication de l'émotion dans l'unité d'énoncé. Il s'agit de savoir si l'émotion est exprimée et perçue différemment en fonction des parties initiale, médiane et finale l'énoncé. Cette question adresse spécifiquement *la particularité de la partie finale de l'énoncé dans la communication émotionnelle*, vu le rôle important de la prosodie de la partie finale de l'énoncé dans la communication de la modalité linguistique. Cette problématique constituant un aspect original de cette thèse, elle sera examinée de nouveau avec le corpus anglais et les données de la parole synthétique dans les chapitres suivants.

Dans le chapitre V, une étude comparative du corpus anglais sera présentée concernant la manifestation de l'émotion de détresse dans la voix et la reconnaissance de cette émotion par l'auditeur en fonction de la présentation des différentes parties de l'énoncé, initiale, médiane et finale. Après une brève introduction des problèmes à aborder dans la partie V.1, la procédure de la construction du corpus anglais est présentée dans la partie V.2. Le corpus anglais est établi à partir des discours spontanés de cinq locutrices Américaines. Chaque discours contient des énoncés produits dans un état émotionnel de détresse et dans un état neutre. Les caractéristiques acoustiques des énoncés émotionnels et des énoncés neutres seront examinées dans la partie V.3. La variation des traits prosodiques à l'intérieur de l'énoncé sera aussi analysée dans cette analyse acoustique. Une expérience perceptive dans la partie V.4 examinera le statut spécial de la partie finale de l'énoncé dans la communication de l'émotion de détresse. Comme dans l'expérience de la partie IV.6, les stimuli des parties initiale, médiane et finale de l'énoncé seront présentées à l'auditeur sous forme isolée, et nous allons voir si la reconnaissance de la détresse en larmes est meilleure dans une partie (par exemple, la partie finale) que dans les autres parties (initiale et médiane) de l'énoncé. Le résultat de cette expérience sera discuté en comparaison avec celui de l'expérience de la partie IV.6.

Dans le chapitre VI, nous allons présenter deux expériences avec des stimuli synthétiques, dont le but général est de vérifier les résultats de nos expériences précédentes. A la suite de la constatation de l'influence des indices acoustiques sur la perception de l'émotion, nous modifierons dans la première expérience (partie VI.2.1) le contour de Fo et la durée d'une partie de l'énoncé au moyen de la resynthèse vocale, et verrons si cette modification produit un effet sur la perception de la positivité émotionnelle. Dans cette expérience, les Coréens et Américains jugeront l'émotion (POSITIVE ou NEGATIVE) des stimuli resynthétisés en ce qui concerne la variation de quatre types de contour de Fo (*montant*, *descendant*, *montant-descendant* et *plat*) et de deux types de durée (*longue* et *courte*). Dans la deuxième expérience (partie VI.2.2), la pertinence de la partie finale de l'énoncé dans la perception émotionnelle sera examinée avec des stimuli modifiés par la resynthèse. Il s'agit de la position efficace des indices prosodiques dans l'énoncé pour communiquer une nuance émotionnelle du locuteur. Dans cette expérience, l'émotion des stimuli sera jugée - soit émotion POSITIVE, soit émotion NEGATIVE - en fonction du placement de l'indice prosodique (e.g. un cliché intonatif *montant-descendant*) dans la position initiale, médiane ou finale de l'énoncé. Dans les deux expériences, les résultats des

Coréens et des Américains seront comparés en ce qui concerne leur similarité ou leur différence perceptuelle.

Dans le chapitre VII, nous allons résumer les points originaux du présent travail, présenter plusieurs applications potentielles de nos résultats dans la technologie vocale et conclure la thèse avec notre perspective du travail futur. Les applications potentielles comprennent l'adéquation de la voix synthétique au contexte par l'ajout d'effet émotionnel et le renforcement du système de la reconnaisance automatique de l'émotion du locuteur. Quelques projets acutels sur la reconnaissance automatique de l'état psychologique du sujet devant l'ordinateur chez une entreprise informatique IBM et dans le laboratoire de multimédia (*Media Laboratory*) à l'institut de la technologie de Massachusetts (*MIT*) seront présentés en tant qu'exemples de la recherche appliquée. En conclusion, des points forts et des points faibles de nos analyses seront notés, et nous discuterons de comment améliorer et comment approfondir la recherche de la *communication de l'émotion dans la parole spontanée*.

Chapitre II

Rappel des études précédentes sur les différents aspects de l'émotion

Résumé

Ce chapitre présente une revue des études précédentes sur la communication de l'émotion à travers la voix dans les domaines sémiotique, philosophique, psychologique, linguistique, et de l'ingénieur. Cette revue comprend les théories de la communication (Bühler et Shannon), l'histoire de l'étude de l'émotion (depuis la philosophie ancienne grecque jusqu'à la psychologie moderne), les théories de l'émotion (Darwin, Tomkins, et Plutchik) et les modèles de la communication de l'émotion vocale (Fónagy, Léon, et Scherer) et les résultats de diverses expériences concernant les indices acoustiques et perceptifs de l'émotion vocale (Lieberman & Michael, Williams & Stevens, Murray & Arnott, etc.). Parmi les modèles de la communication de l'émotion vocale, nous présentons en particulier le modèle de la covariation et le modèle de la configuration que nous allons utiliser pour l'explication de nos résultats d'expériences dans les chapitres suivants. Sur le plan conceptuel, nous nous attardons sur les définitions des divers termes émotionnels comme 'affect,' 'humeur ('mood'),' 'émotion,' 'sentiment,' 'passion' et 'attitude,' et puis explicitons la définition du terme "EMOTION" dans notre travail.

II. Rappel des études précédentes sur les différents aspects de l'émotion

II.1. Préliminaires

Le chapitre II résume les études précédentes sur l'émotion dans les domaines sémiotique, psychologique, linguistique et acoustique. Cette revue présente les diverses théories qui ont inspiré le cadre théorique et les méthodes d'analyse de notre travail.

L'expression vocale de l'émotion est souvent étudiée à la lumière de la notion de communication, concept universel qu'on retrouve aussi bien en sociologie, linguistique, psychologie, qu'en physiologie ou en biologie. Beaucoup de chercheurs s'intéressent au fonctionnement sémiotique de la variation vocale en fonction de différentes émotions, dans la mesure où elle produit une réaction ou un effet sur l'auditeur.

Ce chapitre se compose de quatre parties principales. La première partie présente les théories de la communication, dont les modèles de Bühler et de Shannon et la théorie du codage. La deuxième partie présente l'histoire de l'étude de l'émotion, les définitions des termes émotionnels et deux approches majeures dans la recherche de l'émotion. La troisième partie présente quatre modèles de l'expression vocale de l'émotion, proposés par des psychologues et des linguistes. La quatrième partie fait une revue des études expérimentales sur la voix émotionnelle aux points de vue acoustique et perceptive.

A la fin du chapitre, nous discuterons des problèmes majeurs liés à l'étude de l'émotion et ses potentiels d'application dans le domaine de la technologie moderne. La discussion sur le corpus et le méthode de l'étude de la parole émotionnelle sera continuée dans le chapitre suivant.

II.2. Théories de la communication

Les théories de la communication forment un paradigme interdisciplinaire qui permet d'envisager des connexions cohérentes entre disciplines parfois éloignées. Ces disciplines sont clairement distinctes mais elles ont des procédures externes qui sont susceptibles de les relier, par exemple, la sémiotique, les sciences cognitives, les traitements artificiels de l'information, la métapsychologie et la biologie du comportement.

La *communication* surgit, selon la définition de Miller (1973) citée par Miermont (1991, p1), quand des événements apparaissant dans un lieu ou à un moment donné sont étroitement reliés à des événements apparaissant dans un autre lieu et à un autre moment. Cet 'autre lieu' peut être un niveau psychologique, social ou biologique, éventuellement éloigné du lieu initialement considéré. Le paradigme communicationnel suppose que différents niveaux de modélisation, de décision et d'organisation fonctionnent de concert et il repose sur la démarche sémiologique ou sémiotique (étude des signes), avec la triple polarité : sémantique (le sens), syntaxique (la structure des comportements, des grammaires) et pragmatique (l'effet concret obtenu).

Dans la partie II.2, nous présentons deux modèles de la communication. Le modèle de Bühler identifie les trois fonctions du langage, *représentation*, *expression* et *appel*, dont la terminologie fut beaucoup reprise et modifiée par des psychologues et linguistes. Le modèle de Shannon, plutôt mathématique, propose un système général de communication, qui comprend huit paramètres, la source d'information, la destination, l'émetteur, le récepteur, le signal émis, le signal reçu, le bruit et le message.

II.2.1. Théorie de Bühler

Bühler (en allemand, 1934 ; traduction anglaise par Innis, 1982) propose un modèle triadique, qui représente trois fonctions de la *langue*. Il (ibid., p147) appelle la langue *organum* , un terme de Platon, qui permet à un sujet de communiquer avec l'autre à propos d'un objet. Cet organum possède trois aspects respectivement, *symbole*, *symptôme* et *appel*, avec les fonctions de *représentation* de l'objet ou un événement, d'*expression* de l'état du sujet (émetteur) et d'*appel* d'une réponse de l'autre (récepteur).

Par exemple, un symbole, *signe* en termes saussuriens, a une fonction de la représentation d'un message linguistique comme « Je dis que tu dois aller à l'école ». Ce message porte un aspect de symptôme par la fonction d'expression, voire la façon d'exprimer qui révèle des caractéristiques du locuteur (sexe, âge, personnalité, émotion, etc.). La fonction d'appel résulte de l'effet que le message produit sur l'auditeur, en suscitant une réponse de ce dernier.

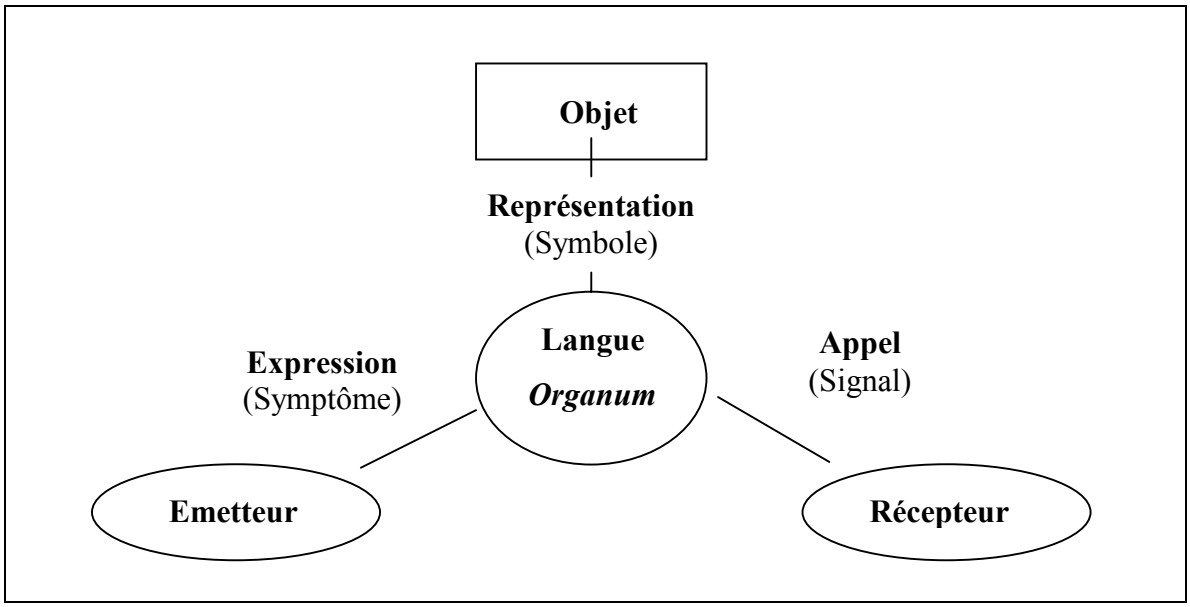


Figure 1. Modèle de Bühler (1934) sur la communication avec trois fonctions de la langue, *représentation*, *expression*, *appel*.

Bühler (1934) note aussi que toutes les caractéristiques du phénomène ou du signe ne sont pas forcément pertinentes et elles peuvent varier beaucoup en fonction de la situation de communication. Il faut préciser que le *symbole* de Bühler est un signe arbitraire, alors que dans la culture européenne le *symbole* suppose un lien motivé entre le signifié et le signifiant : la balance, *symbole* de la justice, par exemple ; ou la croix, *symbole* d'un croisement dans le code de la route (Léon, 1993, p16). Le modèle de Bühler est repris et développé en 1939 par Laziczius (traduction anglaise, 1966), et par Troubetzkoy (1939), puis par Jakobson (1963) dont les modèles seront présentés dans les parties II.2.2 et II.4.3 du présent travail.

II.2.2. Théorie de Shannon

Le modèle de Shannon (Weaver & Shannon, 1949) représente un système général de communication dans lequel huit paramètres sont pris en compte : la source d'information, la destination, l'émetteur, le récepteur, le signal émis, le signal reçu, le bruit et le message. Ce modèle sert de cadre général à différentes approches sur le processus de la communication dans les domaines qui traitent de la transmission d'information à travers divers canaux.

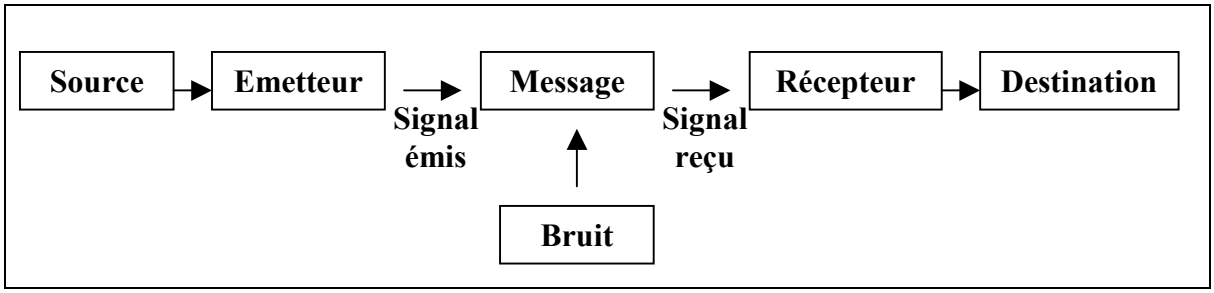


Figure 2. Modèle de Shannon, système de la communication avec huit paramètres, source d'information, destination, émetteur, récepteur, signal émis, signal reçu, bruit et message, d'après l'illustration de Miermont (1991, p2).

Miermont (1991) fournit un excellent aperçu des théories de la communication du point de vue bio-psycho-sociologique, à qui la présentation de cette partie doit largement. Il envisage les théories basées sur le modèle de Shannon, à partir de l'axiome fondamental : *un non-comportement n'existe pas chez l'homme* : avec le corollaire qui s'ensuit : *il est impossible de ne pas communiquer*. Cet axiome s'appuie sur la théorie du 'double-blind', telle qu'elle a été élaborée, développée et affinée par Bateson et ses disciples. Le 'double-blind' repose sur l'interaction entre deux personnes (ou plus) qui sont prises dans un enjeu vital, où les messages échangés dérivent logiquement les uns des autres tout en étant antinomiques, et où il est impossible de sortir de l'interaction (Bateson, 1972). Une telle théorie débouche sur une vision hiérarchisée parmi les niveaux de la communication.

L'appareil humain de communication est une entité fonctionnelle qui caractérise l'ensemble de l'activité physique et psychique. La communication dans l'interaction personnelle s'effectue à travers un circuit des systèmes biologique, sociologique et psychologique, qui sont à la fois indépendants et interdépendants. Chaque système

accomplit sa fonction propre dans l'acte de communication. La fonction biologique équilibre des régulations hormonales, vixcéro-motrices et cérébrales ; la fonction sociologique établit les liens entre la personne et autrui ; la fonction psychologique conduit à la différenciation du soi de la personne dans l'interaction sociale et stabilise son état psycho-affectif.

Par exemple, lorsque deux personnes sont engagées dans la conversation où les concepts sont associés à des images acoustiques à travers des actes individuels. Une partie du circuit de parole est purement psychique, comme l'association, chez l'émetteur, d'un concept à une image acoustique. L'autre partie est physiologique, comme la transmission aux organes de la phonation d'une impulsion corrélative à l'image acoustique, la propagation des ondes sonores de la bouche de l'émetteur au tympan du récepteur et la transmission de l'image acoustique qui est corrélée psychiquement au concept chez le récepteur. Cette corrélation des parties psychique et physiologique révèle la double face du signe linguistique, ce qui évoque les termes Saussuriens, *signifié* et *signifiant*, correspondant respectivement au concept et à l'image acoustique. Le signe linguistique est arbitraire sur le plan individuel, et motivé sur le plan social.

Bateson et Ruesch (1951, p235) distinguent les communications *interpersonnelle* et *intrapersonnelle* par le fait que dans la première il est possible d'évaluer et de corriger l'effet des actions intentionnelles ou expressives tandis que dans la dernière il est beaucoup plus difficile et parfois impossible de le faire. La communication *interpersonnelle* est constituée d'actes expressifs, verbaux et non verbaux des personnes engagées. Ces actes sont en partie volontaires, en partie involontaires. Ils sont perçus, consciemment ou inconsciemment, de manière asymétrique par l'émetteur et le récepteur. La perception réciproque des échanges entre eux crée une relation qui peut modifier la position ou l'attitude d'une personne vis-à-vis de l'autre. La communication *intrapersonnelle* est un cas particulier et indispensable à l'établissement de la communication interpersonnelle. Elle rigidifie la vie fantasmatique, ce qui crée un siège de conflits représentés mentalement, et cristallise toutes les formes de conflits autour d'un scénario unique pour agir. Cependant, les auteurs notent que la distinction entre les communications *interpersonnelle* et *intrapersonnelle* devient souvent ambiguë à cause de facteurs comme la variation de la capacité individuelle à se contrôler et le contexte de la communication.

II.2.3.Théorie du codage

La théorie du codage cherche à trouver par quelles procédures les informations sont transmises et traitées dans un système de la communication. Le *codage* peut être défini comme le processus par lequel un événement est substitué à un autre pour le représenter (Miermont, 1991, p5). Le codage repose sur les régularités constatées dans le recueil interne des données externes et suppose une 'transformation' au sens mathématique, entre l'espace-source et l'espace-but. La théorie du codage permet des travaux comparatifs entre les communications humaines, animales, et leur simulation sur ordinateurs.

Nous considérons trois types de codage, *digital*, *analogue* et *symbolique,* parmi les six identifiés par Miermont (ibid.), en tant que modalités de codage. Le codage *digital* a un caractère discontinu, entre la nature de code et l'apparence perçue du référent. Il suppose une relation biunivoque entre l'élément codé et l'objet référent (par exemple, le mot 'cheval' évoque l'animal appartenant à la famille des équidés). Ce codage est individuellement arbitraire, bien qu'il soit socialement convenu et motivé, et sa fonction est souvent d'indiquer le contenu du message dans le système de la communication.

Le codage *analogique* est fondé sur l'utilisation de grandeurs physiques continues, qui établissent une dénotation plus ou moins ressemblante entre la structure sémiotique du référent et les supports matériels du code. Il existe une continuité entre le référent et le code, ce qui reflète un aspect immédiat et motivé de ce codage. La relation entre l'élément codé et l'objet référent est multivoque, voire équivoque ou ambiguë (par exemple, le fait de miauler ou d'imiter la voix de quelqu'un ne permet pas forcément de connaître l'objet exact ainsi désigné ; il peut s'agir de l'animal ou de la personne). La teneur sémantique de ce codage est riche et la mesure de l'information se fait en 'plus ou moins'.

Le codage *symbolique* est voisin du codage digital par son caractère arbitraire, mais il se situe à un niveau plus élaboré du traitement de l'information. Un système de symboles physiques est un ensemble de formes organisées ('patterns') qui peuvent être transformées par une série de procédures. Ces processus de construction et de transformation sont relativement indépendants du substrat qui les produit. Les mots du langage, les formules

mathématiques, les thèmes musicaux possèdent les propriétés du codage symbolique, irréductibles au niveau des codages digitaux et analogiques qui participent à leur édification. Laver (1994, p17) appelle les *signes linguistiques* les *signes symboliques*. Ces signes reposent sur un lien arbitraire entre les signes et les référents et ce lien est déterminé culturellement et gouverné par des conventions sociales.

Deux aspects de codage sont pertinents dans la communication parlée ; l'aspect de *contenu* (compte rendu) et l'aspect d'*opération* (ordre performatif de la relation). Le *contenu* du message est essentiellement *digitalisé* ou *symbolisé* par l'intermédiaire du langage en servant d'indice à la communication, tandis que l'aspect *opératoire* de la relation est plutôt du type *analogique* non verbal fonctionnant comme métacommunication. Si le *contenu* de la communication obéit habituellement à une logique binaire, en tout ou rien (ce qui permet de lever l'ambiguïté du sens dans l'acte de communication), l'*opération* qui lie l'émetteur et le récepteur par l'intermédiaire d'un message obéit à une logique ternaire. Ces différents aspects de l'information radicalisent l'opposition entre les procédures de jonction et de disjonction dans tous les médias de la communication *digitale* et l'ensemble des échanges non verbaux dans une communication *analogique*. Le problème de la communication est alors de comprendre les relations entre ces aspects d'information, leurs degrés de congruence ou de désaccord, de transformation ou d'exclusion.

La communication *analogique* ou *mimétique*, tout particulièrement en ce qui concerne le domaine de l'affectivité profonde, qui s'accompagne des réactions neuro-végétatives et vixcéro-motrices, se prête mal à un compte rendu verbal. Dire « Je t'aime » n'a pas la même valeur que de le montrer par toute une série d'attitudes, de gestes, de 'preuves'. Dire « Je te déteste » peut signifier un authentique mouvement de haine ou au contraire une provocation amoureuse.

La vocalisation affective chez les primates implique une combinaison de deux systèmes, un *système fermé du signal* qui consiste en un certain nombre de types de la vocalisation, et un *système analogue qualificateur* qui repose sur une variation continue des paramètres du signal en fonction de la variation de l'état d'âme ou de situation (Scherer, 1979, p496). Par exemple, le premier système, étant *digital* ou *symbolique*, construit la structure syntaxique et le champ sémantique de l'expression vocale émotionnelle tandis que le dernier système, étant *analogique*, code la nuance et l'intensité de son émotion.

La figure suivante résume les différents codages et les fonctions du message dans la communication parlée, selon Bühler (1934), Weaver & Shannon (1949), Jakobson (1963) et Scherer (1979). Le message est construit par un codage *symbolique* et un codage *analogique* pour transmettre des informations sur le *contenu* linguistique du message (fonction *référentielle*), sur l'*émetteur* (fonction *expressive* ou *émotive*) et sur le *destinataire* (fonction *appellative* ou *conative*). On constate aussi trois autres aspects fonctionnels du message, selon Jakobson (1963), telles que l'*organisation* esthétique du message (fonction *poétique*), la manière du *contact* avec l'interlocuteur (fonction *phatique*) et l'explication du message à travers la référence des *codes* (fonction *métalinguistique*).

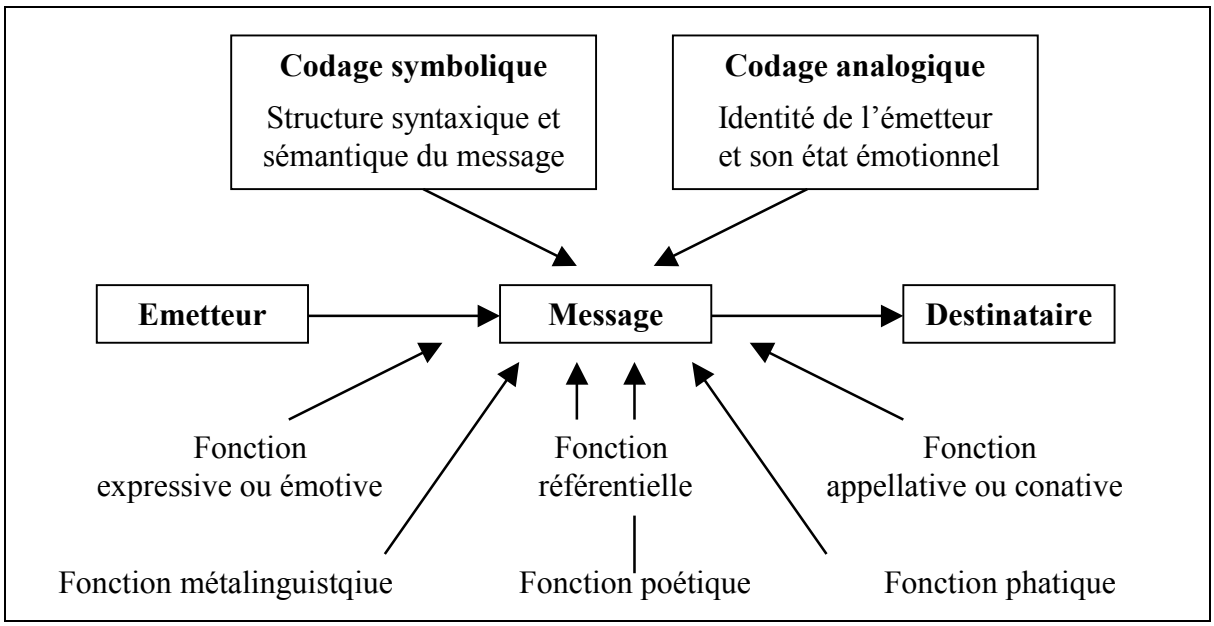


Figure 3. Deux codages et six fonctions du message dans la communication parlée.

II.3. Théories de l'émotion

L'étude de l'émotion remonte, pour son origine, jusqu'à la philosophie ancienne grèque. Le concept d'émotion s'est toujours trouvé controversé à travers les siècles. Les théories de l'émotion appartiennent à deux tendances majeures selon leur point de vue sur la relation entre l'émotion et la cognition. D'un côté, l'émotion est considérée comme une perception des modifications physiologiques par les événements extérieurs et la cognition est un facteur secondaire dans l'expérience émotionnelle. D'un autre côté, les psychologues cognitifs insistent que le processus cognitif est une étape antérieure et indispensable dans la production de l'émotion.

Dans cette partie II.3, nous nous proposons de résumer les études précédentes, particulièrement sur le concept d'émotion, d'expliquer les différents termes relatifs à l'émotion et de présenter deux perspectives majeures dans le domaine psychologique. Nous donnons d'abord un aperçu historique des études philosophiques, psychologiques, biophysiologiques et neurologiques depuis la philosophie ancienne grecque jusqu'aux études modernes. Puis, deux approches différentes dans la recherche des émotions sont présentées. Une approche met l'accent sur l'aspect primaire de l'émotion, tel que le système de l'affect est un système primaire motivationnel qui renforce d'autres processus mentaux. L'autre approche considère l'émotion comme un ensemble d'évaluations du stimulus, dont les expressions sont des traces des fonctions adaptatives, développées au cours de l'évolution.

II.3.1. Histoire de l'étude de l'émotion

II.3.1.1. Philosophes grecques

Dans la théorie classique, Démocrate (460 avant Jésus-Christ, traduction française, 1927) identifie deux composants de l'*âme*, l'une *rationnelle* ayant son siège dans la poitrine, l'autre *irrationnelle* qui est répartie dans tout le corps. Selon sa théorie atomistique, l'*esprit* est un mélange des éléments corporels, dont la fonction est altérée quand ce mélange devient trop chaud ou trop froid, et l'*émotion* est un état pathologique de

l'esprit qui cause cette altération et mène à une conduite irrationnelle. Ainsi, le bonheur de l'être humain, selon lui (ibid. p27), est de garder son esprit dans un équilibre mental et physique.

Le classement binaire des émotions positives et négatives se trouve déjà chez les philosophes anciennes grecques (voir Dumitrache, 1994, p1). Platon distingue la peine et le plaisir : la *peine* naît de la dissolution de l'harmonie originale, et le *plaisir* de sa restitution (Timée). Aristote propose plusieurs oppositions binaires, telles que *placidité* et *colère*, *amour* et *haine*, *confiance* et *peur*, *bienveillance* et *méchanceté* (Rhétorique). Les philosophes stoïques postulent quatre émotions de base comme le *désir* (latin = 'libido'), la *peur* ('metus'), le *plaisir* ('voluptas') et la *peine* ('aegritudo'), alors que Dumas (1923) identifie les quatre émotions fondamentales avec la *joie*, la *tristesse*, la *peur* et la *colère*.

II.3.1.2. Philosophie médiévale

On retrouve la conception pathologique de l'émotion chez les philosophes scolastiques aussi bien que dans la philosophie moderne. Thomas d'Aquin (1225-1274, cité dans Gardiner *et al*., 1937, p106) établit un système des *passions* dans lequel deux états d'âmes s'opposent par les catégories de passion *concupiscible* et de passion *irascible*. Chaque catégorie contient deux sous-ensembles de passion en fonction de leur relation avec le *bien* et le *mal*. Les passions dans la catégorie concupiscible traitent le bien et le mal d'une façon absolue, tandis que celles dans la catégorie irascible le font avec plus d'effort.

Passion			
Concupiscible		Irascible	
Bien	Mal	Bien	Mal
Amour	Haine	Espoir	Peur
Désir	Aversion	Désespoir	Courage
Plaisir (Joie)	Peine		Colère

Tableau 1. Système des passions de Thomas d'Aquin (dans Gardiner *et al*. 1937, p108).

II.3.1.3. Philosophie moderne

En s'inspirant de ce système des passions, Hobbes (1588-1679) explique les émotions par la notion de la *tendance de comportement* ('*endeavor*'). Pour lui, toutes les passions consistent en *appétit* et *aversion*, excepté le plaisir et la peine. L'appétit, ou désir, implique une tendance de rapprochement, et l'aversion, une tendance d'éloignement ; l'objet de l'appétit est le *bien* et celui de l'aversion est le *mal*. L'amour et la haine sont les symptômes de l'appétit et de l'aversion et le plaisir et la peine sont des réalisations pures du bien et du mal eux-mêmes.

Spinoza (1632-1677) reprend des notions stoïques comme le *plaisir*, la *peine* et le *désir*, et conclut que tous les phénomènes affectifs de l'être humain sont des variantes de ces trois émotions dans des circonstances différentes. Il distingue le *plaisir* (joie ou délice, tout ce qui vient des expériences agréables), une émotion qui rend un meilleur 'perfectionnement' de l'esprit, de la *peine* (déplaisir ou tristesse, tout ce qui vient des expériences désagréables), une émotion qui rend un mauvais 'perfectionnement' de l'esprit. Ici, le *perfectionnement* veut dire un degré relatif de la force d'action ou d'existence chez l'être humain. Le *désir*, en tant qu'un appétit conscient, est considéré comme un effort fondamental de l'homme pour maintenir son existence. Il fait la distinction entre l'émotion ('*affectus*') et la passion ('*pathema*'), l'un étant l'affection du corps dans lequel la force d'action augmente ou diminue, l'autre étant un état d'âme confus sans action qui ne fait qu'affirmer l'augmentation ou la diminution de la force d'existence (voir Gardiner *et al*., 1937,183-209).

Pour Kant (en allemand, 1798 ; traduction anglaise par Gregor, 1974), l'émotion est un état irrationnel, appelé une maladie de l'esprit. Il considère le *plaisir* et le *déplaisir* comme les éléments dominants et déterminants de l'émotion. Il définit le plaisir comme une sensation de stimulation de la vie, à l'opposé du déplaisir qui est une sensation de l'empêchement de la vie. Il fait aussi la distinction entre l'émotion ('*Affekte*') et la passion ('*Leidenschaften*') mais dans un sens différent de Spinoza. Selon la description de Kant (ibid. p119), l'émotion est une sensation ('*Gefühl*') de plaisir ou de déplaisir qui ne permet pas de réflexion rationnelle s'il faut refuser ou laisser passer la sensation tandis que

la passion est une inclinaison (‘*Neigung*’) que le sujet peut maîtriser seulement avec grande difficulté ou ne peut pas le faire.

II.3.1.4. Psychologie moderne

Jusqu’au XVIII siècle, les études de l’émotion ont été menées surtout par des philosophes. Leur spéculation est concentrée sur les aspects psychologiques de l’émotion et leur méthode principale est l’introspection consciencieuse de leurs propres états d’âme. A partir de la fin du XIX siècle, ce genre de méthode est remplacé par des méthodes expérimentales dans un laboratoire grâce aux développements méthodologiques scientifiques, surtout dans le domaine biophysiologique. La psychologie moderne commence à établir son domaine de recherche, différent de la philosophie de l’état d’âme.

La théorie de l’émotion proposée par James (1884 & 1894) et connue sous le nom de théorie de James-Lange, postule que l’émotion est une sensation des changements physiologiques qui proviennent directement de la perception des stimuli. Il (1894, p516) insiste le fait que ces changements musculaires et viscéraux, étant des réflexes immédiats à la présence de l’objet, sont les facteurs primaires dans l’expérience de l’émotion. L’évaluation cognitive et l’expression émotionnelle, selon lui, ne sont que des effets secondaires de ces modifications physiologiques. Pour dire d’une manière simplifiée, un homme est triste parce qu’il pleure et il a peur parce qu’il s’enfuit. Cette position théorique forme un des points de vue majeurs sur la relation entre l’émotion et la cognition. L’explication physiologique de James inspire essentiellement les théories d’activation, qui soulignent l’importance du degré d’excitation du centre nerveux dans l’intensité des réactions émotionnelles. Les théories de Tomkins (1962) et de Zajonc (1980) mettent aussi l’accent sur les antécédents physiologiques dans la production des émotions, en s’appuyant sur les théories d’activation.

L’autre point de vue sur la relation entre l’émotion et la cognition, appelé la psychologie cognitive, est initié par Cannon en 1927 avec son objection contre l’argument de James. Sa théorie ou la théorie de Cannon-Bard inverse le postulat de James, en affirmant que l’évaluation cognitive est un préalable de l’émotion et cette évaluation détermine les différentes réactions physiologiques ou expressives de chaque émotion. Donc, l’origine de l’émotion est cognitive, plutôt que viscérale. Cette idée est articulée

d'une manière plus explicite dans la théorie d'évaluation ('*appraisal theory*'), proposée par Arnold (1960) et d'autres : l'émotion résulte des diverses évaluations d'un stimulus externe et se manifeste dans des comportements expressifs. Arnold (ibid.) insiste sur le rôle déterminant de l'évaluation cognitive dans l'organisation des comportements émotionnels : une évaluation positive du stimulus produit une émotion positive qui déclenche un comportement de rapprochement ; une évaluation négative cause une émotion négative avec un comportement d'éloignement. Lazarus (1966, 1982, 1984) montre des évidences expérimentales de l'existence du processus cognitif préalable dans la production de l'émotion, contre des arguments de Zajonc. Frijda (1982, 1986) explique le rôle de l'évaluation cognitive dans les expressions émotionnelles, en s'appuyant sur la théorie d'évaluation aussi bien que sur la théorie évolutionnaire.

La discussion sur la relation entre l'émotion et la cognition est vivement animée dans les années 80 par le débat entre Zajonc et Lazarus, les représentants des arguments pour la primauté de l'émotion ('*primacy of affect*') et la primauté de la cognition ('*primacy of cognition*'). Zajonc (1980, 1982, 1984) propose l'indépendance partielle du système de l'émotion et du système cognitif, qui permet à l'émotion de se produire sans intervention de la cognition, tandis que Lazarus (1982, 1984) insiste sur l'indispensabilité du processus cognitif dans la production de l'émotion, donc la nature postcognitive de l'émotion.

Zajonc (1980, p151) propose que le jugement affectif est indépendant des opérations cognitives et qu'il les précède dans le temps. Cette thèse est démontrée par une série d'expériences, qui montrent le jugement affectif (la préférence ; 'aimer' ou 'pas aimer') dans l'absence totale de mémoire (la reconnaissance de 'nouveau' vs. 'ancien'). Il en conclut que l'émotion et la cognition sont sous le contrôle des systèmes séparés et partiellement indépendants. Ces systèmes s'influencent l'un l'autre mais ils sont des sources indépendantes du processus informationnel. On peut noter trois points essentiels dans sa théorie. (1) La réaction affective est primaire ; le sentiment se trouve le premier et l'évaluation cognitive vient après. (2) La réaction affective est fondamentale et inévitable ; cette réaction ne peut pas toujours être contrôlée volontairement, tel qu'on pourrait contrôler l'expression de l'émotion mais pas l'expérience de l'émotion elle-même. (3) La réaction affective est difficile à verbaliser et elle n'est pas nécessairement dépendante de la cognition ; la réaction est instantanée et automatique ; les expériences affectives ne s'accompagnent pas toujours de la représentation cognitive ou verbale.

Lazarus (1982 et 1984) réfute directement les arguments de Zajonc, en disant que l'activité cognitive est une condition indispensable et suffisante pour la production de l'émotion. En se situant dans la psychologie cognitive, il souligne que l'émotion résulte de l'évaluation cognitive de la relation entre le sujet et son environnement. Cette évaluation joue un rôle de transformateur des stimuli sensoriels à l'émotion engendrée, donc elle est toujours présente dans l'émotion et détermine l'intensité et la qualité de l'émotion produite. Il note aussi que cette évaluation n'est pas nécessairement consciente chez le sujet. Les résultats de ses expériences (1984, p125) montrent que l'évaluation cognitive est présente dans le jugement affectif même primitif (la préférence), ce qui contredit le résultat de l'expérience de Zajonc (1980, p160-165) à savoir que la préférence se produit en l'absence du processus cognitif.

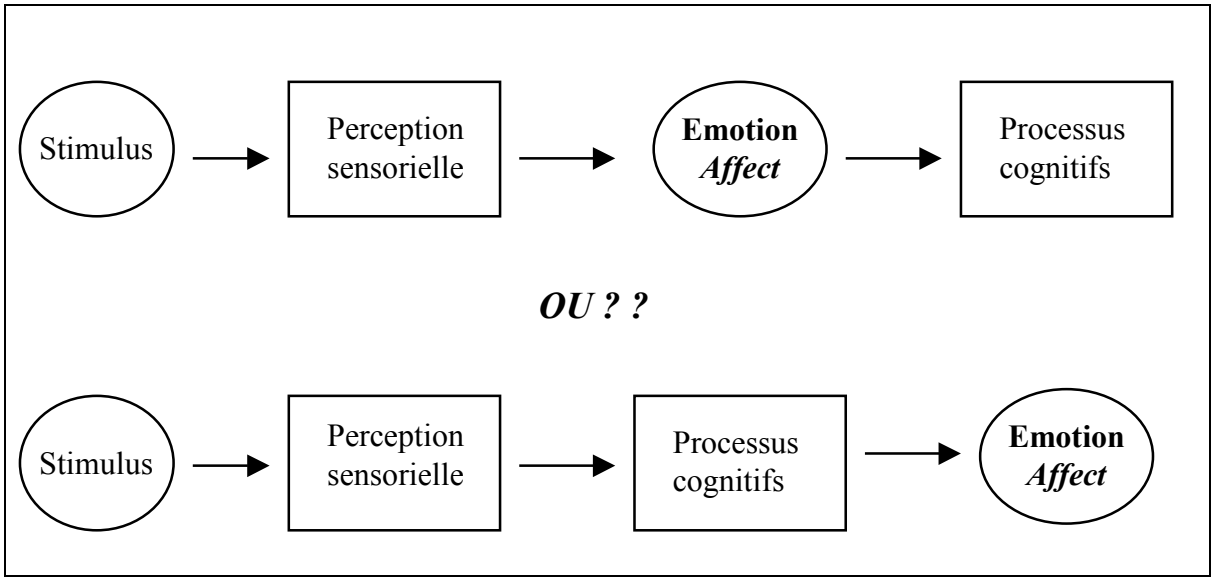


Figure 4. Schéma du débat entre Zajonc et Lazarus sur la production du jugement affectif.

Du point de vue évolutionnaire, l'émotion est une réponse adaptative d'un organisme à l'environnement physique ou social. Darwin (1872, p27-29) considère les symptômes des expressions émotionnelles comme des vestiges d'une activité qui avait un but déterminé dans un contexte primitif mais qui n'est plus utile dans le présent, dû au développement évolutionnaire. Son explication des expressions émotionnelles se base sur trois principes de survie dans l'évolution, qui sont universels pour les primates. Young (1943, p265) définit l'émotion comme une régression biologique, voire un retour à la réaction primitive, autrement dit, les types de réaction qui apparaissent pendant l'excitation

émotionnelle sont biologiquement anciens. Un modèle circulaire de Plutchik (1962, 1980b, 1997) identifie huit émotions primaires par leurs fonctions adaptatives, visant à établir un équilibre des forces opposées dans son environnement. Les approches évolutionnaires seront présentées plus en détail dan la partie II.3.3.

II.3.1.5. Neurophysiologie Moderne

Les théories psychophysiologiques expliquent l'émotion en terme d'activation, le degré de l'excitation du système nerveux. La théorie d'activation, définie au-dessus, soutient que l'émotion ressentie est d'autant plus intense que la physiologie est plus perturbée, mais ni la forme ni la source de cet éveil physiologique n'affectent la qualité de l'émotion (Danzter, 1988, p47). Les scientifiques du XX siècle effectuent diverses expériences pour trouver la corrélation entre les changements neurobiologiques et les émotions (voir Gainotti, 1989, et Scherer, 1986, pour le résumé). Ils associent les émotions négatives, comme la peur et la fureur, au système autonome sympathique, et les émotions positives au système parasympathique (Gardiner *et al*. 1937 ; Arnold, 1961 ; Plutchik, 1980a). L'expérience de Ekman *et al*. (1983) différencie les émotions positives et les émotions négatives par les profils de la variation de la fréquence cardiaque, de la conductance cutanée et de la température cutanée : le bonheur se caractérise par une faible variation de la fréquence cardiaque, la tristesse, par une grande variation cardiaque avec une faible variation de la température cutanée, et la colère, par de grandes variations de la fréquence cardiaque et de la température cutanée. Le développement technique dans les sciences modernes permet une recherche de plus en plus élaborée sur des corrélats physiologiques ou acoustiques des expressions émotionnelles. Leurs résultats sont partiellement appliqués dans les domaines informatiques comme synthèse et reconnaissance vocale par la machine (voir la partie II.5).

La localisation de l'émotion dans les parties du cerveau est un des thèmes des psychoneurologistes[2]. L'essence de la théorie de Cannon (1927) repose sur la fonction du thalamus dans la partie subcorticale en tant que source de l'expérience émotionnelle. Selon lui, l'émotion est la conscience de l'excitation thalamique, ce qui est contre l'argument de

[2] Dans cette partie, les études neurophysiologiques sont présentées relativement plus en détail que les autres études puisque des évidences proposées par ces études suggèrent des explications physiolgoqiues sur les différences des symptômes vocaux entre l'émotion positive et l'émotion négative, ce qui est le sujet de notre travail principal avec des données extraites de la parole spontanée émotionnelle.

James que l'émotion est la sensation des changements viscéraux. Selon Cannon (ibid., p115-124), un stimulus est perçu par un sujet au niveau du cortex cérébral et transmis au thalamus ; l'excitation thalamique rend le sujet sensible et conscient de l'émotion, et sa transmission aux niveaux viscéraux et périphériques produit une expérience émotionnelle chez le sujet, exprimée à travers ses organes physiologiques. Or, l'expérience de Bard et Rioch (1937) précise que c'est en fait l'hypothalamus qui contrôle l'expression émotionnelle et que le thalamus dorsal ne joue un rôle qu'au niveau de la sensation de l'émotion.

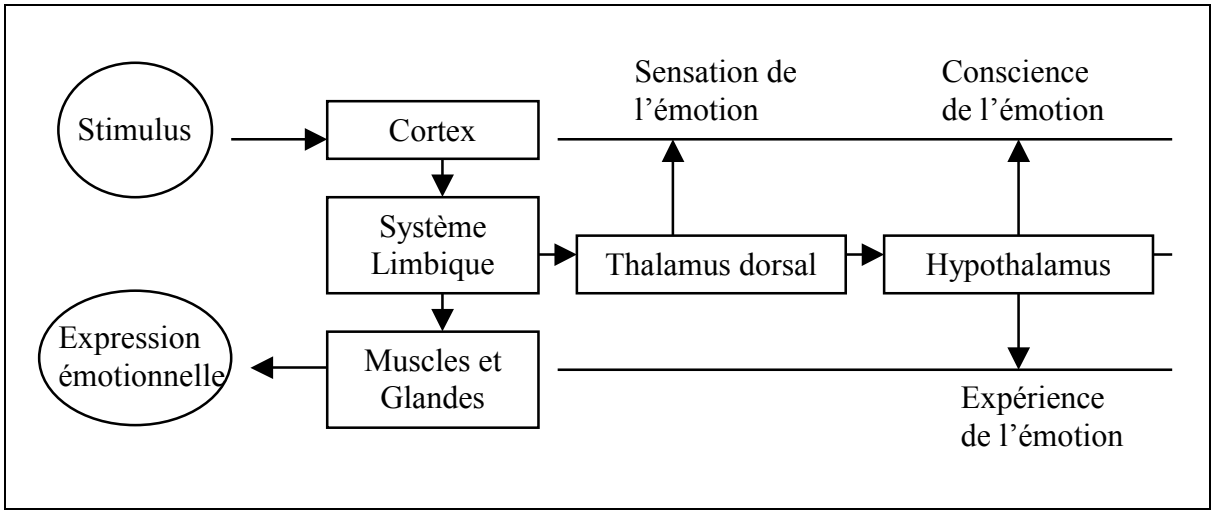


Figure 5. Modèle de la transmission de l'excitation du thalamus dans l'expérience émotionnelle, d'après Cannon et Bard.

Le modèle de circuit (Papez, 1937 ; MacLean, 1949) propose, au lieu de localiser les émotions dans une ou plusieurs structures définies neuroanatomiquement, que l'émotion est contrôlée par un circuit nerveux connectant entre elles plusieurs structures cérébrales, appelées le système limbique. La prédominance du système limbique dans la régulation de l'émotion, surtout le rôle de l'amygdale[3], est démontrée par de nombreux chercheurs (voir MacLean, 1970 ; Rolls, 1986 ; LeDoux, 1989). L'expérience de Delagado (1970) précise que l'émotion négative (colère) est déclenchée par la stimulation du noyau amygdalien droit tandis que l'émotion positive (bonheur), par celle d'autres points des noyaux amygdaliens. Ploog (1986) explique l'expression vocale de l'émotion, définie comme un produit évolutionnaire qui est fondamentalement inné, par des évidences

[3] L'amygdale est une structure limbique située dans la partie antérieure du lobe temporal. Elle reçoit de nombreuses entrées sensorielles d'origine visuelle, auditive, gustative, olfactive et même viscérale, déjà mises en forme par d'autres parties du cerveau (Dantzer, 1994, p76). L'amygdale est idéalement placée pour permettre à l'organisme d'associer les modalités sensorielles des stimuli à leur valeur affective (Rolls, 1986, p128-132).

biologiques dans les systèmes limbiques et les autres. Morris *et al*. (1996) insiste sur le rôle crucial de l'amygdale dans les expressions faciales de l'émotion, tel que la peur implique une grande activation neurale dans l'amygdale gauche, ce qui n'est pas le cas pour le bonheur. A propos de la régularisation des émotions, Bechtereva (1978) observe une augmentation du processus électrique pour les émotions négatives et une diminution de ce processus, pour les émotions positives, par l'intervention des neurotransmetteurs ou des hormones.

Contrairement à la supposition répandue que la cognition est contrôlée par l'hémisphère gauche et l'émotion, par l'hémisphère doit, Davidson (1986) propose qu'une certaine partie de l'hémisphère gauche soit spécialisée pour le processus des émotions positives tandis que la partie correspondante de l'hémisphère droit soit spécialisée pour le processus des émotions négatives. Son idée est inspirée par Sackeim *et al*. (1982), qui ont montré que la lésion de l'hémisphère gauche est associée aux pleurs et celle de l'hémisphère droit est lié aux rires. L'expérience de Davidson (ibid., p40) sur des malades dépressifs et non-dépressifs confirme cette différente latéralisation en fonction de la positivité de l'émotion, par l'observation d'une activation dans la partie frontale de l'hémisphère droite pour les dépressifs, mais d'une activation dans la partie correspondante de l'hémisphère gauche pour les non-dépressifs. Lors de sa deuxième expérience, les sujets normaux ont réagi de la même manière que les malades, vis-à-vis des stimuli de bonheur et de tristesse. Ces résultats n'ont pas toujours été confirmés dans d'autres expériences mais la plupart des études dans ce domaine suggèrent au moins que l'hémisphère droit et l'hémisphère gauche contribuent différemment aux processus des émotions positives et des émotions négatives (Dimond *et al*. 1976 ; Denenberg, 1981 ; Doty, 1989).

II.3.2. Théorie de l'affect (Tomkins)

La théorie de Tomkins ('*Affect theory*' ; 1962, 1963, 1980, 1984) est l'une des approches psychologiques qui s'intéressent surtout aux aspects primaires de l'émotion, connus sous le nom d'*affect*[4]. Tomkins (1962, p20) définit l'affect comme un mécanisme primaire motivationnel, qui consiste en un ensemble des réponses physiologiques à un stimulus interne ou externe, et qui produit des réactions sensorielles sous forme 'acceptable' ou 'non-acceptable'. Ce mécanisme cause l'activation neurale aux centres sous-corticaux, et les différents profils de cette activation caractérisent des affects différents. L'être humain, par sa nature, tente de maximiser l'affect positif et de minimiser l'affect négatif, afin d'obtenir ce qu'il considère comme un état idéal.

L'affect est la source biopsychologique des états *dérivés* ('drives' en anglais) comme le halètement, la faim, la soif et l'excitation sexuelle, et il fournit des calques pour la cognition, la décision, et l'action. Autrement dit, les dérivés doivent être assistés par l'affect pour qu'ils puissent être réalisés et tous les genres de l'activité cognitive sont initialement dirigés par l'affect positif ou négatif. En bref, l'affect est un amplificateur de tous les processus mentaux. Sans amplification de l'affect, rien ne se passe, et avec son amplification tout est rendu possible, ce qui suggère la primauté de l'affect[5]. L'*amplification*, selon Tomkins (1980, p147), signifie une augmentation ou une diminution de l'entrée dans la structure neurologique, ce qui correspond plus ou moins à la notion d'*activation* ou d'excitation ('arousal' en anglais), utilisée dans la psychologie générale.

En s'appuyant sur la théorie de polarité et la théorie d'activation, Tomkins (1980, 1982, 1984) identifie neuf affects primaires ; trois affects positifs *intérêt*, *joie*, *surprise*, et six affects négatifs *détresse*, *colère*, *peur*, *honte*, *mépris* et *dégoût*[6]. Les affects se diffèrent par leur différente sensibilité d'activation neurale et leurs diverses réponses symptomatiques en fonction des stimuli. Tomkins (1982, p244) considère la voix et le visage, plus précisément les muscles du visage, comme les sites principaux des réponses

[4] Dans notre travail, nous utilisons le terme *affect* sans traduction du mot anglais 'affect,' pour garder l'idée originale de l'auteur. L'affect dans les études psychologiques anglophones signifie l'émotion primaire qui se distingue de l'émotion dérivée avec l'évaluation cognitive. Voir Scherer *et al.* (1998) pour l'utilisation française du terme *affect*.

[5] Voir la discussion de Zajonc dans la partie II.3.1.4.

[6] Les deux derniers affects sont considérés comme des variantes d'un seul affect dans la théorie originale (Tomkins,1963).

affectives. La contraction des muscles du visage a, selon lui (1984, p189), une propriété motivationnelle et elle sert aussi une fonction communicative. Le tableau suivant présente les caractéristiques des réponses faciales[7] pour les affects primaires.

Affects primaires	**Réponses faciales**
Joie ou *Jouissance*	Sourire
Intérêt ou *Excitation*	Les sourcils sont baissés et le regard est fixé ou suit un objet.
Surprise ou *Etonnement*	Les sourcils sont relevés et les yeux clignotent.
Détresse ou *Angoisse*	Cri ou larmes
Colère ou *Fureur*	La mâchoire est serrée et le visage devient rouge.
Honte ou *Humiliation*	Les yeux et la tête sont baissées.
Mépris	La lèvre supérieure est relevée dans un sourire.
Dégoût	La lèvre inférieure est baissée et avancée.
Peur ou *Terreur*	Les yeux sont ronds et figés et le regard fixé ou s'éloignant de l'objet de la peur.

Tableau 2. Neuf affects primaires et leurs réponses faciales, selon Tomkins (1984).

En termes neurologiques, les affects se caractérisent par la *densité de l'activation neurale*, c'est-à-dire *augmentation*, *niveau* ou *diminution de la stimulation nerveuse*. Une augmentation de la stimulation peut activer des affects positifs ou des affects négatifs (*surprise*, *peur* et *intérêt*) tandis qu'une diminution de la stimulation active seulement les affects positifs (*rire*, *joie*). Un niveau tenu de stimulation active seulement les affects négatifs (*détresse*, *colère*). Prenons un exemple. Quand un enfant entend un bruit brusque, son activation neurale augmente et il s'étonne, devient intéressé ou a peur, selon la soudaineté de l'augmentation. Si le bruit est fort et long, l'activation neurale dépasse un certain niveau chez l'enfant et il pleure. Si ce bruit continue, l'activation maintient constamment un haut niveau de stimulation et il crie de colère. Par contre, si le bruit diminue soudainement, en faisant diminuer l'activation neurale d'une manière brusque, il va rire ou sourire de plaisir, selon la soudaineté de la diminution. La figure suivante montre les profils de l'activation neurale des affects positifs et des affects négatifs en fonction du temps.

[7] Il a noté la corrélation des réponses faciales et vocales mais n'a pas précisé les caractéristiques vocales des affects.

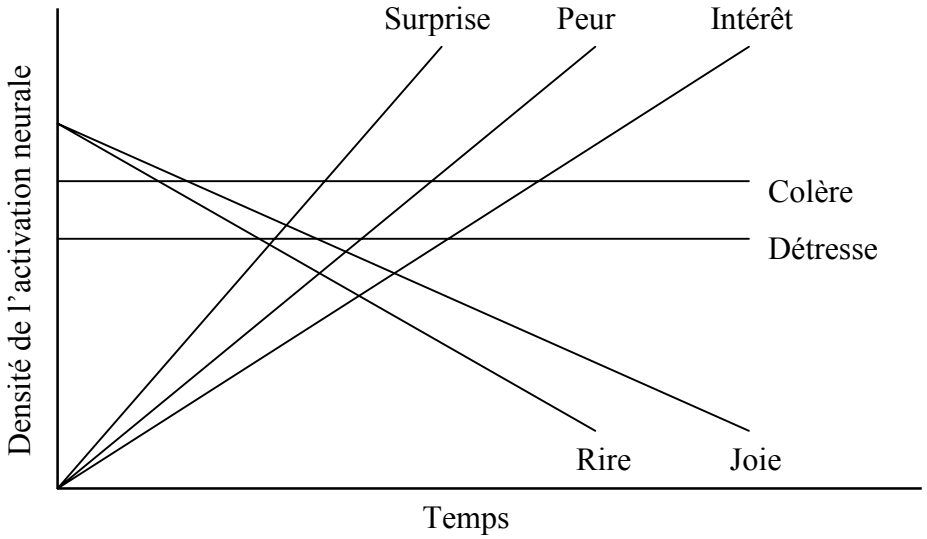


Figure 6. Modèle de l'activation des affects innés[8] selon Tomkins (1984, p169).

La théorie de l'affect insiste sur le fait que l'affect est activé seulement par le profil général de l'activation neurale en absence de l'évaluation cognitive. Ainsi, elle explique que la différente signification des stimuli produit des affects différents à travers le système neural, sans recours à l'évaluation cognitive. Une des évidences de cette absence de la cognition dans la production de l'affect est que le nouveau-né qui émet son cri à la naissance, ne peut pas 'évaluer' le nouvel environnement.

Parmi les affects primaires proposés par Tomkins, nous nous intéressons particulièrement à deux affects, la *détresse* et la *joie*, qui se retrouvent dans nos données du travail principal. Les réponses faciales de ces affects, tels les larmes et le sourire, seront utilisées comme un de nos critères de distinction de l'émotion positive et négative dans nos données à analyser. Nous présentons ici quelques caractéristiques de la détresse et de la joie, selon la description de Tomkins (1984, p173-180).

La *détresse* est un affect primaire et fondamental chez l'être humain, dû à l'ubiquité de la souffrance dans la société humaine. Cet affect se produit par un haut niveau de la densité de stimulation, qui dépasse, en général, un niveau optimal de l'activation neurale. La douleur, soudaine ou prolongée, produit typiquement un cri chez l'enfant. Un coup de pied brusque lui fait pousser un cri perçant alors que la douleur due à une maladie

[8] D'après Tomkins, la *honte*, le *mépris* et le *dégoût* sont des affects primaires mais pas innés. Ces trois affects ont, comme les autres, les propriétés d'amplification et de motivation, mais leur mécanisme de production est différent des autres.

prolongée le fait pleurer longtemps. La détresse peut être aussi produite par une activation neurale d'un niveau non-optimal continu, comme la douleur d'un niveau bas de fatigue, de faim et de froid, à laquelle l'enfant répond par un cri de détresse et l'adulte répond d'une manière muette. Tomkins (1984, p174) remarque la double face de la fonction biologique du *cri*. D'un côté, le *cri* communique à l'organisme lui-même et aux autres qu'il y a quelque chose de mauvais dans son environnement. De l'autre côté, il motive l'organisme et les autres à entreprendre une action afin de baisser le niveau de toxicité de la réponse de cri à un niveau tolérable à soi-même aussi bien qu'à ceux qui entendent le cri.

La *joie* se produit par la diminution soudaine de l'activation neurale, en contraste avec la *surprise* qui est produite par l'augmentation de cette stimulation. La signification du sourire de joie varie en fonction de l'état précédent. Après la douleur, la peur ou la détresse, le sourire est un *sourire de soulagement*. Dans le cas d'une diminution soudaine de la colère, c'est un *sourire de triomphe*. Quand il s'agit d'une diminution soudaine du plaisir comme après un bon repas ou après un orgasme, le sourire est un *sourire de plaisir*. Quand on rencontre une personne par surprise et la reconnaît avec joie, la diminution soudaine de son excitation active un *sourire de joie*, le *sourire de reconnaissance* ou le *sourire de familiarité*. Dans tous les cas, la raideur de la pente de diminution dans l'activation neurale est cruciale pour la production de *joie*. C'est-à-dire, la douleur ou l'excitation doit avoir un niveau d'activation suffisamment haut, afin que sa diminution puisse constituer une pente suffisamment raide pour stimuler la *joie*. Une diminution graduelle de l'activation neurale résulte en un état indifférent.

L'affect est essentiellement inné mais la stimulation et la réponse affective peuvent aussi être apprises à travers des expériences. Par exemple, l'activation de la réponse de sourire implique deux sortes de sourire. Un sourire peut être activé directement par la stimulation du programme de la réponse de sourire dans son cerveau, appelé le sourire 'inné' ou 'non-appris'. Izard ('*Differential Emotions Theory*,' 1984), comme Tomkins et Zajonc, met l'accent sur cet aspect inné de l'affect. Elle souligne que la composante de l'émotion est une fonction directe des perceptions sensorielles qui n'exigent pas de médiation cognitive (ibid., p19). La production d'une autre sorte de sourire peut être expliquée par la théorie de la mémoire. Une expérience consciente d'un sourire d'autrui active une recherche consciente du sourire de la personne elle-même dans le passé. Cette activation rétablit un programme mémorisé, ce qui active un sourire 'appris'.

II.3.3. Théories évolutionnaires (Darwin, Plutchik)

Les psychologues cognitifs considèrent l'émotion comme un ensemble d'états organisés, y compris évaluations cognitives, impulsions de l'action et réactions somatiques (voir Lazarus, 1982, 1984). Les psychologues évolutionnaires replacent l'émotion dans son contexte naturel d'origine et analysent la fonction et l'expression de l'émotion sur la continuité phylogénique et ontogénique chez les primates (voir Plutchik, 1980b). L'approche cognitive évolutionnaire à l'étude de l'émotion suppose l'existence préalable de l'évaluation cognitive dans la production de l'émotion, qui a évoluée au cours de l'histoire animale et humaine pour servir la fonction de survie.

On peut retracer l'analyse fonctionnelle de l'émotion jusqu'à l'œuvre de Darwin (1872), '*L'expression de l'émotion chez l'être humain et l'animal*'. D'après sa théorie de la sélection naturelle, tous les traits des espèces qui existent ont une valeur de survie. La conduite émotionnelle et la morphologie des êtres humains et des animaux peuvent être donc expliquées par leur fonction de survie. Darwin restitue le contexte naturel des expressions émotionnelles. C'est le monde qui entourait l'homme primitif et ses ancêtres. La nature de l'environnement crée certains besoins fonctionnels pour tous les organismes afin qu'ils survivent. Ils doivent chercher la nourriture, avoir un territoire et distinguer un partenaire potentiel d'un ennemi potentiel. Selon Darwin, l'émotion est une activité dirigée vers un but précis qui est utile dans certaines circonstances et qui est automatiquement déclenchée par d'autres situations similaires.

Du point de vue phylogénétique, la conduite spécifique par laquelle des fonctions primaires sont accomplies varie au cours du temps et à l'intérieur d'un groupe d'espèce mais le prototype des conduites pour les fonctions de base reste le même. Dans le même ordre d'idée, Darwin (1872, p28) considère les expressions émotionnelles comme des résidus - 'des membres dispersés' - des actions ancestrales et formule trois lois qui sous-entendent les expressions des émotions comme suit.

1) **Loi des habitudes associées serviables** (*'The principle of serviceable associated Habits'*) : Les actes utiles s'associent à certains états d'âme et seront reproduits même s'ils ne sont pas nécessaires, dans toutes les situations analogues.

2) **Loi de l'antithèse** (*'The principle of Antithesis'*) : Quand un état d'âme qui est contraire à celui d'habitude est déclenché, il y a une tendance forte et involontaire que son mouvement aille dans un sens opposé à la nature. Ce genre de mouvement est souvent hautement expressif.

3) **Loi de l'action directe du système nerveux** (*'The principle of the direct action of the nervous system'*) : Quand un nerf sensoriel est fortement excité, la force est transmise dans un certain sens défini, en fonction de la connexion des cellules nerveuses et partiellement en fonction de l'habitude. Ce sont des effets produits de cette excitation qu'on perçoit comme expressifs. Ces actions sont dues au système nerveux, indépendamment de la volonté et de l'habitude.

Ainsi, le point de vue évolutionnaire a jeté une nouvelle lumière sur l'étude de l'émotion, fournissant diverses explications sur les fonctions de l'émotion dans la vie de chaque organisme. Dans cette approche, l'expression de l'émotion est censée montrer les traces de sa fonction originelle de survie, c'est-à-dire les traits fonctionnels de l'émotion se retrouvent dans les expressions faciales et vocales.

McDougall (1908), un des fondateurs de la psychologie sociale, explique l'émotion en terme d'instinct qui a des valeurs de survie chez un organisme. Selon lui, les émotions primaires se trouvent au centre des instincts et chaque instinct est constitué de trois composants, les composants affectifs, cognitifs et motivationnels. Le composant affectif lie les deux derniers composants et il montre un aspect plus primaire que les autres ; il n'est pas modifiable, tandis que les composants cognitifs de la perception et de la motivation, peuvent être modifiés par l'expérience et l'apprentissage.

La théorie psychoévolutionnaire de Plutchik (1962, 1980a, 1980b, 1990, 1997) considère l'émotion comme un ensemble de réactions adaptatives fonctionnelles pour établir un équilibre social dans l'environnement d'un organisme. Plutchik (1980b, p4) définit l'émotion comme une entité hypothétique et déductive, qui ne peut être analysée que par les traits communs des évidences constatées. Cette définition vient de sa critique des approches subjectives dont la définition de l'émotion est basée sur le rapport verbal des sujets. Il relève les problématiques de cette méthode comme la falsification volontaire ou involontaire de la description, l'influence de la capacité des sujets à se rappeler et à

exprimer ce qu'ils ont vécu et le biais introduit par l'expérimentateur par sa façon de poser des questions sur la description de l'émotion par les sujets. Donc, au lieu de traiter l'émotion en tant qu'un sentiment subjectif, il se base sur un modèle hypothétique de l'émotion, appelé le modèle circulaire ('*Circular Model*'), pour ses analyses de la production et l'expression de l'émotion.

Son modèle, comme d'autres psychologues cognitifs, présuppose l'existence d'une évaluation cognitive dans la production de l'émotion, qui estime si un stimulus est bon ou mauvais (favorable ou défavorable) pour l'organisme. Ce processus cognitif n'est pas toujours conscient ou verbalisable mais il existe toujours pour évaluer le stimulus et prédire la situation, rendant la conduite émotionnelle de l'organisme mieux adapté à son environnement. Autrement dit, l'émotion se produit par suite d'une série de processus cognitifs qui évoluent à la disposition de l'émotion. La figure suivante montre comment l'émotion est développée dans une série d'événements dans un organisme.

Evénement de stimuli	→ **Cognition**	→ **Sentiment**	→ **Conduite**	→ **Effet**
Menace par un ennemi	→ Danger	→ Peur	→ Fuite	→ Protection
Perte des parents	→ Isolation	→ Tristesse	→ Pleur	→ Aide

Figure 7. La séquence des événements dans l'émotion, selon Plutchik (1980b, p11).

Le modèle circulaire conceptualise les émotions dans un champ sémantique selon le degré de la similarité (la proximité) et de l'opposition (la polarité). Quatre paires d'émotions primaires sont identifiées selon leurs fonctions d'origine[9] ; la colère/ la peur (l'attaque/ la fuite), la joie/ la tristesse (la possession/ la perte), l'acceptation/ le dégoût (la prise/ le rejet) et la surprise/ l'attente (l'absence/ la présence de la prédiction). Le mélange de ces émotions primaires produit des émotions dérivées, ce qui est illustré dans la figure suivante. Les traits de la personnalité et les traits psychopathologiques sont aussi considérés comme les dérivatifs des émotions primaires et ils peuvent être décrits dans un champs similaire à celui de l'émotion.

[9] Les orignines fonctionnelles sont notées dans les parenthèses.

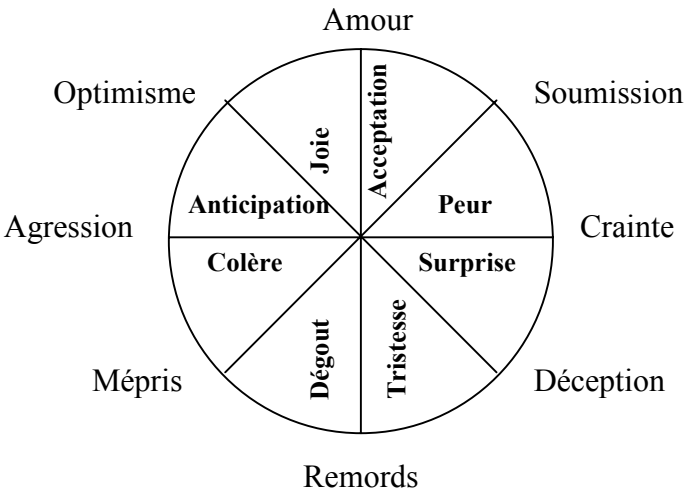


Figure 8. Modèle circulaire de huit émotions primaires (en caractères gras) et leur mélange, selon Plutchik (1984, p205).

Plutchik (1980b, p27-30) explique les émotions primaires en relation avec les problèmes universaux de l'adaptation chez les primates. Premièrement, la colère et la fuite sont liées à la hiérarchie de dominance dans la société humaine ou animale. Deuxièmement, la temporalité de l'être humain ou de l'animal cause les émotions comme la joie et la tristesse. Troisièmement, l'acceptation et le dégoût sont associées aux problèmes d'identité ou d'appartenance à un groupe. Dernièrement, le problème du territoire dans lequel on peut être assuré pour la nourriture et le repos produit les émotions comme la surprise et l'attente.

Concernant l'expression vocale de l'émotion, Andrew (1972) attribue l'origine de la phonation aux réponses physiologiques adaptatives de l'organe, telle que la fermeture ou le rétrécissement de la glotte pour sa protection, alors que Lieberman (1975) insiste que le larynx des primates s'est adapté d'une manière sélective pour la phonation, afin de mieux servir la communication, au détriment de la protection optimale des poumons ou de l'efficacité respiratoire. Scherer (1979, p496) considère la vocalisation affective comme un résultat des pressions sélectives des signaux acoustiques, qui sont adaptés à la communication entre espèces. Scherer & Kappas (1988) soulignent la continuité évolutionnaire dans la phonation expressive émotionnelle chez l'être humain et l'animal. Le modèle de Scherer est présenté plus en détail dans la partie II.4.1.

II.3.4. Différents termes émotionnels

La terminologie de l'émotion est toujours un sujet de controverse parmi les chercheurs. Comme nous avons vu dans les parties précédentes, la définition de l'émotion diffère largement d'une théorie à l'autre, ce qui est étroitement lié aux problématiques de la nature de l'émotion et de la relation entre l'émotion et la cognition. Ici, nous présentons quelques essais de distinction des différents termes affectifs (ex. émotion, passion, sentiment, affect, humeur et attitude), selon leurs propriétés et leurs fonctions chez l'être humain.

Wetzel (1989) distingue l'*émotion*, la *passion* et le *sentiment*, à l'égard de leur production et leurs caractéristiques symptomatiques. L'*émotion* est une réaction affective intense à un événement extérieur et elle a un caractère fugitif dans le temps. La *passion* naît d'un déclic interne et elle est entretenue par des objets externes au cours du temps. Le *sentiment* est une passion vécue sur un mode mineur : le sujet garde la tête froide au lieu de se laisser gagner par le vertige, la frénésie et la complaisance qui caractérisent la vraie passion.

Dantzer (1994, p11) insiste sur le fait que l'émotion est plus qu'une simple réaction pour constituer une véritable attitude, une évaluation de ses propres états en rapport avec autrui. Selon cette conception, trois niveaux de l'émotion sont identifiés : 1) les *émotions fondamentales* qui sont simplement des réactions à des événements extérieurs réels ou imaginaires (par exemple, le dégoût et la peur) ; 2) les *émotions dérivées* qui sont fondées sur l'émotion engendrée par l'image que l'on a de la conscience de l'autre (par exemple, le mépris est un dégoût pour les prétentions d'une autre conscience et la méfiance est une peur du mystère posé par celle-ci) ; 3) les *émotions tierces* qui naissent de la conscience de soi face au regard de l'autre (par exemple, la honte est un mépris de soi et la timidité est une peur de sa propre valeur telle qu'elle est perçue par les autres).

Cette classification des émotions rappelle la catégorisation des *émotions I, II et III* dans la théorie de Buck ('*Prime Theory*,' 1985). Cette théorie considère les émotions comme des dispositions potentielles de l'organisme, qui sont organisées de manière

hiérarchique dans des systèmes primaires motivationnels et affectifs (‘*primes*’). L’*émotion I* est le sentiment le plus fondamental qui est associé à la fonction de l’adaptation corporelle et à la fonction du maintien de l’équilibre d’homéostasie pour la nourriture, l’eau, l’oxygène et la température. L’*émotion II* est un sentiment associé aux conduites expressives dont la fonction est de communiquer avec les autres dans un contexte social. L’*émotion III* est une réflexion interne cognitif du sujet, associée à son expérience subjective émotionnelle.

Batson *et al*. (1992) distinguent l’*affect*, l’*humeur* et l’*émotion*, (‘*Affect*,’ ‘*Mood*’ et ‘*Emotion*’ en anglais) par leur fonction motivationnelle. L’*affect* est un sentiment de base qui est indispensable dans la phase initiale de la motivation. L’*humeur* et l’*émotion* viennent après l’*affect* avec un processus cognitif pour l’évaluation de la relation entre le sujet et l’événement ou le stimulus. L’évaluation cognitive est plus élaborée dans l’*émotion* que dans l’*humeur* parce que la première cause la réaction d’un sujet pour changer sa relation actuelle avec un but spécifique tandis que la dernière reste juste au niveau d’attente du plaisir dans le futur. Ces trois états affectifs sont étroitement liés et ils peuvent être provoqués par un même événement, simultanément ou à la suite. Batson *et al*. (ibid., p298-302) fournissent une explication détaillée de ces trois états affectifs, l’*affect*, l’*humeur* et l’*émotion*, dans le cadre de la théorie motivationnelle comme suit.

L’*affect* est le terme le plus général et le plus fondamental, phylogéniquement et ontogéniquement, parmi les trois termes. Il est présent, sous la forme primitive, dans le cri de l’enfant ou dans le jappement du chien. L’*affect* a des propriétés de valence (‘positive’ ou ‘négative’) et d’intensité (de ‘faible’ à ‘fort’) et ces propriétés semblent être physiologiquement exprimées dans le cerveau moyen (peut-être l’hypothalamus pour la valence et la formation réticulaire pour l’intensité ; voir Buck, 1985, 392-396) mais ce n’est pas encore clairement défini (voir Leventhal, 1980). L’*affect* informe l’organisme sur la valeur d’un état par rapport à celle d’autres, ce qui provoque la préférence d’aimer ou de ne pas aimer (Zajonc 1980, p154). Le changement d’un état avec moins de valeur à un autre avec plus de valeur produit un *affect positif* tandis que le changement inverse produit un *affect négatif*. L’intensité de l’*affect* reflète l’ampleur de la préférence de valeur.

L’*humeur* est un état affectif qui présuppose une attente d’événement dans le futur. Il se produit après le jugement affectif général, l’*affect positif* ou l’*affect négatif* et fait un

ajustement psychologique de ce jugement par rapport à l'événement prévu. Autrement dit, l'*humeur* informe l'organisme sur une expérience de plaisir ou une expérience de souffrance dans l'avenir. Une *bonne humeur* signifie que l'événement ou le stimulus à venir est réjouissant alors qu'une *mauvaise humeur* signifie que cet événement est nuisible. Sur le plan temporel, l'*humeur* peut durer quelques jours, même quelques semaines. Une transition d'attente de plaisir ou de souffrance dans le futur cause un changement de l'*humeur* dans le présent.

L'*émotion*, selon Batson, est un état affectif qui vise un but spécifique et organise une séquence d'actions orientées vers le but. Elle a les propriétés de valence et d'intensité comme l'*affect* et l'*humeur*, mais elle implique un processus cognitif relativement plus compliqué que les deux derniers. Le but de l'*émotion* s'établit à partir d'une évaluation des valeurs des événements ou des stimuli perçus chez un sujet. Cette estimation des valeurs se base sur la préférence relative et le but se produit quand la valeur préférée est menacée ou cette valeur peut être obtenue mais elle n'est pas encore acquise.

Léon (1993, p113) souligne la dualité de l'émotion, qui est à la fois une réaction physiologique et un comportement social et distingue l'*émotion brute* et l'*émotion socialisée*. La première est considérée comme une réponse de l'organisme à une situation donnée et la dernière, comme une activité dirigée dans un environnement social. Il explique que ces deux catégories psychologiques correspondent à ce que les linguistes désignent généralement par *émotion* et *attitude*. A l'égard du contrôle du sujet dans son expression émotionnelle, l'*attitude* et l'*émotion* peuvent être appelées l'*émotion contrôlée* et l'*émotion non-contrôlée*, et entrent respectivement dans le domaine paralinguistique et le domaine extralinguistique (voir Laver, 1994, p21-23). Selon Fónagy (1983a), qui s'intéresse surtout aux aspects expressifs et communicatifs de l'émotion, l'expression de l'émotion fait partie du *style*, défini comme une façon de parler ou une manière de s'exprimer, avec d'autres expressions emphatiques ou de personnalité.

Dans le même ordre d'idée, Péter (1997) distingue trois notions, *affectivité*, *expressivité* et *valeur stylistique*, dans le fonctionnement de la langue. Selon lui, l'*affectivité* est la manifestation de l'état affectif actuel du sujet parlant dans le discours. Cet état affectif est exprimé de manière automatique (avec un minimum contrôle cognitif) dans l'intonation et le lexique du mot, et peut être interprété par les formules "et c'est bon"

/ "et c'est mauvais". Le sujet aussi peut utiliser des moyens langagiers d'une façon inhabituelle et inattendue pour attirer l'attention de l'interlocuteur dans le discours, ce que Péter appelle *expressivité*. Cette utilisation particulière des mots ou des phrases augmente l'efficacité communicative du discours. Le *style* étant une variation d'usage de la langue, la *valeur stylistique* peut être définie comme le marquage des éléments de la langue indiquant la situation communicative, y compris l'activité sociale aussi bien que les sortes de textes. Cette valeur se base sur le rapport causal naturel entre le signifiant et le signifié. En résumé, l'affectivité porte sur *ce qui* est dit, l'expressivité et la valeur stylistique sur la *façon* de dire.

Dans notre présentation, le terme *émotion* est utilisé au sens large, dénotant L'ENTITE AMALGAMEE DES COMPOSANTS AFFECTIFS ET COGNITIFS, EXPRIMEE A TRAVERS LA VOIX OU LE VISAGE AVEC OU SANS CONTROLE DU LOCUTEUR. Ce terme est utilisé comme l'équivalent de l'*expression émotionnelle vocale* dans la plupart des cas de cette thèse, sauf quelques exceptions précisées dans le texte. La terminologie globale de l'émotion s'explique par le fait qu'il est difficile et même dangereux d'employer différentes terminologies catégoriques pour les différentes expressions émotionnelles, vu la nature de nos données acquises à partir de la situation naturelle. Au lieu d'établir la définition conceptuelle pour les différents genres d'émotion, nous les caractérisons par les corrélats acoustiques et perceptifs à travers une série d'expériences, ce qui est présenté dans les chapitres IV. Par exemple, l'effet des pleurs sur la vocalisation est expliqué par les corrélats acoustiques et perceptifs de l'émotion alors que la contribution du contour de Fo à la perception des émotions positives et négatives est expliquée par la configuration conventionnelle des traits prosodiques dans le système langagier. Les modèles sur lesquels nos expériences sont basées sont présentés dans les parties suivantes, II.4 et II.5.

II.4. Modélisations de l'expression émotionnelle

Beaucoup de chercheurs dans l'étude de l'émotion s'intéressent à trouver des invariants dans l'expression émotionnelle et le mécanisme de la communication de l'émotion à travers le visage, la voix ou les mouvements corporels. Dans cette partie II.4, nous présentons quatre modèles qui visent à expliquer l'expression vocale de l'émotion en termes des indices physiologiques et acoustiques dans un système de la communication.

Premièrement, le modèle de Scherer, le *modèle du processus componentiel*, considère l'émotion comme une série de changements adaptatifs de l'organisme au cours de l'évaluation des stimuli et il analyse les facteurs dans cette évaluation, qui contribuent aux résultats de l'expression émotionnelle. Deuxièmement, le modèle de la covariation et le modèle de la configuration reflètent les deux approches majeures dans l'étude de l'émotion vocale ; l'un s'intéressant à une covariation directe entre le degré d'émotion et les changements acoustiques, et l'autre insistant sur l'interaction entre les indices verbaux et non-verbaux. Troisièmement, le modèle phonostylistique de Troubetzkoy et de Léon distingue les aspects volontaires et involontaires dans la parole naturelle et attribue une fonction identificatrice à l'expression émotionnelle involontaire. Enfin, le modèle de Fónagy explique le double codage dans la communication parlée, un codage linguistique et un codage paralinguistique. La relation entre le message linguistique et sa réalisation expressive est bien conçue dans ce modèle.

II.4.1. Modèle du processus componentiel (Scherer)

Le modèle du processus componentiel, proposé par Scherer ('*Component Process Model*'; 1984, 1986) considère l'émotion comme un processus constitué de cinq composants principaux : a) le composant cognitif, b) le composant physiologique, c) le composant motivationnel, d) le composant de l'expression motrice, et e) le composant du sentiment subjectif. Chaque composant accomplit sa propre fonction dans le système affectif, telles que l'évaluation des stimuli, la régulation neurale, la préparation des actions, la communication avec les autres et la réflexion avec soi-même.

COMPOSANTS	FONCTIONS
Processus cognitif du stimulus	Evaluation des stimuli
Processus neurophysiologique	Système de régulation
Motivation et tendances de conduite	Préparation des actions
Expression motrice	Communication de l'intention
Etat subjectif sentimental	Réflexion et contrôle

Tableau 3. Cinq composants et leurs fonctions dans le modèle de Scherer (1984, p297).

Dans ce modèle, Scherer (1986, p147) partage l'idée avec Arnold (1960), Lazarus (1966) et Plutchik (1980a), que l'expérience émotionnelle résulte des évaluations cognitives du stimulus, qui informent l'organisme de la signification du stimulus vis-à-vis des problèmes de survie ou de bien-être. Pour lui, l'émotion est un *syndrome* des divers composants dans une séquence d'évaluations. Il souligne que cette émotion est un ensemble des réponses de l'organisme face à l'événement évalué comme significatif (c'est-à-dire, excitation physiologique *et* expression motrice *et* sentiment subjectif), mais pas une réponse simple de l'un des sous-systèmes de l'organisme (c'est-à-dire, excitation physiologique *ou* expression motrice *ou* sentiment subjectif).

Ainsi, il définit l'émotion comme une série des changements adaptatifs dans les systèmes de l'organisme, par suite des événements antécédents évalués comme significatifs dans le but de l'organisme. Il distingue l'*émotion en tant que processus* et l'*émotion discrète*. La première est une succession des états émotionnels, qui changent rapidement dans le temps, tandis que la dernière est un sous-ensemble dans le changement des états, qui est perçu par des sujets comme une unité identifiable et qui peut être étiqueté avec des termes émotionnels dans les langues naturelles.

L'émotion est l'interface entre l'organisme et son environnement, et le processus d'émotion est constitué de trois aspects majeurs : Premièrement, le processus reflète l'évaluation de la pertinence ou de la signification des stimuli par l'organisme, concernant ses besoins, son plan, et sa préférence. Deuxièmement, il prépare des actions adaptées, aux niveaux physiologique et psychologique, pour que l'organisme soit capable de prendre en compte des stimuli. Troisièmement, le processus d'émotion fait communiquer l'organisme avec les autres dans son environnement social (Scherer, 1984, 295).

Le modèle du processus componentiel se base sur une théorie de la séquence de différenciation émotionnelle, proposée par Scherer (‘*Sequence theory of emotional differentiation*’; Scherer, 1984, 1986). Cette théorie postule que chaque système de la procédure d’information dans l’organisme examine l’entrée du stimulus interne ou externe et effectue une série d’évaluations du stimulus (‘*Stimulus Evaluation Checks*,’ *SECs*), en utilisant les critères qui sont définis en termes de fonctions. Il y a cinq évaluations (*SECs*) dans la théorie et elles sont censées s’effectuer toujours dans le même ordre de succession. Les différentes émotions discrètes sont produites selon les différents résultats des évaluations. Ainsi, la théorie identifie les facteurs déterminants dans la différenciation des émotions discrètes et modélise les changements des états dans les systèmes de l’organisme pour des émotions produites. Voici les cinq évaluations (*SECs*) du processus d’émotion.

1) **Evaluation de la nouveauté** (‘*novelty check*’) : évaluer s’il y a un changement dans les stimuli internes ou externes, en particulier s’il y a un nouvel événement.

2) **Evaluation du plaisir intrinsèque** (‘*intrinsic pleasantness check*’) : évaluer si l’événement est agréable ou désagréable, déterminant une tendance d’approche ou de recul de l’organisme.

3) **Evaluation de la signification du but et du besoin** (‘*Goal/need significance check*’) : évaluer si cet événement est pertinent pour satisfaire un but ou un besoin de l’organisme (‘*relevance subcheck*’), si le résultat est cohérent ou incohérent avec l’état anticipé (‘*expectation subcheck*’), s’il est favorable ou défavorable d’atteindre le but ou de satisfaire le besoin (‘*conductiveness check*’), et à quel degré d’urgence la réponse de conduite est exigée (‘*urgency check*’).

4) **Evaluation de la puissance face à l’événement** (‘*Coping potential check*’) : évaluer la cause de l’événement et le degré de contrôle de l’organisme sur l’événement (‘*control subcheck*’), la puissance relative de l’organisme pour changer ou éviter le résultat (‘*power subcheck*’), et le potentiel d’ajustement des résultats par l’organisme (‘*adjustment subcheck*’).

5) **Evaluation de la compatibilité avec la norme et le soi/norme** (‘*Norm/self compatibility check*’) : évaluer si l’action de l’organisme est conforme aux normes sociales et aux conventions culturelles (‘*external standards subcheck*’), et si cette

action est cohérente avec les normes internes de soi-même (‘*internal standards subcheck*’).

Selon ce modèle, les résultats de chaque évaluation ont des effets spécifiques sur le système nerveux somatique (‘*Somatic Nervous System*’) et le système nerveux autonome (‘*Autonomic Nervous System*’), en produisant des changements distinctifs dans les activités motrices. La plupart des émotions sont accompagnées de l’augmentation des activités motrices, comme l’augmentation de la tension musculaire, ce qui explique l’augmentation de la fréquence fondamentale dans la voix émotionnelle. Les caractéristiques de l’expression vocale sont considérées comme les résultats des effets des évaluations (*SECs*) dans les systèmes neuromusculaires.

Scherer (1986, p148-158) prédit différents types de voix dans l’expression émotionnelle, en fonction des effets physiologiques par les évaluations psychologiques (*SECs*). Sa prédiction repose sur trois dimensions principales, *valence*, *activation* et *puissance*, en se basant sur la théorie dimensionnelle de l’émotion. Concernant la *valence*, l’évaluation du stimulus agréable ou désagréable produit un résultat positif ou négatif et le résultat positif se manifeste en *voix étendue* tandis que le résultat négatif, en *voix étroite*. Concernant l’*activation*, l’ensemble des résultats des évaluations de pertinence, d’attente, d’urgence et de contrôle, influence la qualité de *voix, tendue* ou de *voix relâchée*. Par exemple, plus le but est pertinent, plus l’attente est contradictoire, plus l’action est urgente et plus il est possible de contrôler la situation, plus d’activations neurales s’établissent dans l’organisme et sa voix devient plus tendue. Concernant la *puissance*, les résultats de l’évaluation de puissance affectent l’ampleur de la voix. C’est-à-dire, quand l’organisme perçoit une grande puissance face à l’événement, sa *voix* devient *pleine*, tandis que quand il perçoit une absence de la puissance ou peu de puissance, sa *voix* devient *fine*.

La figure suivante montre comment les différentes qualités de voix sont dérivées à travers les cinq étapes d’évaluation (*SECs*), liées aux changements physiologiques et acoustiques, selon la prédiction du modèle du processus componentiel.

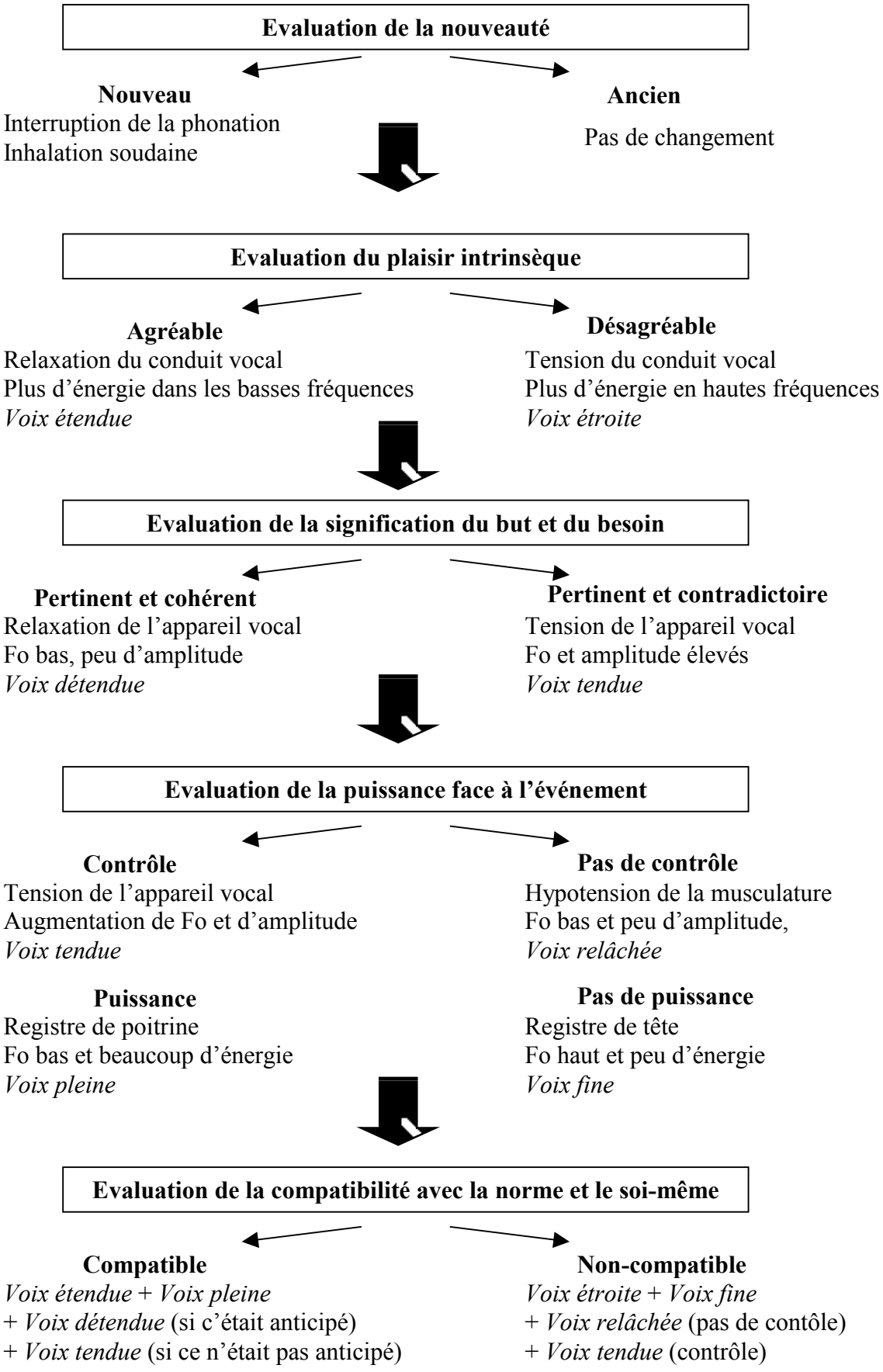


Figure 9. Prédictions des changements vocaux dans les différentes étapes d'évaluation (*SECs*) d'après le modèle de Scherer (1986).

II.4.2. Modèle de la covariation et modèle de la configuration

Dans cette partie, nous présentons deux modèles psycholinguistiques concernant l'expression de l'émotionnelle dans la voix ; le *modèle de la covariation* et le *modèle de la configuration*. La plupart des signaux vocaux sont plurifonctionnels, ce qui implique des déterminants multiples dans la vocalisation. Les différents types de déterminants se manifestent dans la voix émotionnelle par différents effets sur le codage, c'est-à-dire la relation entre le référent sous-jacent et les caractéristiques du signal. Scherer *et al.* (1980) distingue les effets des déterminants internes et externes dans l'expression vocale émotionnelle, appelés l'effet '*push*' et l'effet '*pull*'. Selon lui, les deux modèles se caractérisent par leur intérêt principal sur l'effet '*push*' ou l'effet '*pull*'.

Le *modèle de la covariation* s'intéresse surtout à l'effet '*push*' qui se manifeste dans des changements physiologiques de la voix émotionnelle. Cet effet concerne les processus physiologiques, telle que la tension musculaire, qui 'poussent' la vocalisation dans une certaine direction. Les changements dans les sous-systèmes physiologiques de l'organisme produisent un effet direct sur les paramètres vocaux et les mesures acoustiques sont censées refléter cet effet directement. Par exemple, le stress cognitif ou émotionnel produit une augmentation de la tension des muscles, ce qui fait augmenter la fréquence fondamentale de la voix. Dans ce cas, on s'attend à une relation covariationnelle directe entre la quantité d'augmentation de la tension musculaire, mesurée par électromyographie, et l'augmentation de la fréquence fondamentale. Ainsi, ce modèle insiste sur l'indépendance des indices prosodiques, non-verbaux (comme la fréquence fondamentale, l'intensité et la pause), et le contenu verbal, en cherchant à trouver la covariation entre les paramètres acoustiques et le degré d'émotion du sujet sur une échelle continue, linéaire ou non-linéaire.

L'autre type d'approche, le *modèle de la configuration*, s'intéresse plutôt à l'effet '*pull*' dans l'expression de l'émotion et de l'attitude. Cet effet a trait à des facteurs externes, telles que les attentes de l'auditeur, qui tirent la vocalisation vers un modèle acoustique particulier. Les facteurs '*pull*,' bien que médiatisés par des systèmes internes, sont basés sur l'extérieur et repose sur un aspect 'cognitif' ou 'linguistique' de la

vocalisation. Ils produisent certaines caractéristiques acoustiques de la voix, qui sont définies et valorisées par des membres linguistiques dans une société donnée. Donc, ce modèle insiste sur la dépendance des indices verbaux et non-verbaux et leur perception catégorielle. Les différentes expressions de l'émotion ou de l'attitude se font par une configuration différente des variables dans une structure linguistique catégorielle. Le locuteur utilise une combinaison particulière d'intonations, d'accents, de mots et de structures syntaxiques, pour communiquer son état émotionnel ou la modalité de sa proposition à l'auditeur. Par exemple, le contour final montant forme des questions qui provoquent des réponses de type oui/non (questions fermées) et le contour final descendant forme des questions qui provoquent des informations ou la confirmation de l'interlocuteur (questions ouvertes) ; ils peuvent aussi signaler l'état émotionnel du locuteur par leurs différentes façons de combinaison.

En termes de Bühler (1934), l'effet '*push*' reflète l'aspect de *symptôme* de l'expression émotionnelle et l'effet '*pull*,' les aspects de *symbole* et d'*appel*. Autrement dit, le premier reflète des modifications physiologiques, qui sont totalement involontaires et liées directement à l'état interne du sujet, tandis que le dernier montre l'intention du locuteur dont la communication est culturellement définie et acquise à travers des expériences. C'est ainsi que l'effet '*push*' et l'effet '*pull*' servent de fonctions distinctes dans la parole émotionnelle, en déterminant la production vocale du locuteur, d'une part, et son style de la communication dans un milieu social, d'autre part (voir Scherer *et al*.,1998, p251). Il faut aussi noter que ces fonctions sont mutuellement interdépendantes par nature et qu'il peut y avoir un antagonisme entre l'effet '*push*' et l'effet '*pull*'. Par exemple, une excitation physiologique accrue 'pousse' la fréquence fondamentale à un niveau plus élevé, alors que les tentatives conscientes de contrôle par le locuteur 'tirent' la fréquence fondamentale vers le bas, ce qui mène à la production de messages mixtes, voire contradictoires.

Afin d'évaluer la validité des deux modèles, le *modèle de la covariation* et le *modèle de la configuration* et d'identifier la contribution respective des effets '*push*' et des effets '*pull*' dans l'expression vocale émotionnelle, Scherer *et al.* (1984) ont effectué une série d'expériences avec des stimuli préparés par trois méthodes, filtrage passe-bas,

découpage aléatoire[10] et renversement de phrase. Leur corpus consistait en des extraits de conversations entre des fonctionnaires allemands et deux acteurs qui jouèrent le rôle des clients, et les énoncés du corpus étaient des questions sous forme de reproche ou des questions d'information purement factuelle. Dans le test perceptif, les auditeurs ont jugé quelle émotion reflétaient les énoncés enregistrés, les phrases écrites et les phrases filtrées, en choisissant des adjectifs affectifs.

Leurs résultats confirment la validité des deux modèles par la différente contribution des facteurs '*push*' et '*pull*' aux différents aspects de la parole. D'une part, les auditeurs ont pu reconnaître la signification affective des énoncés même lorsque le contenu des énoncés était rendu inintelligible. Ce résultat montre que la communication affective se fait de façon directe et indépendante du contenu verbal, comme proposé par le *modèle de la covariation*. Selon ce résultat, la qualité de voix joue un rôle plus essentiel que le contour d'intonation dans l'expression émotionnelle. Cependant, d'une autre part, le contour d'intonation influence largement la perception émotionnelle par son interaction forte avec les traits grammaticaux de l'énoncé, ce qui est proposé par le *modèle de la configuration*. L'existence de cette interaction signale l'importance des facteurs linguistiques dans la parole émotionnelle et le danger de l'utilisation des techniques de masquage, qui détruisent la structure linguistique de l'énoncé, dans l'étude de l'expression émotionnelle. En bref, les résultats des expériences de Scherer *et al.* montrent l'existence de plusieurs types d'indice dans la voix émotionnelle, par exemple, la qualité de voix, qui s'opère selon les règles du *modèle de la covariation*, et le contour d'intonation, qui fonctionne selon les règles du *modèle de la configuration*.

Ladd *et al.* (1985) ont mené trois expériences, similaires à celles présentées ci-dessus, afin d'examiner le rôle de trois indices acoustiques, le niveau de la fréquence fondamentale ('Fo range'), le contour de Fo et la qualité de voix, dans l'expression émotionnelle. Ils ont utilisé des techniques de synthèse et de resynthèse digitale pour la manipulation systématique des paramètres à étudier. Leurs résultats montrent, premièrement, que les trois paramètres acoustiques fonctionnent indépendamment l'un de l'autre dans la communication affective, et deuxièmement, que le niveau de la fréquence fondamentale, parmi les trois paramètres, a plus d'effet sur la perception émotionnelle (par

[10] Pour la description détaillée de la technique de découpage aléatoire ('*Random splicing*'), voir Scherer (1971).

exemple, l'augmentation de la fréquence fondamentale est corrélée en fonction linéaire avec l'augmentation de l'excitation émotionnelle perçue par l'auditeur).

Concernant les concepts de la *covariation* et de la *configuration*, Scherer *et al*. (1998) fournissent une interprétation psychobiologique sur les déterminants des signaux vocaux dans l'expression et la communication de l'émotion. D'abord, ils constatent un degré remarquable de continuité phylogénétique dans les expressions émotionnelles chez les primates, par exemple la phonation différentielle des animaux et les différentes qualités de la voix humaine. Puis, ils soulignent le fait qu'un grand nombre des paramètres des expressions émotionnelles est 'domestiqué' dans le système de langage humain par l'utilisation des différents contours d'intonation et des mots affectifs. Enfin, ils concluent que le modèle de covariance serait convenable pour expliquer les cas où on attend l'influence directe des facteurs physiologiques sur la réalisation acoustique de l'expression émotionnelle alors que l'explication du modèle de configuration serait plus puissante dans les cas où les conventions socioculturelles et linguistiques sont dominantes. En résumé, les deux types d'approche se trouvent complémentaires pour rendre compte des différents facteurs dans la parole émotionnelle, telles que l'excitation physiologique, la motivation psychologique et les normes socioculturelles. En même temps, les concepts de l'effet '*push*' et de l'effet '*pull*' nous permettent de faire la distinction entre l'état transitoire physiologique du locuteur et son intention consciente pour la communication émotionnelle.

II.4.3. Modèles phonostylistiques (Troubetzkoy, Léon)

La phonostylistique, selon la définition de Léon (1993, p6), est une étude de l'oralité, qui s'intéresse à tout ce qui a été ou est susceptible d'être oralisé, produisant un effet par rapport au discours attendu. Cette science des styles sonores concerne essentiellement l'expressivité de l'oral dont les multiples facettes construisent les divers canaux de la communication parlée. Toute parole énoncée comporte des significations qui vont bien au-delà des sens véhiculés par les mots et la syntaxe. On identifie ainsi l'âge, le sexe, l'origine géographique et sociale, la personnalité du locuteur, aussi bien que son état émotionnel au moment de la prononciation. Ces informations phonostylistiques sont exprimées dans la parole, avec ou sans conscience du locuteur, et la communication du message linguistique varie en fonction de la modification des paramètres expressifs.

Dans la littérature phonostylistique, Troubetzkoy (1939, traduction française 1957) se trouve être le fondateur de la phonostylistique mais en même temps le premier à exclure l'étude de l'expressivité vocale de la linguistique formelle. Dans l'introduction aux « Principes de Phonologie », Troubetzkoy (ibid.) propose, pour la première fois, le terme *phonostylistique ('Lautstilistik')* pour l'étude des variations expressives dans la parole. Dans un chapitre plus tard, intitulé « Phonologie et phonostylistiques », il distingue la *stylistique phonétique* de la *stylistique phonologique*, qui étudient respectivement les phénomènes phonétiques comme des réalisations individuelles du son et les phénomènes linguistiquement codés comme des signes phonologiques.

En se basant sur le modèle de Bühler (1934), Troubetzkoy (ibid. p17-18) identifie trois fonctions phonostylistiques de la parole; une fonction *représentative* qui symbolise le message à communiquer, une fonction *expressive* qui caractérise le sujet parlant et une fonction *appellative* qui provoque une impression particulière sur l'auditeur. Notant les difficultés méthodologiques dans l'analyse des procédés *expressifs* et *appellatifs*, il justifie la limitation des linguistes à la fonction *représentative* de la langue. Cette argumentation se retrouve chez les fonctionnalistes, Roman Jakobson, André Martinet, puis plus tard chez Noam Chomsky et les générativistes dans la linguistique moderne. Le tableau suivant

montre comment les termes de la *fonction expressive* et de la *fonction appellative*[11], ont été interprétés par différents chercheurs, de Troubetzkoy jusqu'à Léon.

Fonction expressive (Troubetzkoy, 1939)	**Fonction appellative (Troubetzkoy, 1939)**
Fonction expressive (Martinet, 1960)	Fonction impressive (Guiraud, 1953)
Fonction émotive (Jakobson, 1963)	Fonction conative (Jakobson, 1963)
Plan présentatif (Rigault, 1964)	Plan expressif (Rigault, 1964)
Fonction identificatrice (Léon, 1971)	**Fonction impressive (Léon, 1971)**

Tableau 4. Différentes terminologies des fonctions phonostylistiques.

En opposition avec la négligence de l'étude de l'expressivité de la parole dans la linguistique formelle, Léon (1971) insiste sur l'importance des facteurs phonostylistiques dans la communication parlée et propose la notion de *phonostylème* en tant qu'unité d'analyse dans l'étude de la parole expressive. Selon sa définition (ibid., p10), le *phonostylème* est un ensemble des traits phonostylistiques, qui sont nécessaires et suffisants pour identifier les caractéristiques du message expressif de façon pertinente. Un *phonostylème* est constitué de multiples aspects fonctionnels, qui montrent les états psychologiques et sociologiques du locuteur et son intention ou son attitude vis-à-vis l'interlocuteur.

Inspiré par les notions de Bühler (1934), Léon (1971, 1993) distingue deux aspects de la parole, les *signaux volontaires* et les *indices involontaires*, qui servent des fonctions communicatives et expressives. Les *signaux volontaires* consistent en des signes conventionnels, codés dans la langue, (ex. l'accent d'insistance en français) tandis que les *indices involontaires* consistent en des signes non-codés, comme des symptômes pathologiques (ex. la toux, le bégaiement, le rire et le zézaiement). En termes saussuriens, il s'agit de *signes arbitraires*, proprement linguistiques, et de *signes motivés*, expressifs, qui entrent dans les domaines d'étude distinctifs, la *stylistique phonologique* et la *stylistique phonétique*, selon Troubetzkoy (1939). En élargissant le domaine phonostylistique d'une étude des effets purement linguistiques à une étude de l'expressivité

[11] La distinction entre la *fonction expressive* et la *fonction appellative* rappelle la subdivision de la phonologie par Laziczius (1939, traduction anglaise 1966). Il distingue d'abord deux types de variantes dans la parole, les variantes *linguistiques* et les variantes *phonostylistiques* (*expressives* et *emphatique*), puis identifie deux sous-domaines de la phonologie, la *phonologie expressive* et la *phonologie appellative*, en tant que l'étude distinctive entre les *variantes expressive* et les *variantes emphatiques*.

générale, Léon (ibid.) établit une *sémiotique vocale* dont le sujet d'analyse comprend les *indices* émotionnels, d'attitude et de personnalité aussi bien que les *signaux* linguistiques et communicatifs.

« *Les seules fonctions proprement phonostylistiques sont celles des signaux si l'on s'en tient aux règles classiques de la rhétorique, qui n'accorde de valeur stylistique qu'à l'effet conscient. Néanmoins, rien n'interdit d'analyser l'effet produit, même s'il est involontaire, tant dans l'analyse des discours oraux que dans ceux de l'écrit. On étudiera donc aussi bien les indices que les signaux, l'importance restant l'effet interprété par l'auditeur et non l'intention de l'émetteur.* » (Léon, 1979, p56)

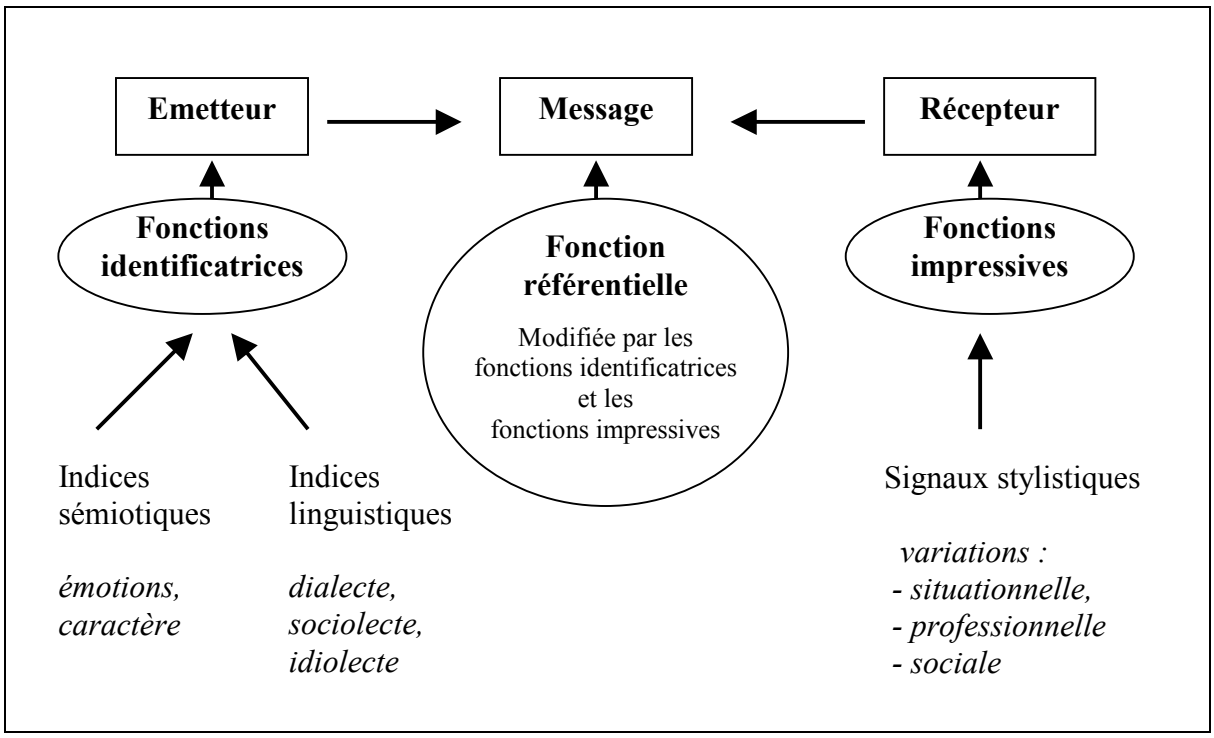


Figure 10. Modèle phonostylistique fonctionnel de Léon (1993).

Le modèle de la communication phonostylistique, proposé par Léon (1993), repose sur un système communicatif qui possède trois fonctions principales pour l'encodage et le décodage des signes. La *fonction référentielle* représente le message linguistique dans le sens lexical des mots et la structure syntaxique. La *fonction identificatrice* révèle des informations sur le sujet parlant, souvent exprimées involontairement, et cette fonction se

compose de deux sortes d'indices, l'une, les indices liés à l'état physiologique ou psychologique du sujet (comme les *émotions* et le *caractère*) et l'autre, les indices connotant l'appartenance du sujet à tel ou tel groupe dialectal ou sociologique (par exemple, l'accent du sud en français). La *fonction impressive* est constituée de signaux stylistiques, qui s'adressent au récepteur (comme les *attitudes*) et se manifestent selon des circonstances particulières. Ces signaux sont relativement bien codés et donc plus faciles à décoder par le récepteur.

Léon (1971, p6) note aussi la transition des indices involontaires aux signaux volontaires. Par exemple, le rire et la toux, d'origine involontaire, peuvent être employés par l'acteur de théâtre pour exprimer une attitude de politesse ou un signal d'appel. Un autre exemple est la nasalité dans la phonation du locuteur, qui est à origine un défaut physiologique, peut devenir un indice de l'émotion, de la personnalité, ou du groupe linguistique (comme pour le 'nasal twang' chez les Américains du mid-ouest). Cette transition explique l'ambiguïté des facteurs extralinguistiques, non-contrôlés, et les facteurs paralinguistiques, contrôlés, dans les phénomènes phonostylistiques.

II.4.4.Modèle du double codage (Fónagy)

Le modèle de Fónagy (1971a, 1978, 1983a, 1986b, 1990) suppose que le message dans la communication parlée est codé par deux actes successifs d'encodage : un encodage linguistique qui transforme un message global, une idée, en une séquence de phonèmes, et un deuxième codage au cours duquel le message secondaire, gestuel, est greffé sur le message primaire. Comme illustré dans la figure suivante, le premier encodeur (Encodeur 1) construit le champs sémantique du message et la structure grammaticale, et le deuxième encodeur (Encodeur 2) modifie son expression stylistique ou émotionnelle en tant que modulateur expressif. Le processus de décodage est l'image miroir du processus d'encodage. Un décodeur (Décodeur 2) révèle le message primaire linguistique et un autre décodeur (Décodeur 1) retrouve le message secondaire expressif par la réinterprétation des ondes sonores.

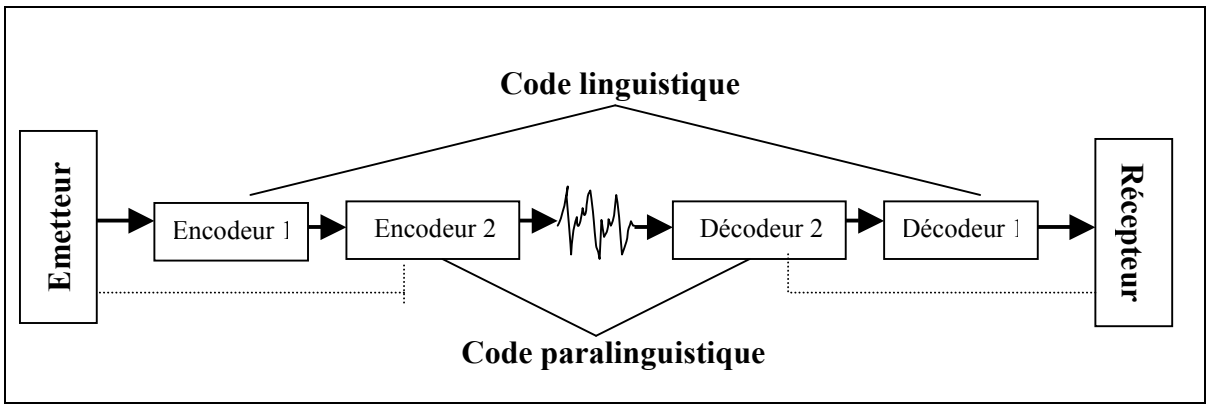


Figure 11. Modèle du double codage de Fónagy (1983a).

La ligne en pointillé dans la figure, liant l'émetteur et Encodeur 2 ou celle liant le récepteur et Décodeur 2, signifie que l'encodage et le décodage des messages secondaires se situent sur un autre niveau mental que les codages des messages primaires. Concernant l'ordre des encodeurs et des décodeurs, on constate que l'encodeur des messages primaires se trouve avant l'encodeur des messages secondaires tandis que le décodeur des messages primaires vient après le décodeur des messages secondaires. Cette contradiction apparente peut être expliquée par le fait suivant. D'abord, la terminologie de Fónagy : l'information linguistique est le message primaire et l'information expressive est le message secondaire est reflétée dans l'ordre Encodeur 1 et Encodeur 2. C'est-à-dire, la notion de modulateur du

deuxième encodeur suppose l'existence préalable d'un message à modifier, le message primaire. L'ordre de Décodeur 2 et Décodeur 1, qui paraît contradictoire à la logique précédente, peut être justifié par le passage suivant, qui nous rappelle l'argument de Zajonc (1980) pour la primauté de l'émotion.

« *Emotional mental processes precede – both ontogenetically and phylogenetically – the development of conceptual thought. Similarly, the child achieves understanding of emotion prosodic features prior to comprehending work* » (Fónagy,1978, p35).

Cependant, le modèle de Fónagy ne précise pas la relation des processus émotionnels expressifs avec les processus cognitifs généraux ou la place des codages phonostylistiques parmi d'autres processus mentaux. Ce modèle s'intéresse surtout à l'existence des deux sortes de codage, codage linguistique et codage phonostylistique, dans la réalisation concrète des phonèmes, et au rôle des codeurs dans la production et la perception de la parole expressive.

Selon l'expression de Fónagy (1983a, p19), "*tous les sons concrets sont expressifs"*. Un phonème, unité abstraite, doit être réalisé, actualisé à l'aide des organes de la parole, la glotte, le pharynx, la langue, les lèvres, pour apparaître dans le discours sous la forme sonore. Tous les phonèmes passent nécessairement par le modulateur pour la réalisation concrète et les différents effets expressifs dans la communication sont donc engendrés par la modulation ou distorsion de la réalisation habituelle du phonème. C'est ainsi que deux sons identiques, représentant le même phonème, peuvent exprimer des messages divergents, telles que menace, tendresse, tristesse et joie, à travers des gestes vocaux. La façon de cette modulation est appelée '*style vocal*' ou '*style*' tout court.

Le *style* consiste en une série de manipulations expressives des phrases engendrées par la grammaire. Il est gouverné par deux sortes de règles d'ordre linguistique et paralinguistique. La grammaire construit des *éléments distinctifs*, avec un nombre limité d'unités *discrètes*, tandis que le modulateur expressif crée des *signes entiers*, dans une dimension *continue*. Le rapport entre le sens et les éléments sonores du mot est généralement *arbitraire* alors que les signes créés par le modulateur sont toujours *motivés*. Ces signes stylistiques n'échappent pas aux *conventions*. Fónagy (1983a, p17-18) formule

des règles qui expliquent la production des signaux gestuels dont les idées principales sont les suivantes.

a) La reproduction volontaire des symptômes vocaux d'une émotion signale la présence de cette émotion. Par exemple, la contraction des muscles du pharynx signale la nausée, le déplaisir, le mépris, la haine, etc.
b) Les organes de la parole peuvent représenter, symboliser d'autres objets animés ou inanimés qui leur sont associés par la ressemblance ou une analogie fonctionnelle. Par exemple, le déplacement de la langue vers l'avant et le haut est similaire au déplacement du bras et l'approche de la langue vers le palais peut représenter des objets proches ou petits.
c) Le principe de l'isomorphisme de l'expression et du contenu exige que les différents degrés d'intensité physiologique ou acoustique correspondent à différents degrés d'intensité psychologique ou sémantique. Par exemple, l'intensité de la tension musculaire reflète l'intensité de l'émotion chez un sujet.

Ces règles déterminant la structure du style vocal s'opèrent à tous les niveaux de la communication verbale et les effets de cette opération se manifestent dans des traits prosodiques de la parole. Selon Fónagy, "*le style vocal est omniprésent*" (1983a, p23). Les différentes expressions stylistiques peuvent être caractérisées par différents traits prosodiques, mesurés avec des paramètres acoustiques comme la fréquence fondamentale, l'intensité, et la durée (voir Fónagy, 1972, 1977, 1980, 1981).

Parmi les effets stylistiques, Fónagy (1987, p82) fait la distinction entre l'*émotion*, l'*attitude* et la *modalité* selon leur degré de régularité conventionnelle. Les *émotions* primaires sont exprimées à tous les niveaux des appareils vocaux par des changements physiologiques, ce qui rend la nature de l'expression émotionnelle essentiellement paralinguistique. L'expression d'*attitude*, s'adressant à l'interlocuteur, repose sur un niveau plus élaboré et plus régularisé en termes de conventions communicatives. L'expression de la *modalité* linguistique est hautement précisée et structurée dans un système grammatical, par exemple l'intonation montante pour la question et l'intonation descendante pour l'affirmation. Le changement des indices motivés paralinguistiques aux signes distinctifs est bien expliqué par Fónagy (1976a, 1976b,1990).

II.5. Etudes expérimentales de l'émotion vocale

Dans la partie II.5., nous résumons les résultats expérimentaux obtenus par des études psychologiques, linguistiques et scientifiques sur les indices acoustiques et perceptifs de l'émotion vocale. L'hypothèse de base est que l'excitation émotionnelle cause des changements de la tension musculaire des appareils vocaux, de la salivation, de la respiration et du registre de phonation du locuteur, ces changements pouvant être mesurés par des paramètres acoustiques, et ainsi les différentes émotions sont exprimées dans la voix et peuvent être identifiées par l'auditeur même en l'absence d'information lexicale. La technologie vocale, synthèse et reconnaissance de la parole, a atteint un niveau tel qu'elle s'intéresse non seulement à la production d'une voix intelligible mais à l'intégration des effets expressifs, émotionnels et stylistiques dans la voix synthétique.

Malgré un grand nombre d'expériences faites sur les expressions émotionnelles, il est encore difficile d'établir un ensemble des indices vocaux qui différencie les diverses émotions et les caractérise de manière prototypique. Cette difficulté vient du fait suivant. D'une part, la nature complexe de l'émotion implique des facteurs physiologiques, psychologiques et sociologiques, dont les effets se manifestent dans des indices différents, même contradictoires. D'une autre part, différentes études utilisent différents termes et différentes méthodes d'analyse, ce qui rend difficile la comparaison ou la synthèse de leurs résultats (Scherer, 1986 ; Murray & Arnott, 1993).

La partie II.5. se déroule en trois étapes. D'abord elle présente des études expérimentales qui mettent en évidence la variation des profils acoustiques en fonction de la signification émotionnelle. Ensuite, elle présente des études perceptives qui concernent des indices non-verbaux dans la communication de l'émotion. Enfin, elle discute l'application des résultats de ces études à la technologie de la parole.

II.5.1.Etudes acoustiques

L'étude expérimentale sur la voix émotionnelle consiste en deux types d'analyse. Le premier type, l'analyse acoustique, cherche à identifier des traits acoustiques ou articulatoires qui varient en fonction de la différente signification émotionnelle, afin d'établir un système de traits prosodiques des émotions. L'autre type, l'analyse perceptive, s'intéresse à la capacité de l'auditeur à reconnaître des émotions avec des indices vocaux seuls, même quand il ne comprend pas le sens du mot ou de la phrase. La plupart des études combinent les deux types d'analyse dans leurs méthodes, en considérant les aspects expressifs et perceptifs de l'émotion comme une entité à double face. Cependant, nous distinguons les études orientées vers l'analyse acoustique et les études orientées vers l'analyse perceptive dans les parties II.5.1 et II.5.2, dans le but de faire une présentation simple et claire.

L'analyse acoustique de l'émotion vocale concerne la variation des trois paramètres principaux, la fréquence fondamentale (Fo), l'intensité et la durée, de la voix émotionnelle par rapport à la voix dite neutre. Une série d'études expérimentales démontrent que les valeurs moyenne, maximale et minimale de Fo, l'amplitude du signal, le rythme et le débit de la parole varient selon les différentes émotions (Skinner, 1935 ; Fairbanks & Hoaglin., 1939, 1941 ; Black, 1961 ; Williams & Stevens, 1972 ; Cosmides, 1983 ; Laukkanen *et al.*, 1996 ; Leinonen *et al*., 1997). L'hypothèse de ces études est que la variation des valeurs acoustiques reflète directement des changements physiologiques du locuteur, qui ressent l'émotion dans une situation affective ou imite (stylise) l'émotion dans un laboratoire, et leur but est de trouver la régularité de cette variation en fonction de l'émotion, c'est-à-dire les invariants acoustiques de l'émotion.

Parmi les indices acoustiques, le Fo est le paramètre le plus souvent étudié et il est considéré comme un indice particulièrement pertinent dans l'expression et la perception de l'émotion. L'augmentation du Fo est étroitement liée à l'élévation de l'excitation émotionnelle ou du stress mental (Hecker *et al*., 1968 ; Scherer, 1979 ; Streeter *et al*., 1983). L'intensité est aussi corrélée à l'excitation émotionnelle (Cowan, 1936 ; Hutter, 1968) mais son augmentation est dépendante de la valeur du Fo (Black, 1961). Fónagy

(1981) précise que l'augmentation de l'intensité amène l'allongement des voyelles et le raccourcissement des consonnes, des liquides et des nasales. Le débit varie en fonction de l'émotion mais c'est la proportion du temps de phonation et du temps de pause, plutôt que le débit lui-même, qui différencie les émotions (Cowan, 1936 ; Fairbanks & Hoaglin, 1941). La perturbation du Fo ('*jitter*') et la forme d'onde de la vibration glottique sont estimées en tant qu'indices des différentes qualités de la voix dans les expressions émotionnelles ou stylistiques (Lieberman, 1961 ; Cummings & Clements, 1995) mais leur validité nécessite plus d'évidences expérimentales (Protopapas & Lieberman, 1995).

La variation des paramètres acoustiques de la voix émotionnelle est souvent décrite en termes du degré de déviation de leurs valeurs par rapport aux valeurs relevées dans la voix neutre. Quelques caractéristiques acoustiques des émotions, considérées comme primaires, sont présentées dans le Tableau 5. L'utilisation des traits vocaux pour déterminer des états psychopathologiques des malades a été proposée par les psychothérapeutes (Roessler & Lester, 1976 ; Smith, 1977 ; Breznitz, 1992 ; Norris *et al*., 1995). Les chercheurs sont pourtant conscients du fait qu'il n'existe pas de corrélat acoustique unique pour caractériser une telle émotion particulière. De plus, les corrélats acoustiques d'une émotion donnée varie considérablement selon ses variantes de différentes formes. Par exemple, une tristesse tranquille fait diminuer le Fo et l'énergie en hautes fréquences de la voix, ce qui est noté dans la plupart des études (voir le Tableau 5), tandis que la tristesse excitée (comme le désespoir ou le détresse) fait augmenter le Fo et l'énergie, ce qui est constaté dans notre analyse acoustique (voir IV.4).

	Domaine fréquentiel	**Domaine temporel**	**Qualité de la voix**
Joie	- Fo moyen élevé - variation du Fo dynamique	- débit rapide - structure rythmique avec accentuation régulière	- intensité élevée mais pas autant que pour la colère - voix légèrement aspirée
Colère	- Fo moyen élevé ou modéré - variation du Fo dynamique - contour de Fo fortement descendant en fin de phrase	- débit plus rapide que la voix neutre mais moins que pour la joie	- intensité élevée - voix aspirée et tendue - grande énergie dans les hautes fréquences
Tristesse	- Fo moyen au niveau de la voix neutre - peu de variation du Fo	- débit lent - rythme avec des pauses régulières	- intensité basse sans variation - articulation moins précise
Peur	- Fo moyen légèrement élevé - grande variation du Fo, mais pas autant que pour la joie ou la colère	- débit plus lent que pour la joie et la colère mais plus rapide que la voix neutre - pauses irrégulières	- intensité basse - articulation précise - peu d'énergie dans les basses fréquences

Tableau 5. Les indices acoustiques des émotions primaires, d'après Abadjieva *et al*. (1993).

II.5.2. Etudes perceptives

Les études perceptives de l'expression émotionnelle soulignent que l'auditeur peut reconnaître ou identifier les différentes émotions à travers la voix sans recours à l'information lexicale des phrases. Leurs expériences s'adressent à la perception émotionnelle avec des stimuli dont le contenu lexical n'est pas pertinent. Pour enlever l'influence de l'information lexicale de l'identification des émotions, les chercheurs se servent de trois techniques majeures ; (a) l'utilisation des énoncés sémantiquement neutres, par exemple les mêmes phrases (ou mots), les lettres alphabétiques, et la parole réitérée, pour les différentes expressions émotionnelles (Kaiser, 1962 ; Bluhme, 1971 ; Davitz & Davitz, 1974) ; (b) l'utilisation d'une langue inconnue des auditeurs dans les expressions émotionnelles (Kramer, 1963a ; Van Bezooyen, 1984) ; (c) la manipulation acoustique des énoncés au moyen de l'ajout du bruit, de la synthèse des stimuli, du filtrage électronique ou du découpage aléatoire de la phrase (Pollack *et al*., 1960 ; Lieberman & Michael, 1962 ; Roger *et al*., 1971 ; Scherer *et al*., 1972).

Le but des études perceptives est de modéliser l'utilisation des paramètres acoustiques, les indices non-verbaux, par l'auditeur dans la communication vocale de l'émotion. Costanzo *et al.* (1969) proposent que l'auditeur se base sur un ensemble d'indices acoustiques pour identifier l'émotion, appelé le *profil vocal* (‘*Vocie Quality Profil*’). Ces indices sont différents selon le type d'émotion. Par exemple, l'auditeur identifie la tristesse et la tendresse d'après leur *profil fréquentiel*, la colère et le mépris d'après leur *profil d'intensité* et l'indifférence, en tant qu'émotion neutre, d'après son *profil temporel*. Les expériences de Scherer (1974) et de Scherer & Oshinsky (1977) mettent en évidence que l'auditeur attribue les différentes significations émotionnelles en fonction des valeurs du Fo, du rythme et de l'amplitude, même aux stimuli de tons pures. L'utilisation des indices acoustiques dans l'identification de la personnalité est montrée par Brown *et al*. (1974) : la modification des valeurs du Fo moyen, de la variation du Fo et du débit, affecte la perception de la personnalité comme la compétence et la bienveillance.

Les facteurs externes de la perception de l'émotion, tels que la différence individuelle, le sexe et la connaissance linguistique et culturelle, ont été examinés par des

expériences avec de diverses techniques . Les résultats globaux indiquent que la perception des émotions primaires reste stable : les auditeurs identifient toujours les émotions avec une précision supérieure à celle qui pourrait être faite au hasard, mais il existe des différences dues aux facteurs externes dans la perception émotionnelle. Dusenberry & Knower (1939) suggère la différence individuelle et la supériorité des femmes dans l'identification des émotions. Kramer (1964), McCluskey *et al*. (1975) et Wallbott & Scherer (1988) montrent les différences de la perception de l'émotion parmi les sujets de différente nationalité. Nash (1974) a démontré que la profession psychiatrique n'implique pas nécessairement une meilleure capacité de l'identification des émotions par rapport aux autres professions. Dans les expériences de la perception de l'émotion, les émotions négatives sont souvent mieux identifiées que les émotions positives (Bluhme, 1971 ; Chung, 1995a). Ce phénomène semble être dû au fait que l'être humain fait attention, de manière prioritaire inconsciemment, aux stimuli négatifs, désagréables, pour une raison de protection. Autrement dit, l'être humain est plus apte à percevoir l'émotion négative que l'émotion positive dans les processus perceptifs. Cette tendance est appelée la *supériorité de l'émotion négative* par Hansen & Hansen (1988), et la *vigilance automatique* par Pratto & John (1991).

II.6. Technologie vocale

Dans la communication parlée, le locuteur exprime son intention communicative, sa personnalité et son état émotionnel dans la voix sans effort particulier, et l'auditeur les perçoit de manière inconsciente. Ces capacités expressive et perceptive paraissent naturelles, même banales, au point de vue humain, mais en fait, elles impliquent un grand nombre de facteurs à être considérés et spécifiés au point de vue technique pour que la machine puisse imiter ou comprendre la parole humaine. Il est bien connu que la richesse ou la complexité de ces capacités est largement due à la *prosodie*, définie comme un ensemble de variations des traits acoustiques tels que le Fo, la durée et l'intensité (Landercy & Renard, 1977, p232). L'importance de la *prosodie* dans les systèmes de la synthèse ou de la reconnaissance automatique de la parole a été souvent soulevée par les chercheurs (Pierrehumber, 1981 ; Rossi, 1988 ; Prevost & Steedman, 1994; Vaissière, 1998 ; Scherer *et al*., 1998) mais l'implantation des traits prosodiques dans les systèmes pour la parole expressive (stylistique et émotionnelle) et la reconnaissance fiable à travers

la variabilité inter- et intra-locuteur, est encore très limitée à cause des problèmes non-résolus aux niveaux conceptuel et pratique (voir Fant *et al*., 1991 ; Homayounpour, 1993 ; Murray *et al*., 1995 ; Batliner *et al*., 1998).

En ce qui concerne la synthèse de la parole émotionnelle, peu de travaux ont été réalisés et la plupart des systèmes commercialisés ne peuvent pas prendre en compte les effets expressifs dans la parole synthétique (voir Scherer *et al*., 1998). En tant qu'essai pionnier, le système de Murray (1989), appelé '*HAMLET*,' développé à la base du synthétiseur DECTalk (version 2.0), produit des expressions émotionnelles dans la voix synthétique, par la modification des paramètres prosodiques comme le contour de Fo, le débit et la qualité de voix. Le système '*HAMLET*" estime d'abord les valeurs prosodiques d'une voix neutre pour une séquence de phonèmes, en s'appuyant sur les règles d'Allen *et al*. (1987)[12], et puis modifie ces valeurs selon ses propres règles, qui sont établies à partir des résultats d'études précédentes dans la littérature de l'émotion[13]. Murray & Arnott (1995) exposent les processus de l'ajout de six émotions à la voix synthétique et soulignent l'avantage de ce système, étant un synthétiseur par règles génératives complètes. Un autre système similaire proposé par Cahn ('*Affect Editor*,' 1990), se base aussi sur les règles mais il nécessite quelques ajustements manuels au cours des processus de synthèse. Dans le même but de synthèse par règles des émotions vocales, Mozziconacci (1998) propose la modification des contours de Fo comme moyen efficace dans la production des différentes impressions émotionnelles, d'après les résultats de ses expériences faites au moyen du système '*IPO Text-To-Speech*'[14]. Pourtant, Mozziconacci (ibid., p72) note également qu'il n'y a pas de relation univoque entre les contours intonatifs et les émotions catégoriques et que d'autres facteurs comme le niveau de Fo, la plage de Fo, le débit, et la durée des segments accentués et non-accentués, sont également importants dans la production des effets émotionnels.

[12] Les règles intonatives du système '*MITalk*' proposé par Allen *et al*. (1987) doivent largement à O'shaughnessy (1976).
[13] Le Tableau 5 d'après Abadjieva *et al*. (1993), présenté dans la partie II.5.1, est un des exemples de résumés des résultats expérimentaux. En fait, ce tableau a été construit pour spécifier les paramètres du système '*HAMLET*".
[14] Le système '*IPO Text-To-Speech*' est développé pour la synthèse de parole en hollandais. Voir 't Hart *et al*. (1990, p176-180) pour la description du système.

II.7. Conclusion du chapitre II.

Dans le chapitre II, nous avons passé en revue les études sur l'émotion faites dans des domaines tels que la philosophie, la psychologie, la linguistique, et les technologies vocales. Cette révision s'est déroulée d'une façon panoramique, à partir de la réflexion du concept de communication générale, jusqu'aux recherches actuelles sur l'émotion, incluant la synthèse et la reconnaissance automatique de la parole expressive.

D'après notre revue, un des problèmes majeurs de l'étude de l'émotion est qu'il n'existe pas de théorie standardisée de l'émotion, qui peut fournir un cadre général aux différentes approches. La grande diversité des théories, surtout à l'égard de la relation entre le sentiment affectif, le processus cognitif, et la réaction physiologique, résulte en une différente signification du même terme 'affectif' selon les études, d'où la difficulté de comparer les résultats. Un autre problème essentiel du point de vue expérimental est que le concept d'émotion échappe souvent à la mesure quantitative, et il est donc difficile d'obtenir les données authentiques sur l'émotion naturelle et d'évaluer la variation de l'expression émotionnelle de manière objective et rigoureuse. En ce qui concerne le choix de paramètres, il est reconnu que la qualité de voix joue un rôle crucial dans l'expression et la perception de l'émotion vocale, mais il n'y a pas de méthodes encore bien établies pour l'estimer en termes de paramètres acoustiques. Le problème de l'authenticité des données et du contrôle expérimental est discuté en détail dans le chapitre suivant.

Malgré les difficultés conceptuelles et méthodologiques, la recherche sur l'émotion permet d'expliquer des phénomènes affectifs dans la communication humaine. Une meilleure connaissance de l'émotion contribuera à l'amélioration de la synthèse vocale pour la qualité de voix plus naturelle, voire plus proche de la voix humaine, et au renforcement du système de la reconnaissance automatique pour qu'il identifie la parole à travers des gens avec différentes personnalités et des voix variables en fonction des états psychologiques (panique, émotion, attitude, etc.) ou physiologiques (fatigue, maladie).

Chapitre III

Emotion dans la parole spontanée

Résumé

Ce chapitre présente la méthodologie de ce travail, particulièrement le choix de la nature des données d'expressions émotionnelles. Nous expliquons ici les raisons pour lesquelles nous avons choisi des données d'expressions émotionnelles acquises à partir de la parole spontanée plutôt qu'à partir du jeu d'un acteur. Dans l'étude de l'émotion, deux types de données peuvent être distingués selon la manière de production : émotions exprimées par un sujet qui se trouve réellement dans un tel état émotionnel (*émotion vécue*) vs. émotions exprimées par un acteur qui imite un tel état émotionnel (*émotion stylisée*). Nous proposons que cette distinction soit similaire à celle proposé par Fagyal (1995) entre la parole spontanée et la parole lue. L'avantage et le désavantage de chaque type de données sont expliqués selon une échelle de l'authenticité des données et du contrôle expérimental. L'émotion vécue et la parole spontanée sont des données favorables pour capter la richesse de la réalité de l'expérience improvisée du locuteur, mais ces données de nature spontanée ne donnent que des possibilités très limitées au niveau du contrôle expérimental des données. Or, l'émotion stylisée et la parole lue sont des données plus fréquemment utilisées dans les études précédentes à cause de la facilité d'acquisition mais leur représentativité de la réalité naturelle est plus ou moins problématique. Dans le présent travail, nous avons choisi d'étudier le premier type de données (spontanées) pour examiner les phénomènes de la parole émotionnelle tels qu'ils sont. Une approche typologique fonctionnelle est adoptée dans cette étude pour que les données soient sélectionnées sur des critères fonctionnellement définis et qu'elles soient examinées en fonction des facteurs systématiquement manipulés dans une expérience donnée.

III. Emotions dans la parole spontanée

III.1. Préliminaires

Dans ce chapitre, nous proposons d'introduire une discussion sur la parole spontanée et l'émotion vécue, afin de mettre en valeur les données de notre travail, qui sont des extraits de la parole spontanée émotionnelle. La structure de ce chapitre se compose de deux comparaisons, l'une entre la parole *spontanée* et la parole *lue* et l'autre entre l'émotion *vécue* (*réelle*) et l'émotion *simulée* (*stylisée*). Dans la linguistique et la psychologie, relativement peu d'études ont été menées sur la parole *spontanée* (Delattre, 1965 ; Lieberman *et al*., 1985 ; Morel & Danon-Boileau, 1995) et sur l'émotion *vécue*, exprimée dans la parole *spontanée* (Hutter, 1968 ; Williams & Stevens, 1972 ; Brown, 1980).

Un des obstacles majeurs dans ce genre d'étude est le problème méthodologique de l'expérimentation. La parole spontanée et l'expression des émotions vécues sont des données, par nature non-contrôlées, tandis que l'expérimentation scientifique exige un contrôle rigoureux sur les données à analyser. Ce conflit fondamental entre la nature des données et la nécessité du contrôle expérimental fait engendrer les notions de parole *lue* et d'émotion *simulée* ou *stylisée*. Vu la difficulté de l'acquisition et de l'analyse des données *spontanées*, les chercheurs analysent souvent des données répétées et enregistrées dans un laboratoire, et les considèrent comme représentatives de la parole ou de l'émotion exprimée dans un milieu naturel.

Nous présenterons d'abord des problématiques dans l'étude de la parole spontanée et l'étude de parole lue, ce qui nous dirigera vers une discussion sur la relation entre la spontanéité du corpus et le contrôle expérimental. Cette discussion est développée dans la partie suivante, concernant les expressions émotionnelles, telles l'émotion vécue et de l'émotion simulée. La conclusion résume les discussions et explique la signification du chapitre III par rapport à nos analyses présentées dans les chapitres suivants.

III.2. Parole spontanée vs. parole lue

La parole *spontanée* est souvent utilisée en phonétique à l'opposé de la parole *lue*. La parole *spontanée* se réfère à l'énonciation sans planification ou indication de ce qui doit être prononcé tandis que la parole *lue* présuppose l'existence d'un texte à lire et la préparation de la lecture, telle la parole oratoire ou la lecture à voix haute.

III.2.1. Parole spontanée

Hagège (1985, p84) distingue le *style parlé*, équivalent à la parole *spontanée*, du *style oral*, symétrique d'écriture. Selon lui, le premier désigne l'usage ordinaire de la parole, qui se produit en situation interlocutoire, alors que le dernier est un genre littéraire, qui consiste en refrains, proverbes et rimes poétiques. Pour Léon (1993, p6), l'*oralité* comprend une gamme plus large, signifiant tout ce qui est parole proférée qu'elle soit lue, récitée, formalisée, stylisée ou non. Il considère la parole *spontanée* comme un type particulier de l'*oralité*, qui est fait de reprises, de fautes et d'interruptions de phrases. Léon (ibid.) note que la parole *spontanée* suit rarement le modèle des énoncés, bien formés d'un locuteur idéal[15] mais les contraintes linguistiques restent pourtant les mêmes dans la parole spontanée. Ses préoccupations phonostylistiques concernent tout ce qui est *oralisé* dans la communication humaine.

Fagyal (1995) définit ainsi la parole *spontanée* : (1) La parole *spontanée n'est pas* de la parole *lue* ; (2) elle consiste en un message *non répété* et *non planifié* à l'avance ; (3) elle est propre aux situations *informelles* ; et (4) elle est *énoncée de mémoire* dans des situations de communication *réelles et naturelles*. Selon cette définition, la parole *spontanée* est l'encodage d'informations linguistiques en temps réel dans lequel l'informalité de la circonstance d'énonciation et l'absence de contrôle ou d'observation directe sont des facteurs importants pour rendre la parole naturelle. Un corpus d'expérience ainsi acquis serait idéal en tant que données de la parole naturelle, mais ce genre de corpus donne peu de possibilité de contrôle à l'expérimentateur. En d'autres termes, les

[15] Cette remarque de Léon rappelle la *performance linguistique*, à l'opposition de la *compétence linguistique*, dans les notions de Chomsky (1965).

phonéticiens veulent avoir des données plus semblables à la parole naturelle, et en même temps intervenir dans la production et l'acquisition des données pour contrôler des facteurs aléatoires. Or, leur intervention peut affecter la naturalité ou la spontanéité de la parole. Ce dilemme ne se trouve pas seulement en phonétique mais aussi dans d'autres domaines comme la sociologie et la psychologie. Labov décrit le problème sous le nom de *paradoxe de l'observateur* :

« *We are then left with the Observer's Paradoxe : the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed ; yet we can only obtain these data by systematic observation* » (1972, p209).

Echelle de l'authenticité et du contrôle expérimental des données

AUTHENTICITE				**CONTROLE EXPERIMENTAL**
Non-spontané	0	LECTURE		**contrôlé**
	1	Logatomes	12	
	2	Mots	11	
	3	Phrases 'réitérées'	10	
	4	Phrases 'probabilitaires'	9	
	5	Phrases ordinaires	8	
	6	Textes	7	
Semi-spontané				**Semi-contrôlé**
		DIALOGUE		
	7	Question-Réponses prédéfinies	6	
	8	Questions avec réponses non définies	5	
	9	Description d'image	4	
	10	Interprétation d'image	3	
	11	Conversation dirigée	2	
	12	Rappel de mémoire	1	
Spontané		……..	0	**Non-contrôlé**

Le degré de l'authenticité des données (échelle de spontanéité de 1 à 12) est inversement proportionnel au degré du contrôle expérimental (échelle de contrôle de 12 à 1).

Figure 12. Echelle de l'authenticité et du contrôle expérimental, construite par Fagyal (1995, p46)[16].

[16] Pour la description de chaque technique inventoriée dans l'échelle, voir Fagyal (1995, 45-53).

Cette situation paradoxale est expliquée par Fagyal (1995) avec une échelle qui représente la relation entre l'authenticité des données et le contrôle expérimental. Elle énumère dans l'échelle les différents types de données de l'expérience phonétique, dont le degré d'authenticité est inversement proportionnel au degré du contrôle expérimental. Selon cette échelle, les données construites par rappel de mémoire ('*recall*' ou '*retelling*' en anglais), sont placées au niveau le plus authentique à la parole naturelle, du fait que le contenu n'est pas planifié et répété à l'avance et le locuteur raconte des histoires d'une façon informelle sous tension émotionnelle. Le rappel de mémoire signifie que l'expérimentateur demande au sujet de raconter des expériences de la vie privée. Le récit ainsi obtenu représente la meilleure approximation des processus d'encodage de la parole naturelle, ce qui rend le maximum degré d'authenticité des données. Pourtant on n'y obtient qu'un minimum de contrôle sur les structures syntaxiques et lexicales du récit. Fagyal, se référant à Labov (1972), précise les circonstances de la *parole réellement spontanée* :

« *Les récits produits sous tension émotive, dont le rappel de l'enfance ou celui des situations de danger de mort sont particulièrement adaptés à servir d'échantillons de parole réellement spontanée : ils ne sont pas prévus et répétés à l'avance, et l'attention du locuteur est capturée par l'émotion, l'empêchant de se concentrer sur le style* » (1995,p51).

Cette remarque attire particulièrement notre attention, car le corpus de notre travail consiste en des extraits des récits où les locuteurs se rappellent leur enfance et racontent leurs problèmes actuels. Un passage plus tard nous inspire des méthodes pour l'analyse de ce genre de corpus, telles l'observation sélective des données, la description paramétrique et la vérification par des tests perceptifs :

« *Il semblerait que la méthode appropriée soit plus l'observation sélective que l'expérimentation contrôlée. Lorsque la manipulation de la situation de communication – aussi imperceptible et indirecte soit-elle – s'avère impossible ou indésirable, il est conseillé de procéder par typologie : classer et comparer les échantillons selon certains critères précis. Les 'preuves' expérimentales peuvent être déduites, par la suite, des tests de perception où les liens entre les paramètres systématiquement modifiés sont établis en fonction des jugements des auditeurs*. » (Fagyal, 1995, p52).

L'étude de la parole spontanée connaît une renaissance à l'heure actuelle (voir Blanche-Benveniste, 1997). La parole spontanée ('*casual speech*') et la parole soutenue ('*careful speech*') sont distinctes par leur durée et la différente réduction des traits phonétiques (Moon & Lindblom, 1994). Les syllabes accentuées jouent un rôle essentiel dans la distinction de la parole spontanée ('*spontaneous speech*') et de la parole lue ('*read speech*') (Blaaw, 1995). En anglais, la parole spontanée a un débit plus rapide, moins de variations de Fo, moins de déclinaison de Fo, plus de perturbations d'amplitude ('*shimmer*') et plus de réduction des voyelles que la parole lue (Laan, 1997). Ces caractères de la parole spontanée semblent être aussi valables pour le français (Di Cristo, 1985 ; Léon & Tennant, 1990 ; Duez, 1995). L'application de ce genre d'étude au développement de la reconnaissance automatique de la parole et de l'appareil acoustique pour les malentendants est suggérée par Eskenazi (1995) dans sa revue des études du style parlé ('*speaking style*').

III.2.2. Parole lue

Dans beaucoup d'expériences phonétiques, les données se composent de syllabes, de mots, de phrases ou de textes, qui sont construits sur certains critères expérimentaux et prononcés à plusieurs reprises. Autrement dit, leurs données sont constituées de parole lue. Etant donné que la structure des données est contrôlée au début et la réalisation phonétique est répétée dans une condition contrôlée, la comparaison des données peut se faire de manière systématique.

Différents types de données sont proposés en tant que moyens expérimentaux de l'étude phonétique, qui sont représentatifs de la parole naturelle mais en même temps, contrôlés par l'expérimentateur (voir Fagyal, 1995, p46). Dans l'échelle de Fagyal (ibid.), les *logatomes* représentent les données les plus contrôlées du point de vue expérimental mais le moins authentique. Ils se composent de syllabes conformes aux règles phonotactiques de la langue. La structure des logatomes est strictement contrôlée pour l'analyse systématique, et elle est censée susciter la réaction naturelle du sujet dans une expérience de perception. Ce genre de données est utile pour étudier l'encodage et le décodage de certaines unités linguistiques (syllabes ou mots) virtuelles dans la langue. Par exemple, à la suite de la comparaison entre la vitesse de l'accès au lexique à partir des mots

et celle à partir des logatomes (non-mots), il a été découvert que la perception des mots nécessite plus de temps d'accès que la perception des logatomes, à cause d'un appel au module lexical.

Au fur à mesure que les données comprennent plus d'informations sémantiques, syntaxiques et pragmatiques, elle sont plus semblables à la parole naturelle, mais moins contrôlables par l'expérimentateur. Fagyal (1995, p47) considère les *phrases probabilitaires*, qui consistent en l'agencement des logatomes selon les règles syntaxiques de la langue en question, comme un meilleur compromis entre l'authenticité des données et le contrôle expérimental. En citant Fónagy (1981), qui a utilisé des phrases probabilitaires pour l'expression des différentes attitudes émotionnelles (colère, tendresse, reproche, menace, coquetterie, etc.) en hongrois, elle insiste sur l'avantage de ce type de données. Les auditeurs jugent les émotions exprimées dans les phrases probabilitaires, uniquement sur base du rythme et de l'intonation, sans information sémantique.

Beaucoup de modèles prosodiques sont établis à partir des phrases lues, avec informations syntaxiques et sémantiques ; par exemple, en anglais (Lieberman, 1967 ; Ladd, 1978 ; Pierrehumbert, 1980 ; Selkirk, 1984), en français (Martin, 1982), en hollandais ('t Hart *et al*., 1990) et en coréen (Koo, 1986 ; Ko, 1988 ; Lee, 1990). Ces modèles sont considérés comme la description générale de la langue donnée mais la validité de ces modèles pour la parole spontanée est à examiner.

III.3. Emotion vécue vs. émotion simulée

La distinction entre l'émotion *vécue* et l'émotion *simulée* est similaire à celle entre la parole *spontanée* et la parole *lue*. Les émotions *vécues* se produisent dans des situations naturelles, d'où vient la parole *spontanée*, tandis que les émotions *simulées* sont faites dans des cadres artificiels, tels que le théâtre et le laboratoire, comme la parole *lue* est enregistrée dans un local insonorisé. L'expression des deux sortes d'émotions paraît semblable puisque la deuxième est censée refléter la première mais du point de vue motivationnel, elles sont tout à fait différentes. L'expression de l'émotion *vécue* se base sur des éléments physiologiques et cognitifs, dont le processus ne peut pas ou peut difficilement être contrôlé par le sujet parlant. Par contre, l'expression de l'émotion *simulée* est généralement sous le contrôle du sujet, avec un but explicite de communication.

Quant à la terminologie de l'émotion, Williams & Stevens (1972) distinguent les émotions exprimées dans la situation réelle (*real-life emotional situation*) et les émotions imitées par l'acteur (*simulated emotion*) pour comparer leurs différences et similarités expressives. Dumitrache (1994, p26) fait la distinction entre l'émotion *vraie* (*vécue*) et l'émotion *scénique* dans son analyse de l'expression émotionnelle avec des données théâtrales. Dans le présent travail, nous utilisons les termes génériques, l'émotion *vécue* et l'émotion *simulée*, pour désigner l'émotion vraie, provoquée par l'expérience réelle et l'émotion imaginaire, exprimée par l'acteur professionnel ou non-professionnel.

III.3.1. Emotion vécue

Dans l'étude de l'émotion vocale, le corpus d'émotions vécues se constitue des émotions spontanées, produites par la transformation des états d'âme du sujet et enregistrées avec le moins d'intervention possible de l'expérimentateur. Ce genre de corpus reflète les meilleures caractéristiques de l'émotion naturelle mais, dans des expériences phonétiques et psychologiques, son utilisation est assez limitée par des problèmes méthodologiques. L'étude de l'émotion se heurte à un dilemme similaire à celui de l'étude de la parole : le but de l'étude est d'analyser les émotions exprimées dans un milieu naturel mais l'acquisition et l'expérimentation de ce genre de données sont extrêmement difficiles.

C'est ce que Labov (1972, p209) appelle le *paradoxe de l'observateur*[17]. Selon l'expression de Fagyal (1995), l'authenticité des données et le contrôle expérimental sont liés par une fonction inverse. Plus les données de l'émotion naturelle sont authentiques, moins les contrôles sont disponibles à l'expérimentateur sur l'organisation et la manipulation des données.

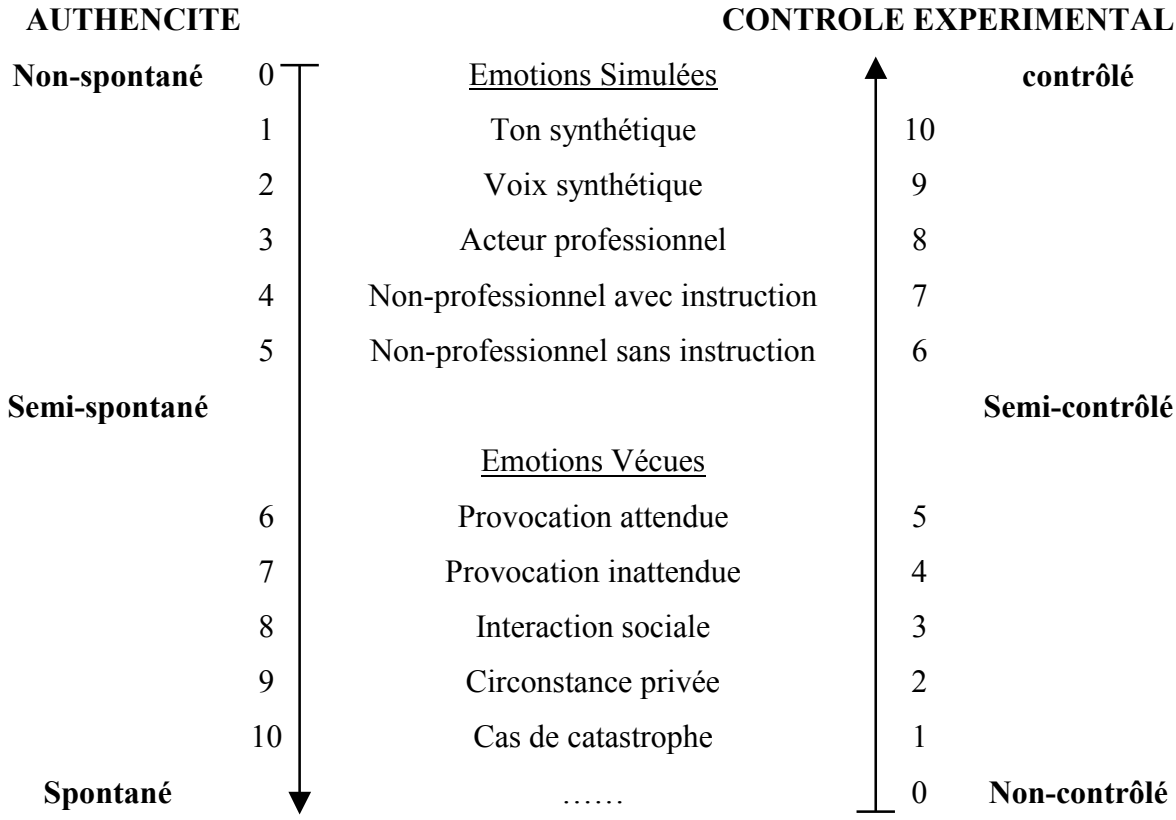
Le problème du contrôle dans l'étude de l'émotion ne se pose pas seulement au niveau méthodologique, mais aussi au niveau expressif. D'abord, il est déjà très problématique d'enregistrer des expressions émotionnelles du sujet en situations naturelles. Même si on y arrive, l'enregistrement est souvent de mauvaise qualité, n'étant pas approprié à des mesures acoustiques (détection de Fo, présentation spectrographique, etc.). De plus, l'interaction des processus psycho-affectifs et des processus cognitif-régulatifs rend difficile de distinguer dans quelle mesure la régulation cognitive influence une expression émotionnelle donnée. Autrement dit, l'expression d'un sujet peut être différente, même contradictoire, de son expérience émotionnelle subjective, parce qu'il peut modifier son expression émotionnelle selon ses règles d'exposition ('*display rules*') et la norme culturelle. Par un exemple, l'excitation émotionnelle chez une personne en colère fait augmenter la fréquence cardiaque et la tension musculaire, son visage devient rouge, sa voix tremble et/ou ses paumes transpirent. La personne arrive pourtant à contrôler son apparence, telle l'expression émotionnelle, pour ne pas montrer ses émotions *brutes*[18] et les transformer en une forme culturellement acceptable. Tomkins (1980, p161) indique que cette régularisation chez les adultes a lieu surtout à travers la suppression de la vocalisation. Scherer *et al*. (1980) expliquent ces deux facteurs biologiques et sociologiques par les notions de l'effet '*push*' interne et de l'effet '*pull*' externe[19].

Les différentes méthodes de construction des données émotionnelles sont présentées dans l'échelle de l'authenticité des données et du contrôle expérimental (Figure 13), dont la structure est empruntée à l'échelle de Fagyal (1995, p46). La même structure est choisie dans le but de faciliter la comparaison entre l'étude de la parole et l'étude de l'émotion vocale. Etant donné que nous étudions les émotions exprimées dans la parole

[17] Voir la partie III.2.1.
[18] Léon (1993, p113) distingue entre l'émotion *brute*, étant le désordre physiologique, et l'émotion *socialisée*, étant une activité dirigée. Ces deux catégories correspondent à ce que les linguistes généralement signifient par l'*émotion* et l'*attitude* (voir II.3.4).
[19] Les facteurs 'push' et 'pull' sont expliqués dans la partie II.4.2.

spontanée dans un cadre phonétique, les problématiques de l'étude de la parole spontanée restent les mêmes dans notre étude et les problèmes de l'étude de l'émotion vécue sont ajoutés aux problèmes phonétiques, donc l'axe d'authenticité et l'axe de contrôle sont encore valables dans l'échelle de la spontanéité des données émotionnelles et de la disponibilité du contrôle expérimental.

Echelle de l'authenticité des émotions et du contrôle expérimental



Le degré de l'authenticité des données (échelle de spontanéité de 1 à 10) est inversement proportionnel au degré du contrôle expérimental (échelle de contrôle de 10 à 1).

Figure 13. Echelle de l'authenticité et du contrôle expérimental des données en cas d'expression émotionnelle, inspirée de Fagyal (1995, p46).

Diverses méthodes sont inventées pour faire un compromis entre l'authenticité des données émotionnelles et le contrôle expérimental. La *provocation attendue*, souvent utilisée par les psychologues, consiste en la provocation des émotions bien définies dans un cadre de laboratoire. L'expérimentateur présente aux sujets des stimuli qui sont censés provoquer certaines réactions émotionnelles (des images, des morceaux de musique ou des textes littéraires) et il enregistre leurs réactions exprimées dans la voix pour comparer la voix émotionnelle avec la voix non-émotionnelle (Skinner, 1935). Cette méthode permet

un grand contrôle sur la production des émotions et une bonne acoustique d'enregistrement mais la provocation des émotions par des moyens artificiels peut être problématique du point de vue psychologique. La *provocation inattendue* de l'émotion se trouve dans un milieu naturel, hors du laboratoire, et dont la procédure est la même que la provocation attendue, sauf que les sujets ne savent pas qu'ils sont enregistrés. Dans l'expérience de Bonner (1943), un signal d'alarme est présenté en plein cours, ce qui rend les étudiants anxieux et stressés. Leurs voix émotionnelles sont enregistrées discrètement et ces expressions naturelles sont comparées avec les expressions des émotions simulées par les mêmes étudiants une semaine après. Le résultat montre qu'il y a des changements prosodiques sous la tension émotionnelle mais la nature des changements varie selon les individus. Par exemple, certaines voix montre l'augmentation du Fo moyen dans l'état anxieux mais d'autres voix montrent la diminution du Fo moyen dans la même situation.

Les émotions exprimées dans des *interactions sociales* sont plus naturelles que les émotions provoquées par des stimuli, reflétant mieux la communication des émotions naturelles. Elles sont produites dans des milieux quotidiens comme la maison, l'école et le lieu de travail, avec une intensité modérée (voir Hutter, 1968 ; Léon, 1970). Les expressions émotionnelles dans une série de conversations entre un employé et son supérieur d'une compagnie d'électricité sont étudiées par Streeter *et al*. (1983). Les auteurs constatent que le Fo moyen et le niveau d'amplitude augmentent en fonction de la difficulté des tâches. Nous plaçons le corpus du présent travail dans cette catégorie d'émotion vécue pour les raisons suivantes. Notre corpus de la parole émotionnelle est constitué à partir d'entretiens télévisés. Six locutrices se sont présentées dans différentes sessions de l'entretien et elles ont conversé avec le présentateur au sujet de leurs problèmes personnels. Malgré le fait qu'elles étaient conscientes d'être télévisées, leurs expressions émotionnelles sont considérées authentiques, à cause de l'urgence de leurs problèmes et de la nature des entretiens concernés.

Les expressions émotionnelles produites dans des *circonstances privées* sont considérées comme les plus authentiques des émotions naturelles, vécues, puisque le sujet se laisse exprimer son émotion telle qu'elle est. La relation intime entre le sujet et l'interlocuteur permet au sujet de s'exprimer librement, sans contrôle excessif. Vu que l'enregistrement de la vie privée est plutôt problématique, ce genre de données n'est guère étudié. Protopapas & Eimas (1997) ont enregistré des pleurs d'enfant et effectué une série

de tests perceptifs avec les stimuli resynthétisés à partir des pleurs originaux. Leurs résultats montrent que les auditeurs distinguent bien les pleurs naturels et les pleurs synthétiques et que le degré perçu de l'émotion négative est lié linéairement aux valeurs élevées du Fo moyen et de la perturbation du Fo ('*jitter*'). Notre corpus a aussi trait à cette catégorie puisque le contenu de la parole de chaque locutrice concerne sa vie privée. De plus, à certains moments de l'entretien, une locutrice fut tellement bouleversée émotionnellement qu'elle ne pouvait plus se contrôler et s'est mise à pleurer.

Parmi les émotions vécues, celles enregistrées *en cas de catastrophe*, (par exemple, la chute de l'avion ou le désastre naturel) montrent une grande spontanéité d'expression émotionnelle. Malheureusement, l'enregistrement dans cette situation est souvent très bruité et ce genre de corpus n'est pas toujours disponible au public. Williams & Stevens (1969) se sont procurés des extraits de voix anxieuse, paniquée, d'un pilote avant la chute de son avion et ont caractérisé cette voix par un Fo moyen élevé et une grande plage de Fo (l'écart entre le Fo maximum et le Fo minimum). Protopapas & Lieberman (1995) ont utilisé le même type de données et ils ont trouvé que le degré du stress émotionnel est corrélé au Fo moyen et au Fo maximum mais pas à la plage de Fo.

En générale, le nombre des études de l'émotion vécue est limité par rapport aux nombreuses études de l'émotion simulée, à cause des difficultés décrites ci-dessus. Cela ne diminue pourtant pas la nécessité ou l'importance de ce genre d'étude ; au contraire, le besoin de connaissance sur l'expression émotionnelle dans des occurrences naturelles devient de plus en plus grand et plus d'analyses expérimentales sur l'expression de l'émotion vécue sont nécessaires pour les développements théoriques et pratiques.

III.3.2. Emotion simulée

Les émotions simulées sont souvent utilisées dans la recherche de l'expression émotionnelle, en raison de la facilité de la manipulation systématique des données. Un des sujets majeurs dans l'étude de l'émotion vocale est de savoir comment les différentes émotions sont identifiées à travers les indices prosodiques, non-verbaux. Pour expérimenter ce genre de question, les chercheurs construisent les données par la simulation des expressions émotionnelles avec l'acteur ou la machine de synthèse vocale. (Fairbanks &

Pronovost, 1938 ; Fónagy & Bérard, 1972 ; Scherer & Oshinsky, 1977 ; Laukkanen *et al*., 1996).

En ce qui concerne la simulation par l'acteur, professionnel ou non, le chercheur détermine d'abord les émotions à étudier, et puis décrit au locuteur le contexte de chaque émotion, qui est censé provoquer une telle émotion. Le locuteur imite les différentes émotions, en modifiant des traits prosodiques (intonatifs) de la voix. Le contenu verbal est contrôlé dans le sens qu'il est soit constant à travers les différentes émotions soit éliminé au moyen de filtrage pour qu'il ne joue pas de rôle dans l'expression ou dans la perception de l'émotion. Les données ainsi acquises facilitent l'analyse comparative, grâce à son organisation systématique. La comparaison porte sur la variation des paramètres prosodiques en fonction des émotions. Ce genre de données est d'un haut degré de contrôle mais peu spontané selon l'échelle de la Figure 13. Vu que l'enregistrement a lieu dans un studio insonorisé, la qualité acoustique est généralement bonne, ce qui est essentiel pour les mesures exactes des traits prosodiques.

De nombreuses études se basent sur les émotions simulées (stylisées) par l'*acteur professionnel* (Dusenberry & Knower, 1939 ; Kramer, 1963b ; Dawkins & Krebs, 1978 ; Fónagy, 1978). La représentativité des émotions théâtrales en tant qu'émotions naturelles est problématique. Cowan (1934, p7) considère la voix actrice comme le type hautement élaboré en anglais dans son analyse des expressions émotionnelles. : « *Since it is generally acknowledged that the artistic speech of actors approaches the highest type of cultivated American speech, material for the study was chosen from the theatre* ». Dumitrache (1994, p26) remarque la spécificité de l'émotion scénique à l'opposition de l'émotion vécue mais en même temps elle conclut que l'effet produit par l'expression vocale de l'émotion scénique est comparable à celui de l'émotion vécue.

Pourtant, d'autres chercheurs montrent que l'expression de l'émotion vécue et l'expression de l'émotion simulée sont différentes des points de vue physiologique et psychologique. Il y a une grande divergence à l'intérieur même des expressions actrices. Les acteurs utilisent différentes techniques pour exprimer des émotions, comme les techniques de Stanislavski, le rappel des expériences personnelles, la modélisation des caractères émotionnels à partir de la comédie télévisée et l'imitation des stéréotypes culturels des émotions (Scherer, 1986, p146), ce qui fait que les expressions émotionnelles

varient largement d'un acteur à l'autre. Duchenne (1862) met en évidence la différence entre l'émotion vécue et l'émotion simulée lors d'une expérience avec des électrodes sur le visage. D'après ses résultats, l'expression du sourire naturel implique des muscles faciaux différents de ceux utilisés pour l'expression du sourire simulé. Les études neuropsychologiques[20] indiquent la différence dans les processus motivationnels entre l'émotion vécue et l'émotion simulée, liée à l'interaction des éléments affectifs et cognitifs dans l'expérience émotionnelle. Murray *et al*. (1995, p74), après avoir passé en revue des problématiques dans la recherche de l'émotion, soulignent la nécessité de l'étude de l'émotion vécue naturelle pour l'amélioration du système de synthèse vocale : « *to ultimately achieve a truly natural-sounding synthetic voice, it must have a good underlying voice quality,… We need improved voice production models, better understanding of the various pragmatic components, and the way they are combined in natural speech* ».

Certains chercheurs de l'expression émotionnelle préfèrent les données d'émotions simulées par des locuteurs *non-professionnels* à celles simulées par des acteurs professionnels (Kaiser, 1962 ; Van Bezooijen, 1984 ; Leinonen *et al*., 1997). Etant conscients des effets spéciaux dus au jeu théâtral dans la voix professionnelle, ils considèrent l'expression émotionnelle non-professionnelle comme plus naturelle que l'expression professionnelle. Quand on demande au locuteur non-professionnel de simuler des émotions *sans instruction* spécifique sur la façon de s'exprimer (voir la Figure 13), il se réfère à ses propres réactions émotionnelles dans le passé, d'où l'authenticité de son expression émotionnelle. La personne peut aussi emprunter les expressions stéréotypées dans la culture mais cela ne détruit pas l'authenticité de ses émotions puisque l'influence culturelle fait aussi partie de l'émotion naturelle. Quand on donne au locuteur des *instructions* sur le contexte émotionnel et la façon de s'exprimer, la structure des données est plus contrôlée mais la spontanéité des données est considérablement diminuée.

La simulation des émotions peut être aussi faite par la machine, le synthétiseur vocal. On crée la *voix synthétique* émotionnelle, en modifiant les paramètres de Fo, d'amplitude et de durée, de la voix naturelle non-émotionnelle. La validité de l'expression émotionnelle dans la voix synthétique est évalué par un test de perception avec des auditeurs naïfs (Lieberman & Michael, 1962 ; Ladd *et al*., 1985 ; Mozziconacci & Hermes,

[20] Voir la partie II.3.

1997). Les systèmes de la synthèse de la voix émotionnelle (Murray, 1989 ; Cahn, 1990) sont fondés sur la même procédure, appelée l'analyse par synthèse, excepté le système de Murray qui se base sur la voix synthétique dès le début.

Scherer (1974) et Scherer & Oshinsky (1977) poussent plus loin l'analyse de la communication non-verbale de l'émotion avec les données de *ton synthétique*. Afin d'étudier comment l'auditeur attribue aux valeurs acoustiques différentes significations émotionnelles, ils synthétisent des tons en modifiant les paramètres acoustiques comme le niveau et la variation de Fo, le niveau et la variation d'amplitude et le rythme. Les stimuli synthétiques sont évalués par les auditeurs sur trois dimensions psychologiques, telles que activation, puissance et valence. Les résultats de leurs expériences montrent que les auditeurs utilisent des indices acoustiques de façon systématique et linéaire pour identifier les différentes significations émotionnelles, d'où le modèle linéaire de l'utilisation des indices dans les jugements ('*linear model of the judges' cue utilization*').

Une des expériences dans le présent travail utilise également des stimuli synthétiques. L'expérience vise à savoir comment le contour de Fo et la durée déterminent la perception de l'émotion positive et l'émotion négative et examine la question avec les stimuli construits par la modification des paramètres au moyen de la synthèse. A partir d'une syllabe en voix naturelle, quatre types de contour de Fo et deux types de la durée sont synthétisés et évalués par les auditeurs de différentes langues maternelles. Le résultat indique un effet significatif de cette modification sur la perception émotionnelle. Les lecteurs sont invités à se référer au chapitre VI pour une description plus détaillée.

III.4. Conclusion du chapitre III.

Dans ce chapitre, nous avons discuté de l'authenticité des données et du contrôle expérimental dans l'analyse de l'émotion vocale. Deux contrastes sont faits sur le critère de spontanéité, l'un entre la parole spontanée et la parole lue, et l'autre entre l'émotion vécue et l'émotion simulée. Nous inspirant de l'échelle de Fagyal (1995), nous avons illustré l'échelle de l'authenticité des données et du contrôle expérimental pour la parole émotionnelle. Les différentes méthodes pour l'expérience phonétique de la parole émotionnelle sont présentées et expliquées en terme de relation antagoniste entre la spontanéité des données et le contrôle expérimental. La relation révèle que plus les données sont authentiques, moins elles sont contrôlées du point de vue expérimental. Pourtant, l'authenticité maximale des données ne signifie pas nécessairement l'absence totale de contrôle expérimental, si les données sont basées sur certains critères fonctionnellement définis. C'est ce que nous avons essayé de faire lors de nos analyses expérimentales sur la parole spontanée émotionnelle, présentées dans les chapitres suivants.

Nous avons aussi discuté, dans ce chapitre, d'un autre type de contrôle dans l'expression émotionnelle, le contrôle cognitif-régulatif. Etant donné que l'émotion est le résultat de l'interaction des éléments affectifs, cognitifs, et physiologiques, l'expression émotionnelle d'une personne varie en fonction de ses règles d'exposition et la norme socioculturelle. Son expression émotionnelle peut être différente, même contradictoire, par rapport à son propre sentiment émotionnel. D'après les diverses études psycho-phonétiques, les traits prosodiques sont trouvés en tant que porteurs principaux de l'émotion vocale et ils sont susceptibles d'être modifiés par le contrôle régulatif de la conscience du sujet. Dans l'analyse prosodique de l'émotion vocale, il est souvent difficile de déterminer dans quelle mesure les changements prosodiques de la voix émotionnelle (par rapport à la voix neutre) sont dus à l'effet direct de l'expérience subjective de l'émotion, et dans quelle mesure ils sont modifiés de manière artificielle par le contrôle conscient du sujet.

Ces difficultés de contrôle dirigent souvent l'étude de l'émotion vers l'utilisation des données d'émotion simulée au lieu des données d'émotion vécue. La remarque de

Williams & Stevens (1972, p1248) est fréquemment citée pour justifier cette utilisation : « *The comparative data and some additional limited data obtained from real-life emotional siutations are not inconsistent with the data obtained from the actors in this study…* ». Ce n'est pas étonnant de constater la comparabilité entre l'émotion vécue et l'émotion simulée vu que la dernière est censée représenter la première, mais cela n'indique pas nécessairement que les données d'émotions simulées peuvent remplacer celles d'émotions vécues. Les émotions simulées par des acteurs sont généralement stylisées, tandis que les émotions exprimées dans nos interactions sociales normales ont trait à une intensité relativement basse en forme discrète.

Il faut noter que la spontanéité des données ne signifie pas nécessairement l'impossibilité du contrôle dans l'étude expérimentale. Il est vrai que les données de la parole spontanée ou de l'émotion vécue excluent la possibilité d'avoir les mêmes phrases à travers différentes conditions, ce qui est faisable avec la parole lue ou l'émotion simulée. Mais la comparaison paradigmatique des variations dans la parole ou dans les émotions est toujours possible avec les données de nature spontanée, à travers le contrôle systématique des facteurs socio-situationnels et psychologiques dans la sélection des données et la vérification expérimentale par un test de perception. Le chapitre suivant est consacré à décrire comment ce genre de contrôle fut effectué dans l'établissement de notre corpus.

L'expression de l'émotion vécue, un phénomène à la fois riche et complexe, intéresse de plus en plus de chercheurs. Actuellement, il s'agit de l'un des sujets à la mode dans la recherche de l'intelligence artificielle. Plus de connaissances dans le domaine des émotions exprimées dans des circonstances naturelles aideront à développer une théorie de l'émotion et favoriser des applications dans le domaine de la technologie vocale. Le présent travail vise à répondre à ce besoin scientifique, en traitant les données d'émotion vécue dans la parole spontanée en coréen et en anglais. Ce chapitre III a fournit un cadre théorique pour expliquer la nature du corpus et le choix de notre expérimentation avec ce genre de corpus. Ces derniers seront exposés dans le chapitre suivant. L'authenticité de nos données et les problèmes de contrôle dans nos expériences seront expliqués à la lumière des notions introduites dans ce chapitre.

Chapitre IV

Etude sur l'expression et la perception de d'émotion dans la parole spontanée en coréen

Résumé

Ce chapitre présente notre étude principale du corpus coréen. Les problématiques étant exposées dans IV.1, la procédure de l'acquisition des données et les critères de sélection des énoncés à analyser sont décrits du point de vue typologique fonctionnel dans la partie IV.2. Le corpus coréen consiste en des extraits de 40 minutes du discours spontané d'une locutrice Coréenne. 110 énoncés ont été choisis d'une façon chronologique pour nos analyses acoustiques et perceptives. Dans IV.3, deux tests de perception sont présentés ; chaque test montrant comment les émotions sont exprimées au niveau prosodique (indice vocal) et au niveau lexical (indice sémantique) des énoncés. Les expressions prosodique et lexicale de l'émotion sont décrites en termes d'intensité émotionnelle (valeur d'activation) et de positivité émotionnelle (valeur de valence). Dans IV.4, les mesures acoustiques des énoncés sont présentées : Fo moyen, Fo maximum, Fo minimum, moyenne des 20% des valeurs les plus basses de Fo ('*Fo Moy Bas*'), plage de Fo, perturbation de Fo ('*jitter*'), perturbation d'intensité ('*shimmer*') et débit de parole. L'excitation émotionnelle de la joie est repérée surtout par l'augmentation du Fo moyen tandis que l'excitation émotionnelle de la tristesse (détresse accompagnée de pleurs) est mieux repérée par la diminution du Fo minimum et du 'Fo Moy Bas'. L'augmentation du Fo maximum et de la plage de Fo est un bon indice de l'excitation émotionnelle générale (soit la joie, soit la tristesse). Les variations du jitter, du shimmer et du débit dues à l'émotion ne sont pas significatives dans notre corpus coréen. Dans la partie IV.5, deux analyses perceptives sont rapportées. La première expérience montre que les indices fréquentiels (comme la plage de Fo) sont plus fortement corrélés à la perception de l'émotion (estimée par la valeur d'activation et la valeur de valence) que les indices temporel et d'intensité. La deuxième expérience a examiné la perception de l'émotion par les Coréens, Français et Américains. Dans le test, ils ont tous identifié la joie et la tristesse de la locutrice Coréenne avec une précision supérieure à celle qui aurait été due au hasard. Cependant, les Coréens étaient significativement plus précis que les Français et les Américains. Ces résultats confirment à la fois l'universalité de l'émotion et l'influence de la connaissance culturelle sur la perception de l'émotion. Dans la partie IV.6, un autre test de perception a été effectué pour savoir si l'émotion est mieux exprimée et mieux reconnue dans certaines parties de l'énoncé que d'autres. Dans le test, 15 énoncés de trois émotions (joie, neutre, et tristesse) ont été divisés en trois parties, initiale, médiane et finale, et leur impression émotionnelle a été évaluée par trente auditeurs coréens, américains et français. Le test a permis de mettre à jour le fait que l'émotion est mieux exprimée et mieux reconnue dans la partie finale de l'énoncé que dans les parties, initiale et médiane. La découverte d'une répartition non uniforme des indices acoustiques de l'émotion tout au long de l'énoncé constitue un aspect original de cette étude. Ce résultat fut mis à l'épreuve et confirmé par d'autres expériences dans les chapitres suivants.

IV. Etude sur l'expression et la perception de l'émotion dans la parole spontanée en coréen

IV.1. Préliminaires

Le chapitre IV présente notre apport personnel, qui est un ensemble d'analyses acoustiques et perceptives sur un corpus coréen. Le but est d'étudier comment l'état émotionnel du locuteur est exprimé dans la parole spontanée et comment l'auditeur identifie l'émotion positive et l'émotion négative à travers les traits prosodiques. Le corpus coréen consiste en des extraits du discours spontané produit par une locutrice Coréenne, où sont exprimées les émotions de la tristesse et de la joie.

Le but de notre étude peut être reformulé en trois tâches d'analyse : (1) identifier les changements acoustiques de la voix émotionnelle, par rapport à la voix neutre; (2) examiner la perception de l'émotion par les auditeurs de culture différente ; et (3) rechercher si la communication de l'émotion vocale varie en fonction des différentes parties de l'énoncé, telles les parties initiale, médiane et finale. La dernière tâche concerne une nouvelle problématique dans ce domaine, constituant un des aspects originaux de cette thèse. Cette problématique sera poursuivie par la suite avec un corpus anglais dans le chapitre V.

Le chapitre IV se compose de cinq parties principales. Les deux premières parties, IV.2 et IV.3, décrivent la construction et les caractéristiques de notre corpus coréen. La partie suivante, IV.4, est consacrée aux mesures acoustiques concernant la fréquence fondamentale, la durée, l'intensité et le spectre. La partie IV.5 examine la perception de l'émotion exprimée en coréen par des auditeurs coréens, français et américains. La perception de l'émotion en fonction des différentes parties de l'énoncé, telles les parties initiale, médiane et finale, est examinée dans la partie IV.6. Enfin, les résultats de chaque partie sont résumés et discutés dans la conclusion du chapitre IV.

IV.2. Sur le corpus coréen

Le corpus[21] se compose de huit échantillons de parole d'une locutrice Coréenne 'WJ,' extraits à partir de l'enregistrement d'entretien télévisé. L'entretien est tiré d'une émission de télévision coréenne, KBS (*Korean Broadcast Station*), intitulée '*Une rencontre dans la matinée*[22]'. Dans cette émission, la locutrice est invitée à parler de ses problèmes personnels dans le but de trouver, au moyen de la diffusion télévisée, une aide ou des informations nécessaires pour résoudre ses problèmes.

IV.2.1. Acquisition des données

L'entretien a été enregistré dans un des studios de KBS et diffusé en direct à la télévision. La copie de l'entretien est disponible au public moyennant une certaine somme, sous forme d'une vidéocassette VHS. Bien que nous n'ayons pas de précision sur l'appareil d'enregistrement utilisé dans le studio de KBS, le spectrogramme du signal de parole révèle une bonne qualité acoustique de l'enregistrement (voir Figure 32). Quelques bruits extérieurs comme bruits des mouvements corporels du sujet parlant et applaudissement de l'auditoire, sont inclus dans l'enregistrement puisqu'il s'agit d'une émission en direct, mais le bruit de fond est négligeable.

Lorsqu'il s'agit de l'étude de l'émotion produite dans une situation naturelle, il est extrêmement difficile d'avoir plusieurs sujets, qui ont la même identité socioculturelle (d'âge, de sexe, d'origine régionale) et produisent les mêmes types d'expression émotionnelle dans la même situation conversationnelle. Pour le cas idéal, les sujets doivent être, entre autres, d'âge, de sexe, d'origine socioculturelle, d'expression émotionnelle comparables, et ils doivent être enregistrés dans des situations de communication similaires. Dans cette étude, nous étudions une seule locutrice dont l'identité socioculturelle est clairement déterminée. Le choix de la locutrice, au lieu du locuteur, n'était pas conçu en début d'étude mais a été décidé après l'observation des entretiens à

[21] Le *corpus* est un ensemble de données, écrits ou oralisés, recueillis pour un sujet d'étude. Les *données* sont définies comme un ensemble d'exemples, constituant les bases d'un problème, selon le dictionnaire français LAROUSSE (édition de 1979). Dans cette présentation, les deux termes sont utilisés avec le même sens, sauf que le terme *corpus* connote souvent un sens plus collectif.

notre disposition. L'entretien de la locutrice WJ a été choisi puisque ses émotions étaient exprimées de manière explicite pendant l'entretien, parmi les sujets dans d'autres entretiens. A propos du risque de la portée limitée des résultats avec les données de la source monotone, Fagyal (1995, p57) cite une remarque d'un biologiste : *"Evidence obtained in this way is never conclusive, though it may be usefully suggestive."* (Beveridge, 1950, p28). Cette étude vise à suggérer de telles preuves pertinentes sur l'expression émotionnelle dans la parole spontanée et à expérimenter la perception de ce genre d'émotion par les auditeurs provenant de différentes cultures.

IV.2.2. Typologie des facteurs concernés

La construction de notre corpus se base sur une approche typologique fonctionnelle. Le contrôle direct de la production des données étant exclu à cause de sa nature spontanée, le contrôle expérimental de notre corpus s'est imposé au niveau de la sélection des données. Notre corpus est choisi dans la considération des facteurs suivants[23] :

Facteurs situationnels : mode de production (lu ou non lu), type de transmission (médiatisé ou non, en direct ou non), sujet de discussion (formel ou informel), et but de la communication (demande d'information, persuasion, et artistique).

Facteurs sociaux : sexe, âge, profession et milieu d'origine.

Facteurs psychologiques : état émotionnel et attitude.

En ce qui concerne les facteurs situationnels, notre corpus peut être identifié comme du discours non lu, spontané (**mode production**), médiatisé en direct (**type de transmission**) au moyen de la télévision. Le présentateur a interrogé la locutrice, de manière improvisée, sur son enfance, ses relations familiales, et ses problèmes actuels. La parole étant produite sans définition préalable ou répétition artificielle, nous avons considéré l'émotion exprimée dans la parole de locutrice comme l'émotion vécue[24]. Le

[22] Le nom coréen de cette émission est [atSimmadaN].

[23] On trouve un excellent aperçu des facteurs situationnels, sociaux et psychologiques du discours spontané dans Fagyal (1995, p62-73). Cette dernière s'intéresse surtout aux facteurs sociaux et situationnels dans son étude de la variation phonosylistique en fonction de l'âge et de la situation conversationnelle, tandis que notre étude de la parole émotionnelle concerne principalement les facteurs psychologiques.

[24] Voir III.3. pour la distinction de l'émotion vécue et de l'émotion simulée.

comportement verbal de la locutrice a pu être influencé par la présence ou l'absence du microphone (Fagyal, 1995, p66). Il est aussi bien connu que les gens ont tendance à se retenir d'exprimer leurs émotions négatives quand ils sont conscients d'être enregistrés pour une transmission de masse (télévision ou radio). Cette considération nous a conduite à choisir l'entretien dans lesquels l'émotion du sujet est suffisamment forte pour être exprimée d'une manière explicite (voir plus loin).

Le **sujet de discussion** de l'entretien de notre corpus est informel, il concerne les souvenirs d'enfance, la situation familiale et les problèmes personnels. Le présentateur s'est adressé à la locutrice de manière informelle, en essayant de créer une atmosphère détendue. Pourtant, aucun entretien n'est entièrement informel, en raison de la nature de l'enregistrement télévisé. Le **but de communication** de l'entretien peut être caractérisé comme une demande d'information ou de conseil au public. La locutrice WJ cherche un conseil, un moyen de faire comprendre à ses parents qu'ils font erreur en faisant objection à son mariage. Le contenu de son discours[25] et les facteurs sociaux de son identité personnelle sont présentés plus en détail dans IV.2.3.

En ce qui concerne les facteurs psychologiques, **trois types d'émotion**, positif, neutre et négatif, sont exprimés dans notre corpus. En termes de sélection des données, l'expression de l'émotion positive fut extraite de moments où la locutrice était joyeuse et celle de l'émotion négative à partir des moments où elle était triste. Nous avons identifié les émotions de la locutrice, en prenant en compte les deux étapes suivantes. D'abord, les états émotionnels de la locutrice sont inférés à partir de l'observation des expressions faciales et du contexte conversationnel. Par exemple, les pleurs ou les sourires sont utilisés en tant que repères dans l'identification de l'émotion positive et de l'émotion négative[26]. L'expression de l'émotion neutre fut extraite de moments où la locutrice parlaient des choses factuelles sans émotion particulière. L'émotion neutre ici indique un état émotionnel non-marqué[27], qui peut servir à une référence de base pour décrire des états émotionnels marqués comme la joie et la tristesse. Ensuite, notre identification des émotions, positive,

[25] La transcription du discours de la locutrice se trouve dans l'annexe.

[26] Le sourire et les pleurs peuvent être produits pour exprimer l'ironie ou pour tromper l'interlocuteur dans l'interaction sociale, mais ce genre de problème n'est pas traité dans la présente étude puisque les expressions émotionnelles de nos locutrices sont considérées comme directes et sincères par rapport à ce qu'elles ressentent

[27] Etant donné que nous considère l'émotion comme omniprésente dans la parole naturelle, suivant différents degrés d'intensité et de positivité émotionnelles, l'émotion neutre correspond à un état d'âme avec le moindre degré d'intensité et de positivité émotionnelles.

neutre et négative, est vérifiée par un jugement de dix auditeurs coréens dans une expérience perceptive. Ces démarches pour identifier trois types d'émotion de notre corpus sont présentées dans les parties IV.2.4 et I.1.1. L'**attitude** de la locutrice WJ, un autre facteur psychologique de notre corpus, peut être caractérisée comme ouverte et communicative. Elle voulait exposer des informations personnelles autant que possible pendant l'entretien, afin que sa situation personnelle soit bien comprise par l'interlocuteur et l'auditoire.

IV.2.3. Locutrice

La locutrice WJ est une célibataire de vingt-neuf ans. Elle travaille dans une boîte d'informatique en tant que secrétaire. Elle parle le dialecte Séoulien, le coréen standard. Elle raconte ses troubles relationnels avec sa famille. Elle aime une personne qui ne plaît pas à ses parents pour des raisons inconnues et ses parents la forcent à quitter cette personne. Ses soeurs et sa tante prennent parti pour ses parents contre le mariage, en mettant la locutrice en situation encore plus difficile. Au moment de l'entretien, elle était rejetée par sa famille. La locutrice s'est présentée dans l'émission dans le but de trouver un conseil pratique pour résoudre son problème familial et éventuellement se faire comprendre par les membres de sa famille, grâce à cette émission télévisée. Divers aspects de sa vie sont racontés dans l'entretien, y compris des souvenirs d'enfance, sa relation avec les membres de la famille, les expériences amoureuses avec la personne et les conflits familiaux en ce qui concerne son mariage. L'enfance et l'amour de la locutrice, dont la locutrice était contente de parler, ont été retracés au début de l'entretien. En se rappelant les meilleurs moments de sa vie, la locutrice était gaie et souriante. Avant de présenter son problème familial, le présentateur et la locutrice ont partagé leurs points de vue sur la vie en général et la locutrice s'exprimait sans émotion particulière. Le problème de la locutrice a été présenté à travers une série de question-réponses. En se rappelant les moments de conflit et d'opposition, elle est devenue triste parce qu'elle était désolée de détruire la paix de sa famille à cause de son mariage. Au fur à mesure qu'elle détaillait le problème en question, son sentiment de tristesse s'est renforcé. A partir d'un certain moment, ses mots ont commencé à se mêler aux larmes et elle a fini par pleurer et à parler en sanglotant.

Il faut noter que la tristesse est exprimée d'une manière plus explicite que celle de la joie dans notre corpus. Cela est dû au fait suivant. Les larmes (considérées comme l'indice de la tristesse dans la présente étude) sont souvent refoulées et on évite de les montrer dans un certain milieu public, donc l'apparition des larmes signale que la force émotionnelle l'a emporté sur la force du contrôle cognitif, c'est-à-dire, une grande intensité émotionnelle du sujet parlant. Or, cela n'est pas nécessairement le cas pour l'expression du sourire. Le sourire (considéré comme l'indice de la joie dans la présente étude) peut être exprimé plus facilement dans le milieu public, avec moins de contrôle régulatif cognitif.

Rappelons l'échelle de l'authenticité et du contrôle expérimental pour les données de parole et celle pour les données d'expression émotionnelle dans le chapitre III. Dans la première échelle pour les données phonétiques, notre corpus correspond à un niveau entre '**conversation dirigée**' et '**rappel de mémoire**,' dont le degré d'authenticité se trouve entre 11 et 12 (sur 12) et le degré de contrôle expérimental entre 1 et 2 (sur 12). Dans la deuxième échelle pour les données émotionnelles, notre corpus consiste en '**émotions vécues**' produites en '**interaction sociale**,' dont le degré d'authenticité est de 8 (sur 10) et le degré de contrôle expérimental de 3 (sur 10). La grande authenticité est un des aspects avantageux de notre corpus. Ce genre de corpus est censé refléter la réalité de parole émotionnelle en tant que telle, sans manipulation artificielle. Cependant, peu de contrôle dans la production des données entraîne des inconvénients dans l'expérimentation de ces données. Par exemple, la tristesse exprimée dans notre corpus, n'est pas une émotion pure, mais une émotion complexe, faisant intervenir différents facteurs motivationnels de la locutrice. Dans les pleurs de la locutrice sont mêlés la tristesse due à la séparation, le chagrin éprouvé envers les aimés, la détresse due aux tentatives veines pour trouver des solutions, et la pitié de soi dans une telle situation. Etant donné que l'expression de l'émotion vécue a été toujours étudiée avec les données de la voix d'homme[28], notre corpus avec la voix de femme apportera un nouvel aspect dans ce domaine de travail.

[28] Par exemple, Williams & Stevens, 1969 ; Léon, 1970 ; Streeter *et al*.,1983 ; Protopapas & Lieberman, 1995.

IV.2.4. Sélection des données

Afin de limiter les données à analyser à huite minutes, nous avons prélevé une minute de parole[29] toutes les cinq minutes à partir de 40 minuites d'entregistrement de l'entretien (Figure 14). Vu la transition émotionnelle de la locutrice, d'un état positif vers un état négatif, le prélèvement de cette manière nous a rendu les trois types d'expression émotionnelle. Ainsi, huit extraits sont obtenus en tant que données à analyser, et enregistrés sur une cassette de DAT ('*Digital Audio Tape*') à partir de la vidéo-cassette. La vidéo à quatre-tête de la marque *HITACHI (DA4, MA423)* et le magnétophone de DAT de la marque *MARANTZ* sont utilisés pour cet enregistrement.

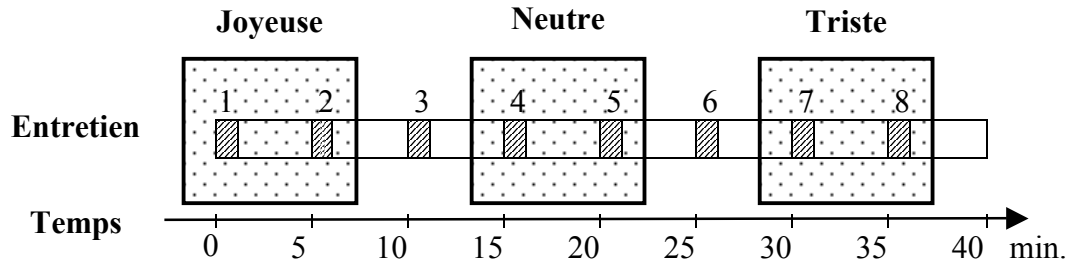


Figure 14. Prélèvement de 24 extraits de parole (boîtes rayées) de la locutrice WJ, à partir des moments où elle est joyeuse, neutre et triste (boîtes pointillées).

A partir des extraits enregistrés sur la cassette de DAT, nous avons échantillonné le signal de parole à 22kHz[30] sur un disque informatique (*PC pentium I*), au moyen du logiciel 'Mev'[31]. La parole du présentateur, les bruits extérieurs et le silence entre les discours sont exclus, dans la mesure possible, de cet enregistrement. La durée de parole échantillonnée varie de 21,6 à 55,8 secondes et la durée totale des extraits est d'environ 5 minutes (voir Tableau 6). La parole des extraits est segmentée en énoncés selon les règles

[29] Nous avons prélevé, en principe, une minute de parole au début de chaque cinq minutes. Pourtant, si la parole de la locutrice n'est guère présente au début mais dans une autre des autres quatre minutes pendant les cinq minutes en question, le morceau d'une minute, qui contient le plus de parole de la locutrice, est choisi pour notre corpus.

[30] L'échantillonage économique pour les données de parole, fait dans la plupart des études phonétiques, est de 10kHz, 12kHz puisque les différents segments de la parole peuvent être distingués avec l'information du signal au-dessus de 5kHz. Etant intéressés à la voix émotionnelle, nous avons choisi un échantillonage plus détaillé, de 22kHz, pour rassurer l'analyse de la qualité de voix, peu traitée dans la présente étude mais conçue pour l'étude ultérieure.

[31] Le logiciel 'Mev' est dévélopé par Mertus (1992) dans le département des sciences cognitives et linguistiques à l'université de Brown. Ce logiciel est similaire à d'autres logiciels commercialisés pour le traitement de la parole, ayant les fonctions d'édition du son, d'extraction de la fréquence fondamentale, de spectrograme et de filtrage.

explicitées dans la partie suivante (IV.2.5). L'expression faciale de la locutrice (l'image vidéo) fut numérisée dans d'autres fichiers informatiques, au moyen du logiciel 'Adobepro,' dont quelques exemples sont présentés dans l'annexe.

IV.2.5. Segmentation des énoncés

En ce qui concerne l'unité de segmentation, divers termes ont été proposés avec leurs propres critères de segmentation, tels que l'*énoncé* (*utterance*, Streeter *et al*., 1983 ; Nakajima & Allen, 1993), la *proposition* (*clause*, Menn & Boyce, 1982), la *suite sonore* (Duez, 1991), et la *séquence sonore* (Fagyal, 1995). Le discours de la locutrice WJ dans notre corpus est segmentée en **énoncés** (*utterances* en anglais) pour nos analyses acoustiques et perceptives. L'énoncé est une séquence des mots, proférés par un locuteur dans une prise de parole[32]. Il est considéré comme l'unité d'information discursive dans l'étude pragmatique et il est déterminé des marques prosodiques (comme les changements de Fo et d'intensité et la présence de pauses silencieuses ou sonores) dans l'analyse acoustique. La construction et l'interprétation de l'énoncé se basent non seulement sur les informations lexicales et grammaticales, mais aussi sur les traits prosodiques comme l'intonation, le rythme et la variation d'intensité. De ce fait, la frontière de l'énoncé ne correspond pas nécessairement à la frontière de l'unité syntaxique, à la différence du syntagme ou de la phrase. Selon l'expression de Lehiste (1970, p154-156), l'*unité phonologique*[33] (tel l'énoncé), qui est l'*unité de performance*[34], n'est pas dans une relation univoque avec l'*unité morpho-syntaxique* (telle la phrase), qui est l'*unité de compétence*. L'énoncé peut consister en un mot ou plusieurs phrases, et inversement, une phrase peut être coupée en plusieurs énoncés par des pauses silencieuses ou sonores. La taille de l'énoncé dépend de la nature du discours, du débit du locuteur, de l'état psychologique du locuteur et de la situation du discours. Nakajima & Allen (1993, p198-199) proposent quatre principes de déterminer l'énoncé dans le discours dialogique :

[32] La *prise de parole* est un passage monologique du locuteur à l'intérieur de l'unité *dialogue* (Fagyal, 1995, p76).
[33] Son terme 'phonologique' peut être interprété comme 'phonétique' dans ce contexte puisqu'il s'agit de la réalistion concrète du son de la langue (*performance linguistique*).
[34] Les termes, *performance* et *compétence*, ont l'origine de la théorie générative transformmelle. Pourtant, l'aspect de performance est souvent ignoré dans cette théorie, au prix de son intérêt principal à l'aspect de compétence, tel la grammaire universelle.

(1) **Principe pragmatique** : L'énoncé correspond à l'acte de parole. C'est-à-dire, il représente l'intention de base du locuteur.

(2) **Principe conversationnel** : Placer une frontière de l'énoncé à la frontière de la prise de parole (changement de locuteur).

(3) **Principe grammatical** : La frontière de la phrase syntaxique est un point virtuel de la frontière de l'énoncé.

(4) **Principe prosodique** : Placer une frontière de l'énoncé avant une pause dont la durée est plus longue que le seuil de coupure.

La segmentation de notre corpus se base principalement sur les principes cités au-dessus. L'acte de parole (ex. question, réponse, confirmation, etc.) de la locutrice est identifié d'abord. Ensuit, la prise de parole de la locutrice est identifiée, dont la frontière souvent correspond à celle de l'acte de parole. Le principe grammatical et le principe prosodique sont pris en compte simultanément à la décision de la frontière de l'énoncé. La frontière de la phrase syntaxique est devenue la frontière de l'énoncé, à moins que deux phrases soient liées avec une pause silencieuse moins de 100ms. Quand il y a une pause plus longue que le seuil de coupure à l'intérieur de la phrase, les frontières de l'énoncé sont placées avant et après la pause (voir Figure 15). Le seuil de coupure est déterminé selon le débit de parole de la locutrice WJ. Il est calculé comme deux fois de la longueur moyenne de la syllabe de la locutrice[35], étant de 312,5ms.

La longueur du seuil de coupure varie sensiblement selon les études. Notre seuil est plus petit que le seuil de Nakajima & Allen (1993), qui est de 750ms, mais plus grand que d'autres seuils proposés. Selon le compte-rendu de Fagyal (1995, p81), le seuil de 200ms à 300ms paraît le plus fréquent : les seuils de 200ms, de 250ms et de 270 ms sont utilisés dans les études de Goldman-Eisler (1968), de Grosjean & Deschamps (1973) et de Kowal *et al*. (1985) respectivement. Le seuil de Barik (1975) est plus long que les autres (610ms), tandis que le seuil de Hieke *et al*. (1983) n'est que de 130ms. Cette diversité de seuils est due surtout à la différente nature des discours utilisés dans les analyses, d'où la difficulté de comparer les études.

[35] Afin d'obtenir la longueur moyenne de la syllabe, 15 phrases relativement faciles à repérer sont prises, de manière aléatoire, à partir des extraits échantillonnés. Le débit moyen de la locutrice (le nombre des syllabes par seconde) est d'abord calculé, puis la durée moyenne de la syllabe est estimée par le convertissement du débit moyen en millisecondes.

Les pauses peuvent être sonores, tels que l'hésitation sonore (ex. [E]), l'allongement vocalique et les faux-départs (reprise répétitive d'un même segment). Les clichés personnels[36] (ex. [glNg'a][37] 'donc', [cEgi] 'là' et [tege] 'très') sont inclus à l'énoncé sauf qu'ils sont séparés de l'énoncé par une pause plus longue que le seuil de coupure (voir Figure 16).

(1) { NP AdvP NP VP }
(2) [op'aga tSclm // Emmaril pwengejo]
(3) 323
(4) // **WJ42a** (780ms)------- // **WJ42b** (1013ms)----------//
(5) // lui, la première fois // mère, a rencontré //
(6) 'Il a rencontré ma mère pour la première fois.'

Figure 15. Exemple de la division des énoncés (**WJ42a** et **WJ42b**) par une pause plus longue que le seuil de coupure ('//' indique la frontière de l'énoncé).

(1) Etiquetage: NP (syntagme nominal), AdvP (syntagme d'adverbe),VP (syntagme verbal), PS (pause sonore), A (allongement vocalique).
** Le symbole '{ }' indique l'unité de la phrase syntaxique.
(2) Transcription phonétique des énoncés coréens.
(3) Durées des pauses silencieuses et de la pause sonore (ms).
** le symbole '%' indique la prise de souffle.
(4) Les énoncés **WJ42a** et **WJ42b** et leurs durées (ms).
(5) Correspondance avec la traduction française
(6) Traduction française.

A la suite de notre segmentation, 110 énoncés de la locutrice WJ sont obtenus[38], dont la durée moyenne (écart-type) est de 1762,3 (268,9). Le nombre des énoncés et leur durée moyenne dans chaque extrait du discours sont présentés dans le Tableau 6. Ces énoncés sont encore segmentés en syllabes et en phonèmes. Vu le désaccord entre l'unité orthographique et l'unité orale, l'effet de resyllabation et l'effet de coarticulation sont pris en compte dans la segmentation. Il faut préciser que le comptage syllabique de notre analyse se base sur la réalisation phonétique et non sur la représentation phonémique.

[36] Les *clichés personnels* sont des mots répétitifs dans le discours, prononcés par l'habitude du locuteur. Ils sont purement stylistiques et sémantiquement vides (sans signification lexicale).
[37] Le mot d'exemple coréen est transcrit en alphabet international phonétique (API) entre crochets.
[38] La transcription orthographique (en coréen) et la transcription phonétique (en API) des énoncés se trouvent dans l'annexe.

(1) NP VP CP PS (FD) AdvP NP AdvP VP IL
(2) [modInire jEIZENZEgigu gINg'a cigimin // cigIm saNhwaNi dege ErjEunig'anjo // je//]
(3) 0 121 411
(4) // **WJ65** (2093ms) ---------- // **WJ66** (1717ms) ---------- //
(5) toutes les choses, (il) fait son mieux, maintenant // maintenant, la situation, très, est difficile // oui //
(6) 'Il fait son mieux pour toutes les choses' // 'Maintenant, la situation est très difficile' //

Figure 16. Exemple de la segmentation des énoncés, dont la frontière est indiquée par '//'.

(1) Etiquetage: NP (syntagme nominal), VP (syntagme verbal), CP (cliché personnel), PS (pause sonore), FD (faux départ), AdvP (syntagme d'adverbe), IL (intervention de l'interlocuteur).
(2) Transcription phonétique des énoncés coréens.
(3) Durées des pauses silencieuses (ms) : le symbole '%' indique la prise de souffle.
(4) Segmentation des énoncés **WJ65** et **WJ66**.
(5) Correspondance avec la traduction française
(6) Traduction française.

Extraits	**Durée (s) de l'extrait**	**Durée moyenne (ms) de l'énoncé**	**Nombre d'énoncés**
WJ1	27,3	1776,9	9
WJ2	21,6	2158,1	8
WJ3	34,2	1973,4	9
WJ4	29,1	1588,6	12
WJ5	49,3	1934,7	19
WJ6	30,4	1803,8	10
WJ7	26,6	1593,1	13
WJ8	55,8	1515,9	30
Moyenne (ET)	34,3 (11,9)	1762,3 (268,9)	13,6 (7,5)
Total	274,3		110

Tableau 6. Durée de l'extrait, la durée moyenne (ms) de l'énoncé à l'intérieur de l'extrait et le nombre des énoncés compris dans l'extrait du discours de la locutrice WJ.

IV.3. Analyse descriptive

La partie IV.3 présente deux expériences perceptives, qui décrivent comment les émotions de la locutrice WJ s'expriment au niveau prosodique et au niveau sémantique des énoncés. Vu la nature spontanée de notre corpus, les différents états émotionnels de la locutrice ne peuvent être identifiés que par l'observation de son visage et de sa voix, et par un compte-rendu du contexte dans lequel les expressions émotionnelles sont produites. L'identification des émotions, réalisée d'une telle manière, nécessite une validation objective qui est présentée dans les parties I.1.1 et IV.3.2. Ces deux parties consistent en deux tests de perception, dans lesquels les auditeurs évaluent l'émotion des stimuli en termes d'intensité émotionnelle et de positivité émotionnelle. Les stimuli sont présentés à l'oral dans le premier test et à l'écrit dans le deuxième test. Les résultats de chaque test montreront la contribution respective du composant verbal (lexical) et du composant non-verbal (prosodique) à l'expression émotionnelle.

Depuis la proposition de Scholsberg (1954), les différentes émotions sont souvent identifiées dans trois champs sémantiques : d'*activation*, de *valence* et de *puissance*. L'axe d'activation représente l'activité neurale déclenchée par l'excitation émotionnelle, ce qui s'étend du sommeil à la tension. L'axe de valence décrit la positivité de l'émotion, dont les deux extrêmes sont l'émotion positive et l'émotion négative. L'axe de puissance distingue entre l'émotion initiée par le sujet et l'émotion provoquée par l'environnement, en allant du dégoût à la peur et à la surprise. Les chercheurs trouvent ces concepts dimensionnels utiles pour expliquer la nature de l'émotion et décrire les différentes expressions émotionnelles (Osgood *et al*., 1957 ; Udall, 1964 ; Pakosz, 1982 ; Tomkins, 1984 ; Scherer, 1986). Grâce à la généralité des concepts d'activation, de valence et de puissance, l'évaluation de l'émotion basée sur les dimensions a un avantage sur celle qui a recours à des termes émotionnels (ex. joie, colère, tristesse, dégoût, etc.). Vu que la définition des émotions et l'étiquetage catégorique sont toujours problématiques dans ce domaine de travail (voir II.3.), l'utilisation des termes émotionnels risque de causer une confusion parmi les sujets. Cela est surtout vrai quand il s'agit de l'évaluation par des sujets qui proviennent de différentes cultures et parlent différentes langues. Etant donné que notre étude traite le sujet de la perception émotionnelle multiculturelle dans l'une de nos expériences (présentée dans

IV.5.1) et que nous voulons avoir une comparabilité entre les expériences, les tâches des sujets dans tous nos tests de perception consistent à évaluer l'intensité et la positivité de l'émotion sur les axes d'activation et de valence. L'émotion n'étant pas une quantité mesurable, les réponses des sujets sont prises en tant que mesures de l'émotion sur la dimension d'*activation* ('relâché – tendu') et la dimension de *valence* ('positive – négative'). Ces mesures conceptuelles seront comparées aux mesures acoustiques de l'expression vocale de l'émotion, ce qui est présenté dans IV.4.

I.1.1. Expérience 1 : Trois catégories d'émotion, positive, neutre et négative

Le but de l'expérience 1 est de décrire comment les émotions sont exprimées dans la voix de la locutrice WJ et d'identifier trois catégories d'émotion - positive, neutre et négative - par l'évaluation perceptuelle des auditeurs coréens. Cette expérience suppose une relation directe entre l'expression émotionnelle et la perception émotionnelle. C'est-à-dire que l'état émotionnel du sujet parlant serait exprimé dans sa voix et il serait reconnu par l'auditeur lors de l'évaluation de l'émotion des stimuli sur les axes d'activation et de valence. L'expérience 1 se compose d'un test de perception dans lequel l'émotion des énoncés de la locutrice est évaluée en termes d'intensité émotionnelle et de positivité émotionnelle. La variation de ces valeurs perceptuelles en fonction des extraits de parole est estimée par une analyse statistique, ce qui décrit le changement de l'état émotionnel de la locutrice pendant les entretiens. A la fin, les extraits sont regroupés en trois catégories émotionnelles (positive, neutre et négative) selon leurs valeurs de positivité émotionnelle, évaluées par les auditeurs.

Vu que l'état émotionnel de la locutrice fut identifié au moment de notre sélection des données (voir IV.2.4), on pourrait questionner la redondance de ce genre d'expérience. Pourtant, l'expérience 1 consiste en une étape de recherche indispensable dans notre présente étude, parce qu'elle fournit une description objective de nos données (basée sur le jugement multiple des auditeurs naïfs, au lieu du jugement singulier de l'expérimentatrice). Le résultat de l'identification des trois catégories émotionnelles sert de référence à nos analyses suivantes (par exemple, la sélection des stimuli pour les expériences perceptives suivantes sera basée sur les trois catégories d'émotion définies dans cette expérience).

IV.3.1.1. Préparation des stimuli

Tous les énoncés du corpus coréen (présentés dans la partie IV.2.5) sont compris dans l'expérience 1, à part dix énoncés qui contiennent trop de bruits extérieurs. Donc, 100 énoncés de la locutrice WJ sont pris en tant que stimuli du test de perception. La transcription phonétique et la durée des stimuli sont présentées dans l'annexe.

IV.3.1.2. Test de perception

Nous avons fait passer le test de perception à dix Coréens (5 hommes et 5 femmes). Ce sont des étudiants et des chercheurs à l'université de Brown aux Etats-Unis. Leur âge varie entre 25 et 32 ans. Ils s'étaient portés volontaires pour le test. Les auditeurs ont passé le test de perception individuellement dans une chambre insonorisée. Chaque auditeur devait accomplir deux tâches dans des sessions différentes. La première tâche était d'évaluer l'intensité émotionnelle sur l'échelle à cinq points avec deux pôles, 'non émotionnel' et 'très émotionnel'[39]. La deuxième tâche était d'évaluer la positivité de l'émotion. L'auditeur devait choisir l'une des trois cases[40], étiquetées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'. Les instructions étaient indiquées au début du questionnaire de la façon suivante :

« Vous allez écouter les stimuli vocaux, qui consistent en segments de la parole d'une locutrice Coréenne. Les stimuli peuvent être des phrases ou des mots, donc leur longueur varie largement. Vous avez deux tâches à accomplir : (1) Après avoir écouté un stimulus, estimez quel degré d'émotion est exprimé dans ce stimulus, d'après votre impression subjective. Dès que votre décision sera prise, mettez une croix ('x') sur l'un des cinq points, marqués comme 'non-émotionnel,' 'peu émotionnel,' 'émotionnel,' 'assez émotionnel' et 'très émotionnel'. (2) Après avoir écouté un stimulus, identifiez quelle sorte d'émotion (positive, neutre ou négative) est exprimée dans ce stimulus, d'après votre impression subjective. Dès que votre décision sera prise, mettez une croix ('x') sur l'une

[39] Cette échelle désigne l'axe d'activation mais le terme d'activation n'était pas mentionné dans le questionnaire. Il en est ainsi de l'axe de valence pour la deuxième échelle. Un exemple du questionnaire se trouve dans l'annexe.

[40] Etant donné que les deux échelles, l'échelle d'activation et l'échelle de valence, sont indépendantes l'une de l'autre, il n'est pas nécessaire de leur donner le même nombre de points. Trois points, au lieu de cinq points, sont choisis pour l'échelle de valence, en vue d'établir des contrastes nets entre les trois catégories, positive, neutre et négative.

des trois cases, marquées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'. Vous avez trois secondes pour répondre entre chaque stimulus. »

Les auditeurs ont écouté les stimuli par l'intermédiaire d'un haut-parleur, et cela une fois, dans un ordre aléatoire. Vu la grande variation de la durée des stimuli, les stimuli sont subdivisés en trois groupes de durée, 1-1000ms, 1001-2000ms et 2001-3000ms, et l'ordre aléatoire est appliqué à l'intérieur du groupe. L'ordre de la session pour les deux tâches a aussi été contrebalancé entre les auditeurs. Le test a duré 40 minutes, réparties en deux sessions de 15 minutes et une pause de 10 minutes.

IV.3.1.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 2000 réponses des auditeurs coréens au total. La moitié de ces réponses résulte de la première tâche (décider de l'intensité émotionnelle), et nous les appelons **réponses d'activation** (1000 réponses = 100 stimuli x 10 sujets), et l'autre moitié résulte de la deuxième tâche (décider de la positivité émotionnelle), et nous les appelons **réponses de valence** (1000 réponses = 100 stimuli x 10 sujets). Les réponses sont entrées dans les données d'analyse statistique, sous forme numérique. C'est-à-dire que les réponses de valence, étant nominales, sont transformées en chiffres de la manière suivante. La réponse positive est remplacée par la valeur de '+1', celle de neutre par la valeur de '0' et celle de négative par la valeur de '-1'[41]. Les réponses d'activation sont prises telles qu'elles sont marquées dans le questionnaire ; ayant des valeurs de '1' (signalant que l'énoncé est non émotionnel)[42], '2' (peu émotionnel), '3' (émotionnel), '4' (assez émotionnel) ou '5' (très émotionnel). Nous avons calculé la moyenne des dix réponses d'activation et la moyenne des dix réponses de valence pour un stimulus donné (énoncé), afin d'estimer le degré d'intensité émotionnelle et le degré de positivité émotionnelle de chaque stimulus. C'est ce que nous appelons la **valeur**

[41] Les trois valeurs de valence sont considérées comme des valeurs continues puisqu'une échelle continue peut être construite avec un nombre de valeurs égal ou plus grand que trois. Cette considération nous permet de traiter les réponses de valences par l'analyse de variance (ANOVA). L'analyse de variance est l'une des analyses paramétriques qui exigent, comme condition d'utilisation, que la variable dépendante soit exprimée en valeurs continues, qui reposent sur une échelle d'intervalle ou de rapports. Les autres conditions, telles la distribution normale et l'homogénéité de la variance intra-groupe, sont supposées par le nombre des données qui est relativement grand.

[42] On aurait pu faire l'échelle de '0' ('non-émotionnel') à '4' ('très émotionnel'), afin d'avoir une même valeur de '0' pour la signification de l'énoncé 'non-émotionnel'. Pourtant, cela ne nous paraît pas pertinent, vu que les valeurs d'activation et les valeurs de valence, de toute façon, ne sont pas comparables à cause de la différence de leur échelle.

d'activation et la **valeur de valence** de l'énoncé. Ainsi ont été obtenues 100 valeurs d'activation et 100 valeurs de valence pour les 100 énoncés de nos stimuli.

Nous avons effectué deux analyses de variance (ANOVA) pour voir si les valeurs d'activation et les valeurs de valence varient en fonction de l'extrait. Deux variables dépendantes, ACTIVATION (valeurs d'activation) et VALENCE (valeurs de valence), sont examinées en fonction de la variable indépendante, EXTRAIT (huit extraits de la locutrice). Le logiciel statistique 'SPSS ' fut utilisé pour l'analyse statistique.

IV.3.1.4. Résultat

Une ANOVA à un facteur à huit niveaux montre un effet significatif du facteur EXTRAIT sur les valeurs d'activation ($F(7,92)=36,14$, $p<0,01$), ce qui indique que la valeur d'activation varie en fonction de l'extrait. Etant donné que les huit extraits sont tirés de l'entretien de façon chronologique, le changement des valeurs d'activation en fonction des extraits représente l'évolution de l'intensité émotionnelle de la locutrice au cours de l'entretien. Les valeurs d'activation sont plus élevées vers la fin, tandis qu'elles sont variables au début et au milieu de l'entretien. En vue d'une présentation illustrative, les valeurs sont étiquetées en trois catégories d'intensité émotionnelle, '**basse**,' '**moyenne**' et '**haute**' (voir la Figure 17). Ces catégories sont déterminées par la division de la plage de variation[43] en trois gammes proportionnelles. D'après un test post-hoc (scheffe, $\alpha=0,05$), la différence entre les valeurs extrêmes appartenant à la catégorie basse et à la catégorie haute est significative, tandis que la différence entre ces valeurs et les valeurs intermédiaires de la catégorie moyenne ne l'est pas toujours. L'identification des catégories d'intensité émotionnelle n'est pas directement basée sur le test post-hoc parce que le test a produit trois sous-ensembles de valeurs homogènes dont la frontière n'est pas facile à identifier (certaines valeurs appartiennent à deux sous-ensembles de valeurs à la fois). Donc, l'identification des catégories d'intensité émotionnelle est faite par une division arithmétique de la plage de variation en trois gammes proportionnelles, ce qui fournit une distinction nette entre les différentes catégories.

[43] La plage de variation est calculée par la différence entre les valeurs extrêmes, valeur maximum et valeur minimum, chez une locutrice donnée.

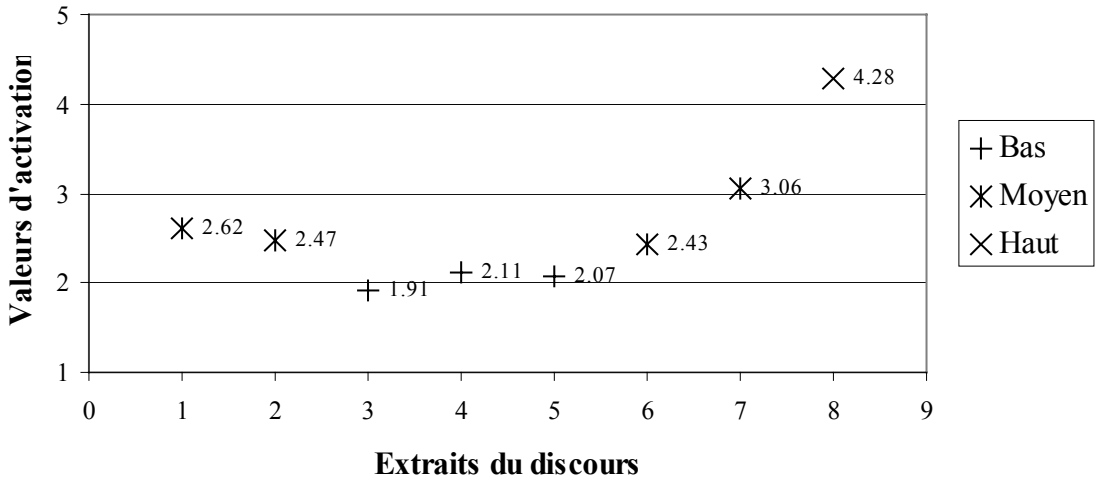


Figure 17. Moyennes des valeurs d'activation (degrés d'intensité émotionnelle) pour les huit extraits[44] de la locutrice WJ.

Une autre ANOVA à un facteur à huit niveaux montre un effet significatif du facteur EXTRAIT sur les valeurs de valence ($F(7,92)=71,53$, $p<0,01$), en indiquant que la valeur de valence varie en fonction de l'extrait. Dans le but de la présentation illustrative de ce résultat, les valeurs de valence sont étiquetées en trois catégories d'émotion - '**positive**,' '**neutre**' et '**négative**' - déterminées par la division de la plage de variation[45] en trois gammes proportionnelles (voir la Figure 18). La différence entre les valeurs extrêmes, appartenant à la catégorie positive et à la catégorie négative, est significative (selon un test post-hoc de scheffe, $\alpha=0,05$), tandis que ce n'est pas toujours vrai pour la différence entre ces valeurs et les valeurs intermédiaires de la catégorie neutre. La raison pour laquelle l'identification des catégories d'émotion est basée sur la division arithmétique, au lieu d'être basée sur la significativité de différence d'après le test post-hoc, a été expliquée ci-dessus.

[44] L'augmentation des valeurs d'activation de l'extrait indique l'augmentation de l'intensité émotionnelle perçue, ce qui est l'équivalent de l'augmentation de l'excitation émotionnelle de la locutrice au moment de cet extrait.
[45] Voir note n°43 pour le calcul de la plage de variation.

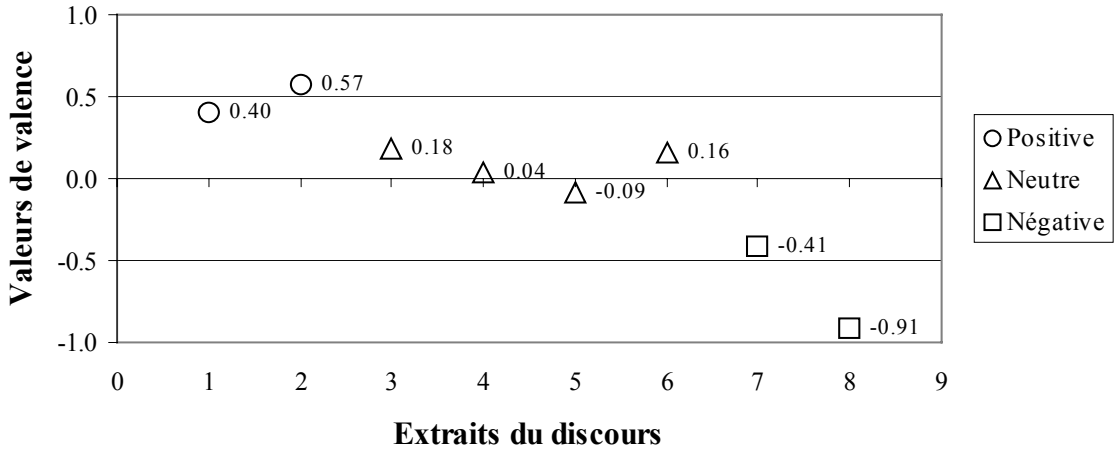


Figure 18. Moyennes des valeurs de valence[46] (degrés de positivité émotionnelle) pour les extraits du discours de la locutrice WJ.

Vu la transition distincte de l'émotion au cours de l'entretien, les extraits du discours sont regroupés en trois catégories émotionnelles, positive, neutre et négative, de la manière suivante. Les deux premiers extraits sont compris dans la catégorie d'émotion positive, les quatre extraits suivants dans la catégorie d'émotion neutre, et les deux derniers extraits dans la catégorie d'émotion négative. D'après l'ANOVA, les trois catégories émotionnelles sont distinguées de manière significative, en termes des valeurs d'activation ($F(2,97)=88,61$, $p<0,01$) et des valeurs de valence ($F(2, 97)=117,92$, $p<0,01$).

Catégories d'émotion	Positive	Neutre	Négative	Moyenne (ET)
Valeurs de valence	0,48 (0,30)	0,04 (0,30)	-0,75 (0,32)	-0,08 (0,62)
Valeurs d'activation	2,56 (0,44)	2,13 (0,51)	3,92 (0,76)	2,87 (0,93)
Nombre d'énoncés	16	42	42	33,3 (15,0)

Tableau 7. Moyennes (et écart-types) des valeurs de valence et des valeurs d'activation pour les trois catégories d'émotion : positive, neutre et négative.

En résumé, les résultats de l'expérience 1 montrent que l'émotion de la locutrice WJ est renforcée vers la fin de l'entretien et que son état émotionnel change d'un état positif vers un état négatif. La théorie d'activation prédit que l'excitation émotionnelle fait augmenter l'activité neurale du sujet parlant, donc l'intensité de l'émotion positive ou

[46] Plus la valeur de valence est proche de '+1,' plus l'émotion de l'extrait est perçue comme positive et plus la valeur de valence est proche de '-1,' plus l'émotion de l'extrait est perçue comme négative.

négative serait plus élevée que celle de l'émotion neutre, non-émotionnelle. Cette prédiction est confirmée dans nos résultats, où les valeurs d'activation des extraits dont l'émotion est identifiée comme positive ou négative sont plus élevées que celles des extraits dont l'émotion est identifiée comme neutre[47]. La valeur d'activation de l'émotion négative de notre locutrice est considérablement plus grande que celle de son émotion positive, ce qui reflète la différence de l'intensité émotionnelle du sourire et des pleurs en général (voir l'explication dans IV.2.3).

IV.3.1.5. Discussion

L'expérience 1 a démontré les caractéristiques des émotions exprimées dans le corpus coréen à travers un test de perception. D'après les résultats du test, trois types d'émotion sont exprimées dans la voix de la locutrice WJ et ils sont reconnus par les Coréens lors de l'évaluation de l'émotion des stimuli vocaux sur l'axe d'activation (intensité émotionnelle) et l'axe de valence (positivité émotionnelle). L'évaluation des auditeurs montre que l'intensité émotionnelle de la locutrice est élevée vers la fin de l'entretien et que son émotion change d'un état positif vers l'état négatif. Les extraits du discours sont regroupés en trois catégories d'émotion - positive, neutre et négative - selon leurs valeurs de valence. Le fait que les émotions de la locutrice, identifiées par les auditeurs sur base des indices vocaux, correspondent à celles identifiées au moment de notre sélection des données (voir IV.2.4) confirme la validité de notre choix des expressions émotionnelles, positive, neutre et négative, au niveau de la construction du corpus coréen.

En général, le sourire et les larmes exprimés dans le visage de la locutrice sont bien reconnus dans la voix par les auditeurs. Tatter (1980) explique la reconnaissance à travers la voix de l'expression faciale du sourire en termes des corrélats acoustiques du sourire. Selon elle, le mouvement musculaire du sourire entraîne une modification du conduit vocal du sujet parlant et fait augmenter la fréquence fondamentale et la fréquence des formants. C'est ce qui est perçu par l'auditeur dans la reconnaissance du sourire. D'après notre expérience, la reconnaissance des larmes est plus facile que celle du sourire, ce qui semble être dû au fait que la voix devenait cassée et tremblante quand la locutrice parlait en

[47] La présente étude suppose la correspondance entre l'intensité émotionnelle perçue et l'excitation émotionnelle exprimée dans la voix de la locturice.

sanglotant. Protopapas & Eimas (1997) caractérise l'effet des pleurs sur la voix de l'enfant par l'augmentation du Fo moyen et de la perturbation de Fo ('jitter'). Leur expérience montre que les pleurs de l'enfant servent une fonction communicationnelle et les valeurs du Fo moyen et de la perturbation de Fo dans la voix pleurante de l'enfant sont corrélées au degré de la gravité négative, perçu par l'adulte. Les traits prosodiques de la voix souriante et de la voix pleurante de la locutrice WJ sont décrits par des mesures acoustiques dans la partie IV.4. Avant d'examiner les traits prosodiques de l'expression émotionnelle de notre locutrice, il est nécessaire de savoir dans quelle mesure les émotions de la locutrice sont exprimées au niveau lexical de ses énoncés. C'est ce qui est adressé dans notre expérience suivante.

IV.3.2. Expérience 2 : Contribution lexicale de l'énoncé à l'expression de l'émotion

Le but de l'expérience 2 est de déterminer comment l'émotion de la locutrice WJ est exprimée au niveau lexical de ses énoncés. Dans cette expérience, nous allons effectuer un test de perception de la même manière que celui de l'expérience 1, mais les stimuli seront présentés aux évaluateurs à l'écrit, à la différence de la présentation orale des stimuli dans l'expérience 1. Les résultats de l'expérience 2 (représentant l'évaluation émotionnelle des énoncés sans indices vocaux) seront interprétés en comparaison avec les résultats de l'expérience 1 (représentant l'évaluation émotionnelle des énoncés avec les indices vocaux). Ainsi seront estimées les expressions de l'intensité et de la positivité émotionnelles dans le lexique des énoncés. C'est ce que nous appelons la *contribution lexicale* de l'énoncé à l'expression émotionnelle dans le corpus coréen.

Dans la conversation naturelle, l'expression vocale (émotionnelle) du sujet parlant suit le contenu de son discours. Quand il est heureux, en parlant de choses agréables, sa joie est exprimée dans sa voix ; quand il est triste, en racontant de histoires tristes, sa tristesse est exprimée dans sa voix. Or, l'information lexicale et l'expression vocale ne sont pas toujours en relation univoque, puisqu'elles appartiennent à des codages indépendants dans la communication parlée, tels le *codage linguistique* et le *codage paralinguistique*, selon les termes de Fónagy (sa théorie du double codage a été présentée dans II.4.4). L'indépendance des deux niveaux, lexical et prosodique, de la communication émotionnelle a été démontrée par nombre d'études expérimentales dont le but principal était de montrer l'importance des indices prosodiques dans l'expression émotionnelle. Les différentes émotions peuvent être exprimées par l'acteur ou par le synthétiseur vocal, en utilisant différentes configurations prosodiques avec un énoncé constant (sémantiquement neutre), et les émotions peuvent être reconnues par l'auditeur dans la condition où il n'y a que des indices vocaux, le contenu lexical étant masqué (Fónagy & Bérard, 1972 ; Scherer *et al.,* 1972 ; Leinonen *et al.,* 1997). Or, l'évidence de la contribution totale de la prosodie à la communication émotionnelle (dans la condition expérimentale) n'exclut pas la possibilité de la contribution lexicale de l'énoncé à la communication de l'émotion (dans la situation naturelle). Vu cette possibilité dans notre corpus, nous allons d'abord examiner

dans quelle mesure la sémantique de l'énoncé peut contiruber à l'expression émotionnelle dans notre corpus, en comparant l'évaluation des stimuli vocaux et celle des stimuli écrits. La réponse à cette question suggérera de la même façon la réponse à une autre question : dans quelle mesure la prosodie de l'énoncé a-t-elle contribué à l'expression émotionnelle de notre locutrice.

IV.3.2.1. Préparation des stimuli

Les stimuli de l'expérience 2 sont construits à partir de la transcription en coréen des 100 énoncés qui ont été utilisés en tant que stimuli de l'expérience 1. Les mots répétés et les expressions raccourcies[48] étant des caractéristiques de la parole spontanée, ils sont transcrits en tant que tels, afin de refléter la prononciation réelle des énoncés. Cependant, la chute des consonnes et des voyelles dans un mot et la permutation des consonnes ou des voyelles à l'intérieur d'un mot ou entre des mots successifs sont rétablies pour que les énoncés soient compréhensibles à l'écrit.

IV.3.2.2. Test de perception

Le test de perception de l'expérience 2 est effectué de la même manière que celui de l'expérience 1, à part le fait que les stimuli transcrits sont utilisés au lieu des stimuli vocaux. Dix Coréens (5 hommes + 5 femmes), qui n'avaient pas participé à l'expérience 1, ont passé le test de perception en tant que volontaires. Ce sont des étudiants à l'université de Brown aux Etats-Unis, dont l'âge varie entre 22 ans et 29 ans. Les sujets ont passé le test de perception individuellement ou à plusieurs dans une salle de l'université. Les tâches des sujets étaient les mêmes que celles de l'expérience 1 ; (1) évaluer l'intensité émotionnelle des stimuli sur l'échelle à cinq points avec deux pôles, 'non émotionnel' et 'très émotionnel' ; (2) et évaluer la positivité émotionnelle des stimuli en choisissant une des trois cases, étiquetées comme 'émotion positive,' 'neutre (état non-émotionnel)' et 'émotion négative'. Les instructions étaient indiquées au début du questionnaire comme suit :

[48] Un exemple coréen est [kIndE] pour [kIrOndE]. Un exmple similaire en français est 'j'ai pas', pour 'je n'ai pas'.

« Vous allez lire une transcription des extraits de parole d'une locutrice Coréenne. Vous avez deux tâches à accomplir : (1) Après avoir lu un stimulus, estimez quel degré d'émotion est exprimé dans ce stimulus, d'après votre impression subjective. Dès que votre décision sera prise, mettez une croix ('x') sur l'un des cinq points, marqués comme 'non-émotionnel,' 'peu émotionnel,' 'émotionnel,' 'assez émotionnel' et 'très émotionnel'. (2) Après avoir lu un stimulus, identifiez quel sorte d'émotion (positive, neutre ou négative) est exprimée dans le stimulus, d'après votre impression subjective. Dès que vous aurez décidez, mettez une croix ('x') sur l'une des trois cases, marquées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'. »

Les stimuli ont été présentés à l'écrit, dans un ordre aléatoire, sur le questionnaire. Dans cette expérience, à la différence du test dans l'expérience 1, nous n'avons pas regroupé les stimuli en fonction de la longueur puisque la différence des longueurs des stimuli n'est pas pertinente dans ce mode de présentation (présentation à l'écrit). Le test a duré 40 minutes, réparties en deux sessions de 15 minutes et une pause de 10 minutes.

IV.3.2.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 2000 réponses des sujets Coréens (100 stimuli x 10 sujets x 2 tâches), dont la moitié sont les **réponses d'activation** (l'évaluation de l'intensité émotionnelle), et l'autre moitié sont les **réponses de valence** (l'évaluation de la positivité émotionnelle). Le traitement des données (réponses des auditeurs) de l'expérience 2 est similaire à celui de l'expérience 1. Les réponses d'activation sont entrées dans les données d'analyse statistique, ayant les valeurs '1' (non-émotionnelle), '2' (peu émotionnelle), '3' (émotionnelle), '4' (assez émotionnelle) et '5' (très émotionnelle), et les réponses de valence sont transformées en valeurs '-1' (négative), '0' (neutre) et '+1' (positive)[49]. Les valeurs d'activation des énoncés sont estimées individuellement, en calculant la moyenne de dix réponses d'activation pour un énoncé donné. Les valeurs de valence des énoncés sont estimées de la même façon que les valeurs d'activation.

[49] Les trois valeurs de valence sont considérées comme des valeurs continues (voir note n°41).

Les analyses de variance (ANOVA) sont effectuées au moyen du logiciel statistique SPSS, pour voir si les émotions de la locutrice peuvent être distinguées par les sujets avec les stimuli écrits, en termes de valeurs d'activation et de valeurs de valence. La variance des deux mesures, ACTIVATION (valeurs d'activation) et VALENCE (valeurs de valence), en fonction du facteur EXTRAIT (huit extraits), est estimée par deux ANOVAs à un facteur à huit niveaux. Etant donné que le but de l'expérience 2 est de voir si l'évaluation émotionnelle avec les stimuli écrits est différente de celle avec les stimuli vocaux, les résultats de cette expérience sont comparés aux résultats de l'expérience 1, en fonction du facteur concerné. Vu la même échelle d'évaluation dans les deux expériences, les moyennes des valeurs d'activation dans les deux résultats sont comparées par un test-t indépendant et la corrélation des deux résultats est évaluée par une analyse de corrélation de Pearson. Il en est de même pour les valeurs de valence.

IV.3.2.4. Résultats

L'ANOVA montre un effet significatif du facteur EXTRAIT sur la variation des valeurs d'activation ($F(7,92)=2,59$, $p<0,05$), ce qui indique que différents degrés d'intensité émotionnelle sont attribués aux extraits du discours lors de l'évaluation émotionnelle des stimuli écrits. Dans le but de la comparaison directe entre les résultats de l'expérience 1 et les résultats de l'expérience 2, nous présentons les valeurs d'activation de l'expérience 2 (voir la Figure 19) en employant le même type de présentation que pour les valeurs d'activation de l'expérience 1 (voir la Figure 17). Les valeurs sont étiquetées en trois catégories d'intensité émotionnelle, '**basse**,' '**moyenne**' et '**haute**,' qui sont déterminées par la division de la plage de variation en trois gammes proportionnelles. En comparaison des valeurs d'activation de l'expérience 1, les valeurs d'activation de l'expérience 2 sont relativement basses et la plage de variation des valeurs est rétrécie. Cette différence entre les deux résultats d'expérience est confirmée par un test-t indépendant, qui montre une différence significative entre la moyenne des valeurs d'activation de l'expérience 1 et celle de l'expérience 2, qui sont respectivement de 2,94 et 2,28 ($t(198)=5,56$, $p<0,01$). Autrement dit, l'intensité émotionnelle évaluée avec les stimuli vocaux est plus grande que celle évaluée avec les stimuli écrits, ce qui n'est pas surprenant, vu la grande expressivité de la voix humaine. La variation des valeurs d'activation en fonction de l'extrait est aussi différente entre les deux résultats. Par exemple, la valeur d'activation de l'extrait 8 était considérablement plus haute que celle des autres dans

l'expérience 1 tandis que ce n'est pas le cas dans l'expérience 2. L'analyse de corrélation de Pearson montre que les valeurs d'activation de l'expérience 1 et celles de l'expérience 2 ne sont pas corrélées, bien qu'il s'agisse de la même source de stimuli (énoncés de la locutrice WJ) et de la même échelle d'évaluation (échelle d'intensité émotionnelle) ($r(100)=0,04$, $p>0,05$). L'absence de corrélation des valeurs d'activation dans les deux résultats indique l'indépendance entre l'évaluation émotionnelle basée sur l'information lexicale et celle basée sur l'information vocale.

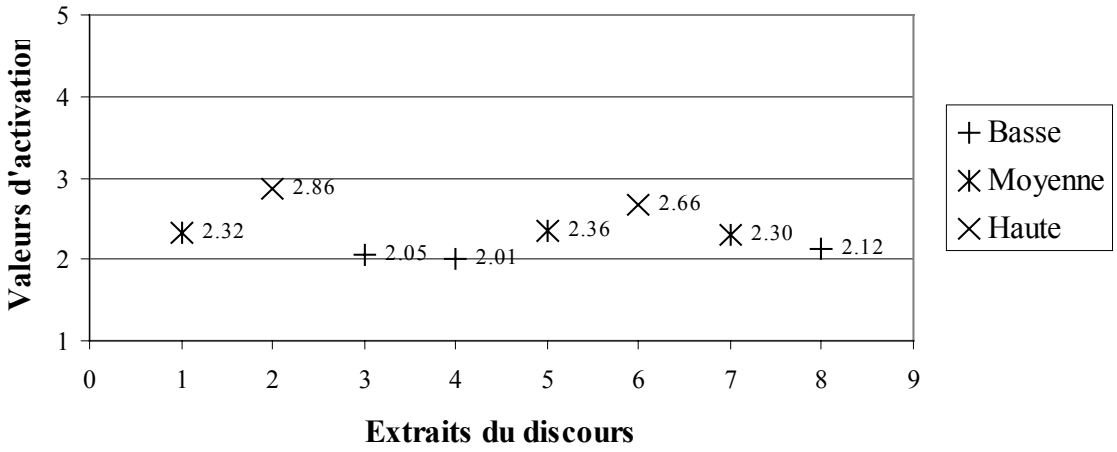


Figure 19. Moyennes des valeurs d'activation (degré d'intensité émotionnelle) pour les huit extraits du discours, d'après l'évaluation émotionnelle des stimuli écrits.

L'effet du facteur EXTRAIT sur les valeurs de valence est aussi trouvé significatif ($F(7,92)=2,59$, $p<0,05$), ce qui indique que l'émotion des extraits est différenciée par la valeur de valence. Dans la Figure 20, les valeurs de valence sont identifiées en trois catégories d'émotion, '**positive**,' '**neutre**' et '**négative**,' comme il a été fait pour les valeurs de valence de l'expérience 1 (voir la Figure 18). La positivité émotionnelle de l'extrait évaluée dans cette expérience n'est pas comparable à celle évaluée dans l'expérience 1, sauf que l'émotion de l'extrait 2 est identifiée comme positive dans les deux résultats. Or, l'émotion de l'extrait 1 est identifiée comme négative dans cette expérience, alors qu'elle était identifiée comme positive dans l'expérience 1. De plus, l'émotion des extraits 7 et 8 est identifiée comme neutre dans cette expérience (l'évaluation de l'émotion des stimuli écrits), tandis qu'elle était identifiée comme considérablement plus négative que celle des autres extraits dans l'expérience 1 (l'évaluation de l'émotion des stimuli vocaux).

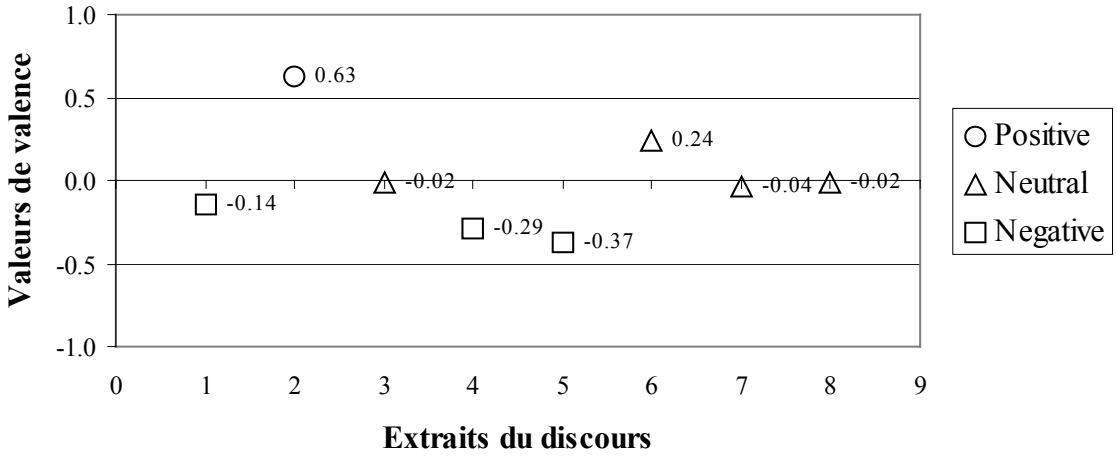


Figure 20. Moyennes des valeurs de valence[50] (degrés de positivité émotionnelle) pour les huit extraits du discours, d'après l'évaluation émotionnelle des stimuli écrits.

La sémantique de son discours à la fin de l'entretien (correspondant aux extraits 7 et 8) n'est pas particulièrement plus négative que celle du discours au début ou au milieu de l'entretien, tandis que l'émotion de sa voix à la fin de l'entretien a été perçue comme considérablement plus négative que celle au début ou au milieu de l'entretien d'après l'expérience 1 (voir la Figure 18). La neutralité de l'émotion évaluée avec les stimuli écrits, par rapport à celle évaluée avec les stimuli vocaux, est démontrée par un test-t indépendant. Le test montre une différence significative des valeurs de valence entre l'expérience 1 et l'expérience 2, dont les moyennes sont de respectivement -0,23 et -0,06 ($t(198)=-2,26$, $p<0,05$). La différence des moyennes signifie que l'impression globale émotionnelle des énoncés est beaucoup plus négative dans l'évaluation des stimuli vocaux que dans l'évaluation des stimuli écrits. Les valeurs de valence de l'expérience 1 et de l'expérience 2 ne sont pas corrélées, d'après l'analyse de corrélation de Pearson ($r(100)=0,19$, $p>0,05$).

Afin de tester si les trois catégories d'émotion - positive, neutre et négative - déterminées dans l'expérience 1 (voir le Tableau 7), peuvent être distinguées au niveau lexical de l'énoncé, les moyennes des valeurs d'activation et de valence, évaluées dans cette expérience, sont calculées pour les trois catégories dans le Tableau 8. D'après une analyse ANOVA, les valeurs d'activation des trois catégories émotionnelles ne sont pas significativement différentes ($F(2,97)=2,55$, $p>0,05$). La valeur de valence de la catégorie

[50] La signification des valeurs de valence des extraits est expliquée dans la note n°46.

d'émotion neutre est différente de celles des catégories d'émotion, positive et négative ($F(2,97)=3,57$, $p<0,05$), mais les valeurs de valence des deux dernières catégories émotionnelles, positive et négative, ne sont pas différentes du point de vue statistique (test post-hoc, $\alpha=0,05$).

Catégories d'émotion	Positive	Neutre	Négative	Moyenne (ET)
Valeurs de valence	0,19 (0,53)	-0,18 (0,56)	-0,03 (0,39)	-0,06 (0,50)
Valeurs d'activation	2,55 (0,49)	2,28 (0,51)	2,18 (0,65)	2,28 (0,59)
Nombre d'énoncés	16	42	42	33,3 (15,0)

Tableau 8. Moyenne (et écart-types) des valeurs de valence et des valeurs d'activation pour les trois catégories émotionnelles, positive, neutre et négative, d'après l'évaluation de l'émotion des stimuli écrits.

IV.3.2.5. Discussion

D'après l'expérience 2, les valeurs d'activation et de valence évaluées avec les stimuli écrits sont différentes de celles évaluées avec les stimuli vocaux dans l'expérience 1. La différence majeure entre les deux résultats d'expérience se trouve dans la différente évaluation émotionnelle des deux derniers extraits du discours (extraits 7 et 8). L'émotion de ces deux derniers extraits était identifiée comme explicitement plus négative que celle des autres extraits dans l'expérience 1, tandis qu'elle est identifiée comme peu négative ou neutre dans l'expérience 2. En termes d'expression, l'émotion négative de la locutrice WJ aux moments d'où les extraits 7 et 8 sont tirés est exprimée dans la voix d'une manière explicite mais peu exprimée au niveau lexical de ses énoncés. Vu que les catégories d'émotion, positive et négative, ne sont pas distinguées par les valeurs d'activation et de valeur dans l'expérience 2, nous arrivons à la conclusion que la contribution lexicale de l'énoncé à l'expression émotionnelle est relativement indépendante et peu pertinente dans notre corpus coréen.

IV.4. Analyse acoustique

Dans la partie précédente, nous avons vu comment les émotions de la locutrice WJ sont exprimées au niveau lexical de l'énoncé. Dans cette partie, nous étudions comment ses émotions sont exprimées au niveau prosodique de l'énoncé. Vu que les trois catégories d'émotion - positive, neutre et négative - sont bien distinguées avec les stimuli vocaux dans l'expérience 1, mais qu'elles sont peu distinguées avec les stimuli écrits dans l'expérience 2, l'information prosodique des énoncés de notre corpus semble être essentielle dans l'identification des émotions. Notre tâche ici est de décrire le changement prosodique des énoncés en fonction de l'état émotionnel des locutrices avec des mesures acoustiques. Les valeurs comme le Fo moyen, le Fo maximum, le Fo minimum, la moyenne des 20% des valeurs les plus basses de Fo ('*Fo Moy Bas*'), la plage de Fo, la perturbation de Fo ('*jitter*'), la perturbation d'intensité ('*shimmer*') et le débit de parole sont mesurées pour 100 énoncés[51]du corpus coréen, au moyen du logiciel 'Winpitch'[52].

L'étude prosodique de la voix émotionnelle est une *analyse de tendance* (Fónagy, 1990, p307). Les caractères des traits vocaux sont des phénomènes statistiques, qui sont généralement, mais pas toujours, présents dans une telle circonstance. Les traits vocaux de la voix émotionnelle sont décrits par les chercheurs en terme du degré de déviation typique, par rapport aux traits de la voix neutre (voir II.5.). Notre analyse acoustique se sert de cette méthode comparative, en se basant sur l'estimation statistique (ANOVAs) de la variation des valeurs acoustiques en fonction de l'émotion de la locutrice.

IV.4.1. Mesures acoustiques

Avant de présenter les résultats de nos mesures acoustiques, la fiabilité de nos mesures basées sur l'enregistrement de vidéo mérite d'être mentionnée, en citant l'expérience de Doherty & Shipp (1988). Ces derniers ont comparé les mesures de Fo,

[51] Ce sont les mêmes énoncés qui ont été utilisés en tant que stimuli de l'Expérience 1 puisque tous les énoncés du corpus coréen, sauf dix énoncés qui ne sont pas acoustiquement propres, ont été inclus dans l'Expérience 1.

[52] Le logiciel 'Winpitch' est dévelopé et commercialisé par Martin (1995) pour l'analyse de la prosodie. Il a des fonctions similaires à celles d'autres logiciels de l'anaylse phonétique, telles l'édition du son, le spectrogramme et le filtrage du signal. Il a aussi une fontion de resynthèse, permettant à l'expérimentateur de vérifier les résultats des paramètres modifiés, ce qui était essentiel dans la préparation des stimuli de notre Expérience 7 (voir VI.2.1).

d'amplitude, de jitter et de shimmer des mêmes ondes sinusoïdales, prises à partir de différentes conditions d'enregistrement : l'enregistrement direct du signal sur l'ordinateur et les enregistrements indirects via audio bande ('reel-to-reel tape'), audio-cassette et vidéo-cassette (PCM/VCR). D'après leurs résultats, les valeurs du jitter et du shimmer peuvent être correctement estimées quand le signal est numérisé directement sur l'ordinateur ou via la vidéo cassette, tandis qu'elles peuvent être surestimées quand le signal est numérisé via la bande audio ou l'audio cassette. Ce genre de divergence selon les différents formats d'enregistrement n'est pas constaté dans l'estimation des valeurs du Fo et de l'intensité moyennes.

IV.4.1.1. Fo moyen, Fo maximum, Fo minimum, 'Fo Moy 20% Bas' et plage de Fo

Le Fo est le paramètre le plus étudié dans les analyses acoustiques de l'émotion, à cause de sa représentativité du changement de la voix dû à l'effet émotionnel. Le Fo moyen est une estimation de la variation globale des valeurs de Fo dans une unité d'analyse. Dans notre analyse, il est calculé par la moyenne des valuers de Fo de chaque énoncé, et puis le Fo moyen de l'extrait est estimé par la moyenne des Fo moyens des énoncés dans chaque extrait du discours. La moyenne des 20% des valuers les plus basses de Fo (appelée ci-dessous '*Fo Moy Bas*') est une autre façon d'estimer la variation globale des valeurs de Fo dans une unité données, qui est moins influencée par la présence des valeurs extraîmement hautes que le Fo moyen. Le Fo maximum et le Fo minimum sont les valeurs extrêmes, qui sont la plus grande et la plus petite parmi les valeurs de Fo de l'énoncé. La plage de Fo signifie la différence de ces deux valeurs extrêmes. La détection du Fo dans notre analyse se base sur la méthode de la combinaison rapide ('*Fast comb*')[53].

D'après les études précédentes, le Fo augmente en général avec l'augmentation de l'excitation émotionnelle mais elle peut diminuer chez certains individus (Bonner, 1943 ; Hecker *et al*., 1968 ; Scherer, 1982). Ainsi, le stress émotionnel du locuteur peut se manifester dans la voix par l'augmentation ou la diminution du Fo, tandis que l'auditeur perçoit le degré du stress émotionnel toujours basé sur une corrélation positive entre le Fo et le degré du stress émotionnel (Streeter *et al*., 1983). En ce qui concerne la variation du

[53] La méthode de la combinaison rapide ('*Fast Comb*') est une variante de la méthode de la combinaison classique ('*Classic Comb*'), qui cherche à trouver une structure harmonique dans le spectre. Elle ne prend en compte que des fréquences au-dessous de 2000Hz. Au-dessus de ce seuil, elle utilise l'information différentielle pour mieux détecter le Fo. Cet algorithme permet une détection rapide du Fo et son résultat est plus fiable (voir Martin, 1995).

Fo pour l'émotion de la joie, le Fo augmente quand il s'agit de la forme forte (comme l'allégresse), tandis qu'elle diminue quand il s'agit de la forme atténuée (comme le plaisir et le contentement). En ce qui concerne la variation du Fo pour l'émotion de la tristesse, on retrouve les deux sortes de variation : le Fo augmente quand la tristesse est exprimée sous ses formes excitées (comme le désespoir et la détresse) et elle diminue quand la tristesse est en forme tranquille (ce qui a été étudié dans la plupart des études précédentes). La variation du Fo est étroitement liée à celle de l'intensité moyenne ; l'intensité moyenne du signal de parole augmente dans la première forme et diminue dans la dernière (Pittam & Scherer, 1993).

Au sein de notre corpus, le Fo de la voix de la locutrice WJ est plus élevée dans ses émotions de joie et de tristesse que dans son émotion neutre. Chronologiquement, le Fo est plus élevé au début (correspondant aux extraits 1 et 2) et à la fin (correspondant aux extraits 7 et 8) qu'au milieu (correspondant aux extraits 4 et 5) de l'entretien (voir la Figure 21). Une ANOVA montre que la différence du Fo moyen entre les extraits est significative, ce qui est principalement dû à l'augmentation du Fo moyen pour l'émotion de joie, exprimée dans les extraits 1 et 2 ($F(7,92)=2,74$, $p<0,05$). Une autre ANOVA est effectuée pour voir si les trois catégories d'émotion - positive, neutre et négative - déterminées dans l'expérience 1, se différencient en terme du Fo moyen, et le résultat montre un effet significatif du Fo moyen entre les trois catégories émotionnelles ($F(2,92)=5,86$, $p<0,05$). Selon un test post-hoc de Scheffe ($\alpha=0,05$), le Fo moyen de la catégorie d'émotion positive (249,8Hz) est significativement différente de celles d'émotion négative (229,8Hz) et d'émotion neutre (222,1Hz), tandis que les Fo moyens des deux dernières catégories émotionnelles ne sont pas différentes du point de vue statistique, ce qui suggère l'efficacité de la mesure de Fo moyen pour repérer l'émotion positive vocale. Or, l'absence de différence significative du Fo moyen entre l'émotion négative et l'émotion neutre n'est pas cohérent avec notre prédiction, vu que la tristesse de la locutrice WJ était exprimée en forme excitée (comme la détresse) dans notre corpus coréen. L'excitation émotionnelle générale (soit la joie, soit la tristesse) semble être mieux expliquée par les valeurs du Fo maximum que par celles du Fo moyen (voir plus loin).

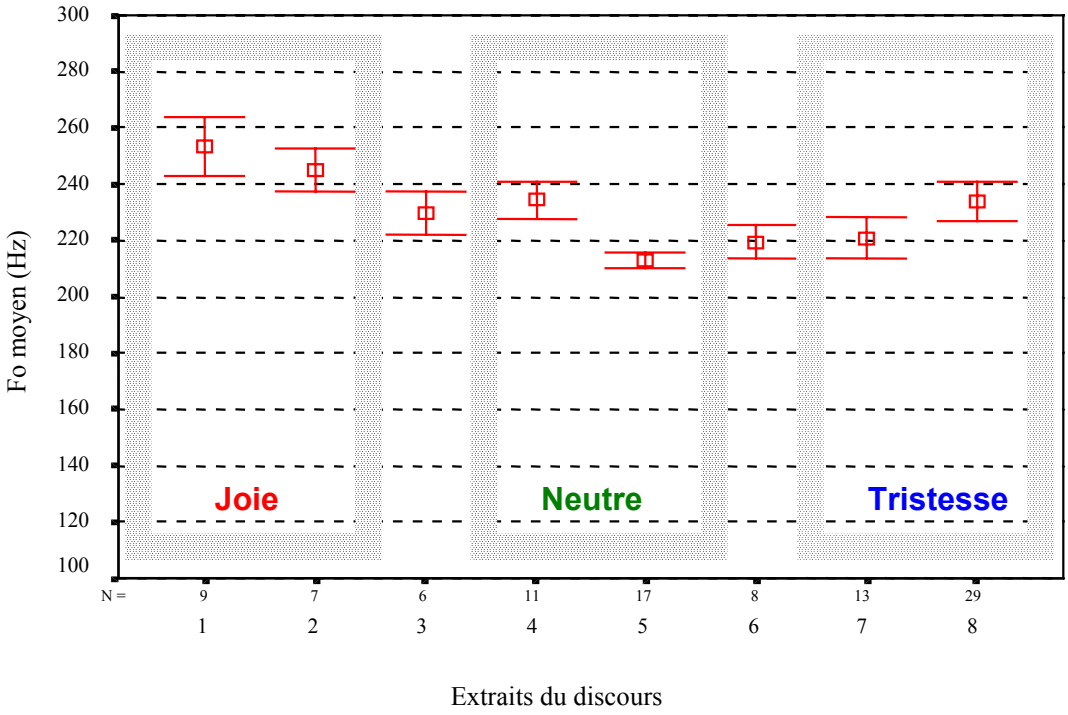


Figure 21. Moyennes et écart-types du Fo moyen (Hz) des énoncés dans les huit extraits du discours exprimant la joie, la neutre ou la tristesse.

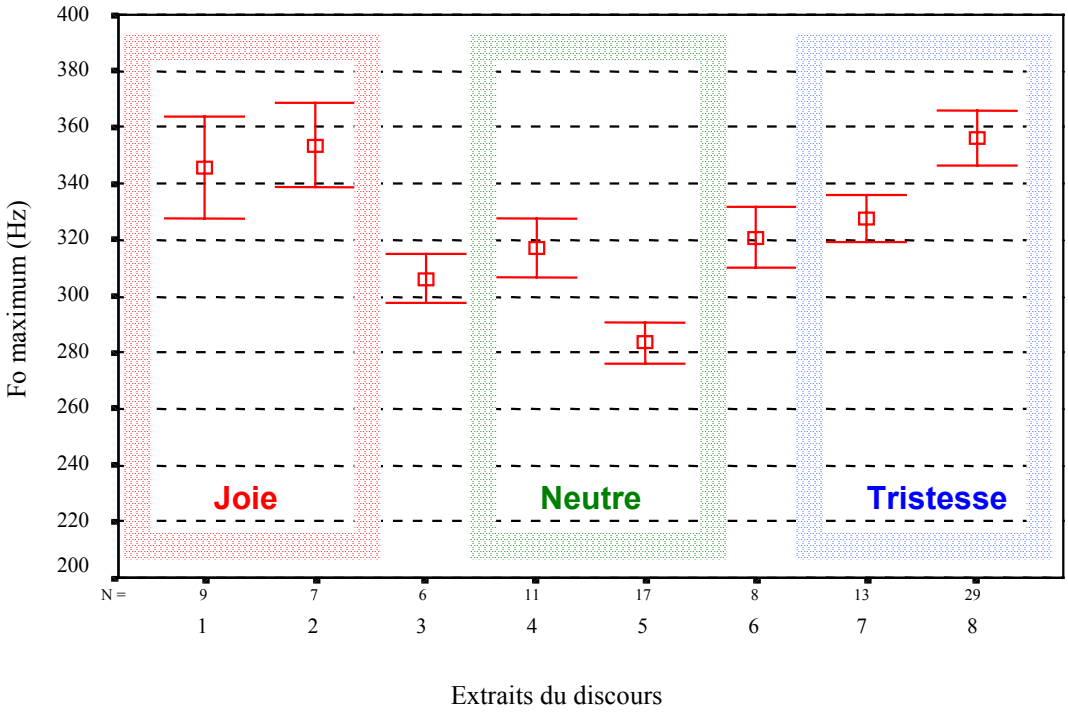


Figure 22. Moyennes et écart-types du Fo maximum (Hz) des énoncés dans les huit extraits du discours exprimant la joie, la neutre ou la tristesse.

Dans la Figure 22, les Fo maxima des extraits 2 (353,4Hz) et 8 (356,0Hz) sont significativement plus hauts que celle de l'extrait 5 (283,4Hz). C'est-à-dire que le Fo maximum est significativement plus élevé quand la locutrice est joyeuse ou triste que quand elle parle sans émotion particulière ($F(7,92)=5,76$, $p<0,01$). En ce qui concerne le Fo maximum des trois catégories émotionnelles, déterminées dans l'expérience 1, le Fo maximum est plus élevé dans les catégories d'émotion positive (348,9Hz) et d'émotion négative (347,2Hz) que dans celui d'émotion neutre (302,6Hz) ($F(2,97)=13,7$, $p<0,01$). Ainsi est repérée l'excitation émotionnelle de l'émotion positive et de l'émotion négative par l'élévation du Fo maximum par rapport au Fo maximum de l'état neutre.

La voix triste (en larmes) semble être mieux caractérisée par la variation des basses valeurs de Fo que par celle des hautes valeurs de Fo. Le Fo minimum des extraits 7 et 8 (tirés à partir des moments où la locutrice WJ parlait en pleurant) est significativement plus bas que celui des extraits 4 et 5 (tirés à partir des moments où la locutrice parlait sans émotion particulière) ($F(7,92)=13,8$, $p<0,01$), ce qui est montré dans la Figure 23. Le Fo minimum des extraits 1 et 2 (tirés à partir des moments où la locutrice WJ parlait en souriant) n'est pas très différent de celui des extraits 4 et 5. En ce qui concerne les valeurs de Fo minimum des trois catégories émotionnelles, le Fo minimum de la catégorie d'émotion négative (76,6Hz) est significativement plus bas que celui des catégories d'émotion positive (124,7Hz) et d'émotion neutre (109,7Hz) ($F(2,97)=30,5$, $p<0,01$).

En résumé, les émotions positive et négative sont distinguées de l'émotion neutre par le Fo maximum ; l'émotion positive est distinguée des émotions négative et neutre par le Fo moyen ; et l'émotion négative est distinguée des émotions positive et neutre par le Fo minimum. Vu le différent repérage des émotions par la variation des hautes valeurs de Fo et celle des basses valeurs de Fo, nous avons calculé un autre type de Fo moyen ('*Fo Moy Bas*'[54]) en ne prenant en compte que 20% des valeurs les plus basses de Fo repérées à partir d'un histogramme des valeurs de Fo de l'énoncé. Le 'Fo Moy Bas' peut rendre une meilleure estimation globale des valeurs de Fo de l'énoncé que le Fo moyen, en ce sens qu'il est moins susceptible à la présence des valeurs de Fo extrêment hautes, qui sont éventuellement dues à l'erreur de la détection automatique de la période du signal.

[54] La mesure du *Fo MoyBas* a été suggérée par Scherer dans une communication personnelle pendant les Journées d'Etudes sur la Parole, qui a eu lieu à Martigny en Suisse du 15 au 19 juin 1998.

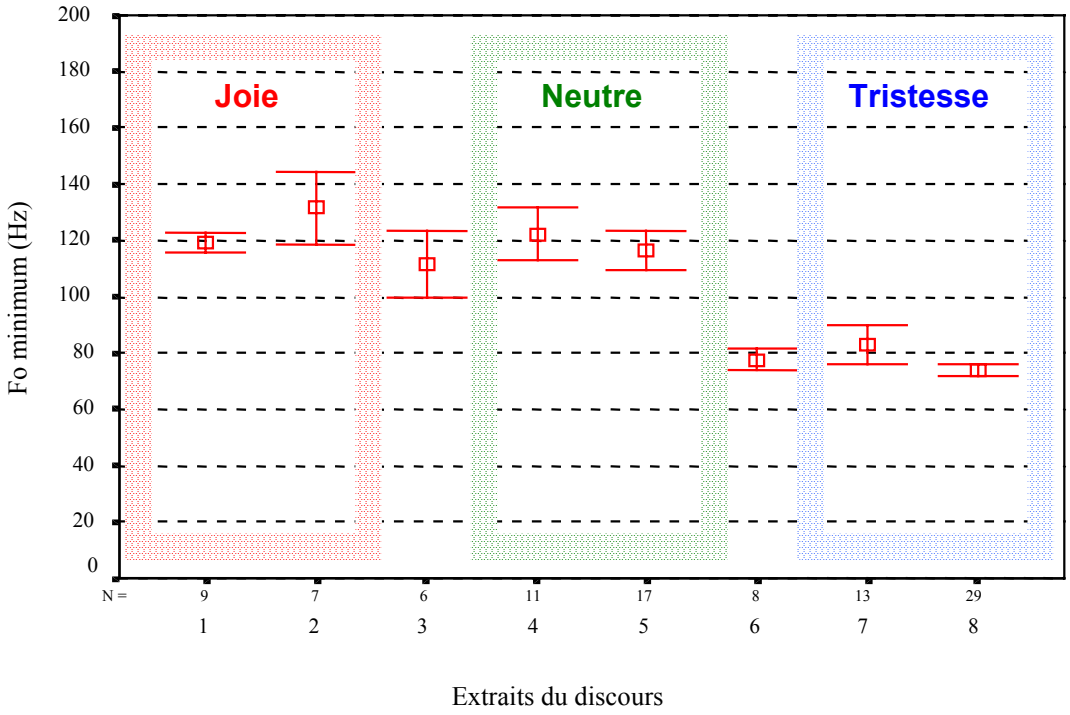


Figure 23. Moyennes et écart-types du Fo minimum (Hz) des énoncés dans les huit extraits exprimant la joie, la neutre ou la tristesse.

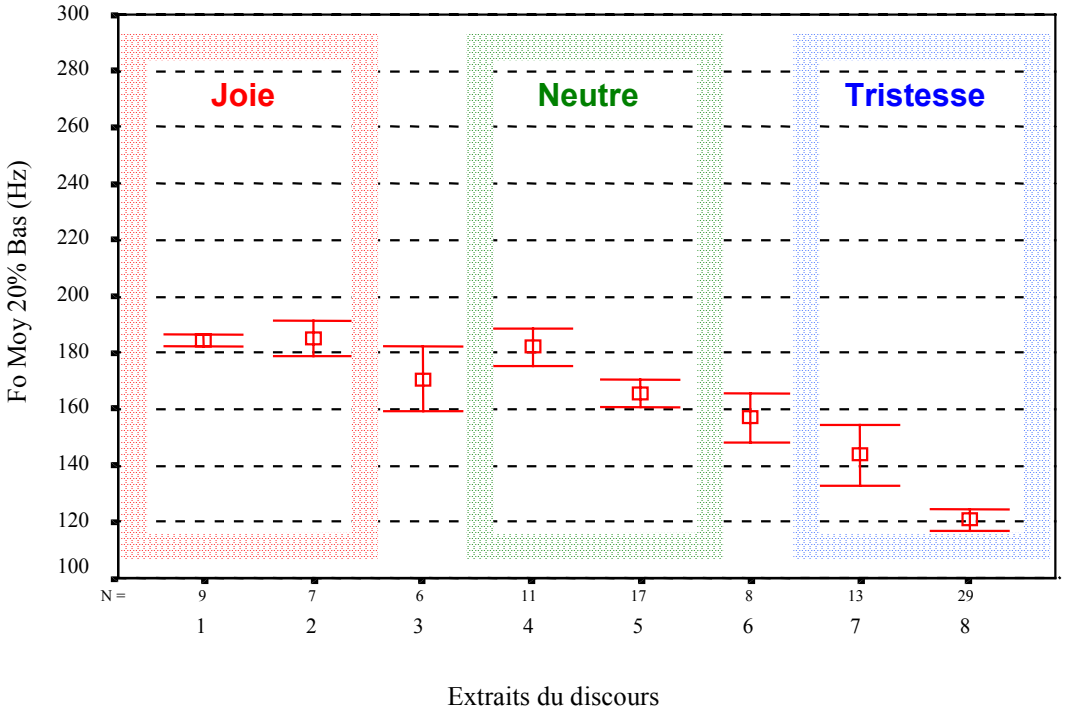


Figure 24. Moyennes et écart-types du 'Fo Moy Bas' (Hz) des énoncés dans les huit extraits du discours exprimant la joie, la neutre ou la tristesse.

Etant donné qu'il s'agit de la moyenne des basses valeurs de Fo, le phénomène vocal lié au ralentissement de la vibration des cordes vocales, comme la glottalisation, serait bien capté par cette mesure du 'Fo Moy Bas'. Au sein de notre corpus, le 'Fo Moy Bas' diminue progressivement du début vers la fin de l'entretien, ce qui reflète le changement de la vibration glottale vers la voix pleurante, apparue à la fin de l'entretien (voir la Figure 24). Le 'Fo Moy Bas' des extraits 7 et 8 est significativement plus bas que celle des extraits 1 et 2 ($F(7,92)=15,6$, $p<0,01$). En ce qui concerne la distinction des trois catégories émotionnelles par le 'Fo Moy Bas,' ce dernier est significativement plus bas pour l'émotion négative (127,8Hz) que pour l'émotion positive (184,7Hz) et l'émotion neutre (169,0Hz) ($F(2,97)=43,5$, $p<0,01$).

La plage de Fo ('*Fo range*' en anglais) est aussi distinctive en tant qu'indice de l'émotion. La plage de Fo est plus grande dans l'extrait 1 (226,0Hz) et l'extrait 8 (282,3Hz) que dans l'extrait 5 (167,3Hz), ce qui indique que la plage de Fo est plus grande pour la joie et la tristesse que pour le neutre dans notre corpus ($F(7,92)=10,6$, $p<0,01$). La plage de Fo de la voix triste de notre locutrice est particulièrement plus grande que celle de sa voix joyeuse, à cause du différent changement fréquentiel en hautes valeurs de Fo et en basses valeurs de Fo dans la voix en larmes. C'est-à-dire, le Fo maximum augmente et le Fo minimum diminue dans la voix pleurante, alors que dans la voix souriante le Fo maximum augmente mais le Fo minimum reste presque le même par rapport aux valeurs de la voix neutre. En ce qui concerne la distinction des trois catégories émotionnelles par la plage de Fo, la plage de Fo de l'émotion négative (270,6Hz) est significativement plus élevée que celles de l'émotion positive (224,3Hz) et de l'émotion neutre (192,9Hz) ($F(2,97)=24,1$, $p<0,01$). Or, l'élévation de la plage de Fo de l'émotion positive, par rapport à la valeur de l'émotion neutre, n'atteint pas la significativité statistique[55]. En terme de la relation des indices acoustiques et perceptuelles, la valeur de la plage de Fo est positivement corrélée au degré de l'émotion perçue (Hutter, 1968). Cette corrélation est confirmée dans les valeurs de la plage de Fo des énoncés de notre locutrice, telle que la variation de la plage de Fo pour les huit extraits du discours (voir la Figure 25) correspond à celle des valeurs d'activation de ces extraits, perçues par les Coréens dans l'expérience 1 (voir la Figure 17).

[55] L'absence de significativité de la différence entre la valeur de l'émotion positive et celle de l'émotion neutre s'explique par la valeur relativement élevée de l'émotion neutre, ce qui est largement dû à la prise en compte de la valeur de l'extrait 6 dans l'estimation de la plage de Fo de la catégorie d'émotion neutre (voir la Figure 25).

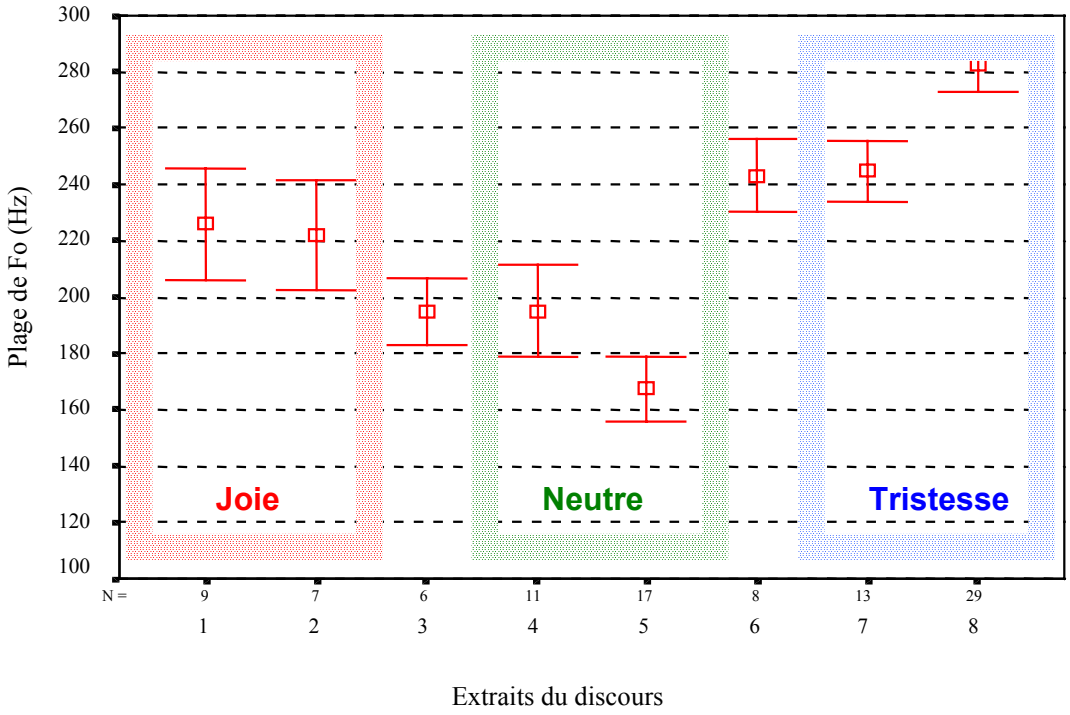


Figure 25. Moyennes et écart-types de la plage de Fo (Hz) des énoncés dans les huit extraits du discours exprimant la joie, la neutre ou la tristesse.

IV.4.1.2. Perturbation de Fo (Jitter)

La perturbation de Fo (‘*jitter*’ en anglais) réfère à la micro-variation du Fo, c’est-à-dire la variation du Fo, cycle par cycle, dans une unité acoustique. La mesure de jitter est proposée par Lieberman (1961) en tant que description quantitative de la qualité de la voix, surtout en ce qui concerne la voix émotionnelle. Il s’agit de la variation des périodes adjacentes dans un signal acoustique, dont l’estimation peut être accomplie à travers diverses méthodes d’analyse (Horii, 1982 ; Klingholz & Martin, 1985 ; Schoentgen & Guchteneere, 1995). Dans notre analyse, le jitter est mesuré par l’estimation du pourcentage de la différence des valeurs de Fo entre les cycles adjacents, dont le calcul mathématique peut être exprimé comme suit :

$$\text{jitter } (\%) = \frac{100\,N \sum_{1}^{N-1} |F0_{i+1} - F0_i|}{(N-1) \sum_{1}^{N} F0_i}$$

Les résultats précédents sur le paramètre du jitter dans l'émotion vocale ne sont pas toujours congruents. Lieberman & Michael (1962) ont montré que la valeur de jitter influence l'identification de l'émotion, tandis que Protopapas & Lieberman (1995) ont constaté que le jitter n'est pas lié à la perception de l'émotion mais à la perception de la qualité générale de la voix, comme l'enrouement de la voix[56]. Pourtant, Protopapas & Eimas (1997) ont trouvé une corrélation forte entre le jitter du cri d'enfant et la perception de l'émotion négative par l'adulte, et Bachorowski & Owren (1995) ont démontré la variation significative du jitter en fonction de l'expression de l'émotion positive et de l'émotion négative.

Au sein de notre corpus, le jitter de l'extrait (calculé par la moyenne des valeurs de jitter des énoncés dans l'extrait) est différent dans les huit extraits du discours mais aucune différence n'est significative ($F(7,92)=1,40$, $p>0,05$). Les valeurs de jitter des trois catégories émotionnelles - positive, neutre et négative - (calculées par les moyennes des valeurs de jitter des extraits dans la catégorie) sont de respectivement 0,6, 0,8 et 0,9%, dont la différence n'est pas significative non plus ($F(2,97)=1,39$, $p>0,05$).

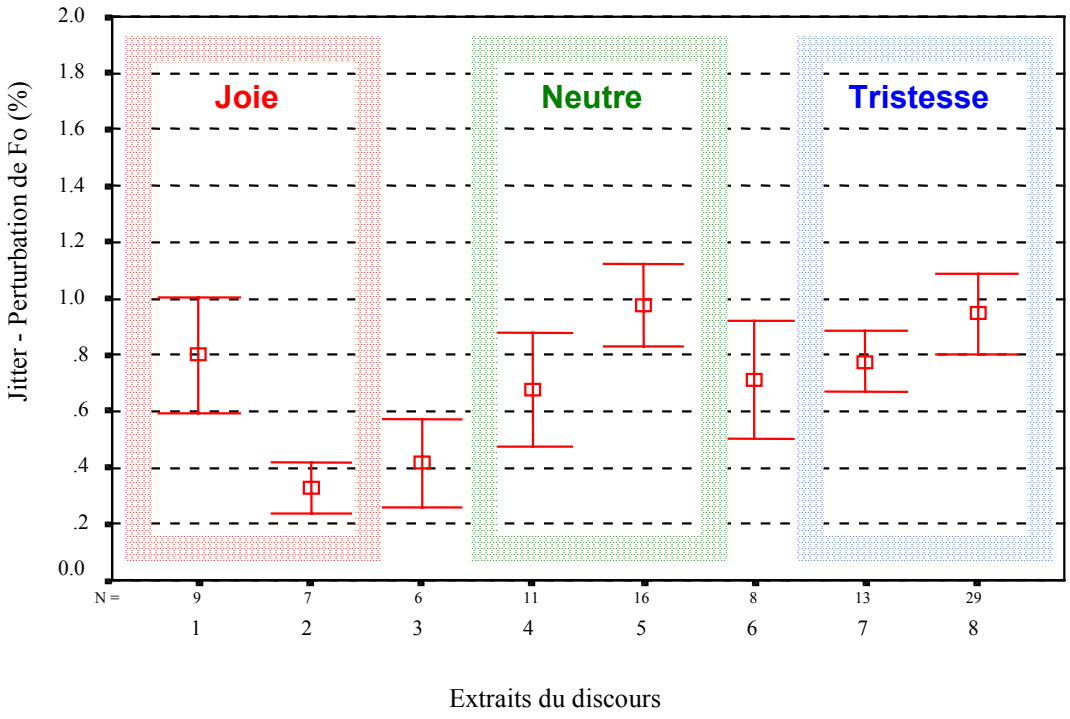


Figure 26. Moyennes et écart-types de la perturbation de Fo (jitter, %) des énoncés dans les huit extraits du discours exprimant la joie, la neutre ou la tristesse.

[56] A cet égard, Hillenbrand (1988) a démontré une corrélation forte entre le jitter et le degré perçu de la rudesse vocale ('*perceived roughness*') dans son expérience utilisant la voix synthétique.

Même si la moyenne des valeurs de jitter n'est pas distinctive entre les différentes émotions du point de vue statistique, l'observation des valeurs de jitter individuelles de chaque extrait nous indique que la valeur maximum de jitter est plus élevée dans la tristesse (extrait 8) que dans la joie (extrait 1) et le neutre (extrait 5) (voir la Figure 27). L'augmentation du jitter dans l'expression de la tristesse de notre locutrice est différente des résultats de l'expérience de Lieberman (1961) dans lequel le jitter a diminué dans l'expression émotionnelle (comme le bonheur ou la peur). Lieberman a expliqué cette diminution du jitter par la présence du contrôle du sujet parlant sur la production vocale puisqu'il s'agit de la simulation (stylisation) de l'émotion. Notre résultat, l'augmentation du jitter dans le cas de la tristesse, peut être expliqué par la perte du contrôle conscient sur la production vocale, du fait qu'il s'agit de l'expression de l'émotion vécue, produite par une expérience réelle. Ce genre d'irrégularité de la vibration vocale dans la voix émotionnelle a été remarquée dans l'expérience de Hecker *et al.* (1968, p1000). Une autre raison possible de la différence entre notre résultat et le résultat de Lieberman peut être attribuée à la différence des émotions concernées, la tristesse et la peur, dans les deux expériences. Le jitter de la voix joyeuse est plus bas que celui de la voix neutre dans notre expérience, de même que le jitter de la voix heureuse est plus bas que celui de la voix neutre dans l'expérience de Lieberman. Pourtant, toutes ces différences du jitter entre les différentes émotions de notre corpus sont masquées par la majorité des valeurs basses de jitter dans l'estimation du jitter de l'extrait, qui fut calculé par la moyenne des valeurs de jitter des énoncés dans l'extrait. Les valeurs de jitter de la voix de la locutrice WJ varient considérablement en comparaison avec les valeurs de jitter trouvées dans les autres études. D'après le compte-rendu de Klingholz & Martin (1985), la valeur de jitter de la voix normale (émotionnellement neutre) varie entre 0,3 et 3,2% (Tableau 9). Les valeurs de jitter de la voix de notre locutrice se trouvent aussi entre 0 et 3,2% (Figure 27).

Etudes précédentes	**Jitter (%)**	**Shimmer (%)**
Hiki, Sugawara & Oizumi (1968)	0,8 / 3,2	---
Koike, Takahashi & Calcaterra. (1977)	0,4 / 1,1	0,6 / 5,0
Horii (1980)	0,4 / 1,1	---
Klingholz & Martin (1985)	0,3 / 1,6	0,7 / 6,0

Tableau 9. Valeurs (minimum / maximum) de jitter (%) et de shimmer (%) de la voix normale dans les études précédentes, d'après Klingholz & Martin (1985, p172).

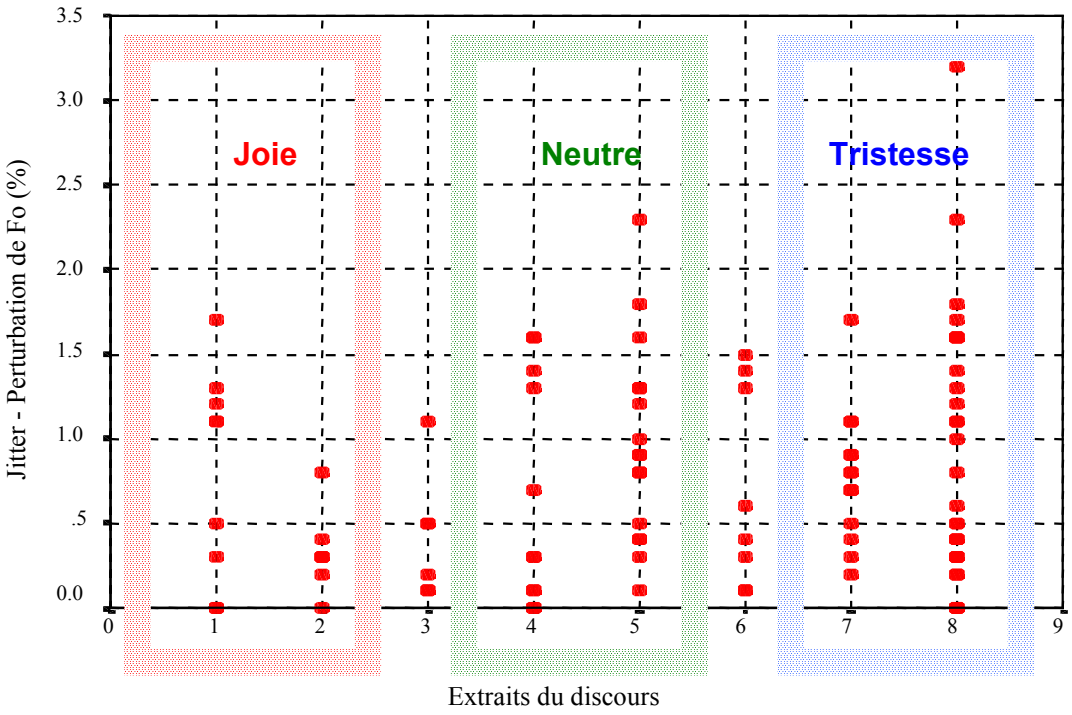


Figure 27. Valeurs de la perturbation de Fo (jitter, %) des énoncés dans les huit extraits du discours[57].

IV.4.1.3. Perturbation d'intensité (Shimmer)

La perturbation d'intensité ('*shimmer*' en anglais) montre la variation de l'intensité, cycle par cycle, dans une unité acoustique. Le shimmer, comme le jitter, varie selon les différentes phonations et influence la perception de la qualité de la voix (comme le degré de rudesse vocale), mais la variation du shimmer est moins pertinente dans l'expression et la perception de l'émotion que la variation du jitter (Heiberger & Horri, 1982 ; Klingholz & Martin, 1985). Les valeurs du shimmer et du jitter sont trouvées corrélées dans l'expérience de Takahashi & Koike (1975). Dans notre analyse, le shimmer est mesuré par l'estimation du pourcentage de la différence des valeurs d'amplitude ('A') entre les cycles adjacents, dont le calcul mathématique peut être exprimé comme suit :

$$\text{shimmer}\ (\%) = \frac{100\,N \sum_{1}^{N-1} |A_{i+1} - A_i|}{(N-1) \sum_{1}^{N} A_i}$$

[57] Le point dans le graphe représente la concentration d'un certain nombre de valeurs identiques.

Le shimmer de l'extrait est estimé comme la moyenne des valeurs de shimmer des énoncés de l'extrait et le shimmer de la catégorie émotionnelle est estimé comme la moyenne des valeurs de shimmer des extraits de la catégorie émotionnelle. Le shimmer de l'extrait 8 est plus élevé que celui des autres extraits mais cette élévation n'atteint pas la significativité statistique ($F(7,92)=0,71$, $p>0,05$) (voir la Figure 28). Les valeurs de shimmer des trois catégories émotionnelles - positive, neutre et négative - sont de respectivement 0,03, 0,03 et 0,07%, dont la différence n'est pas statistiquement significative non plus ($F(2,97)=1,60$, $p>0,05$).

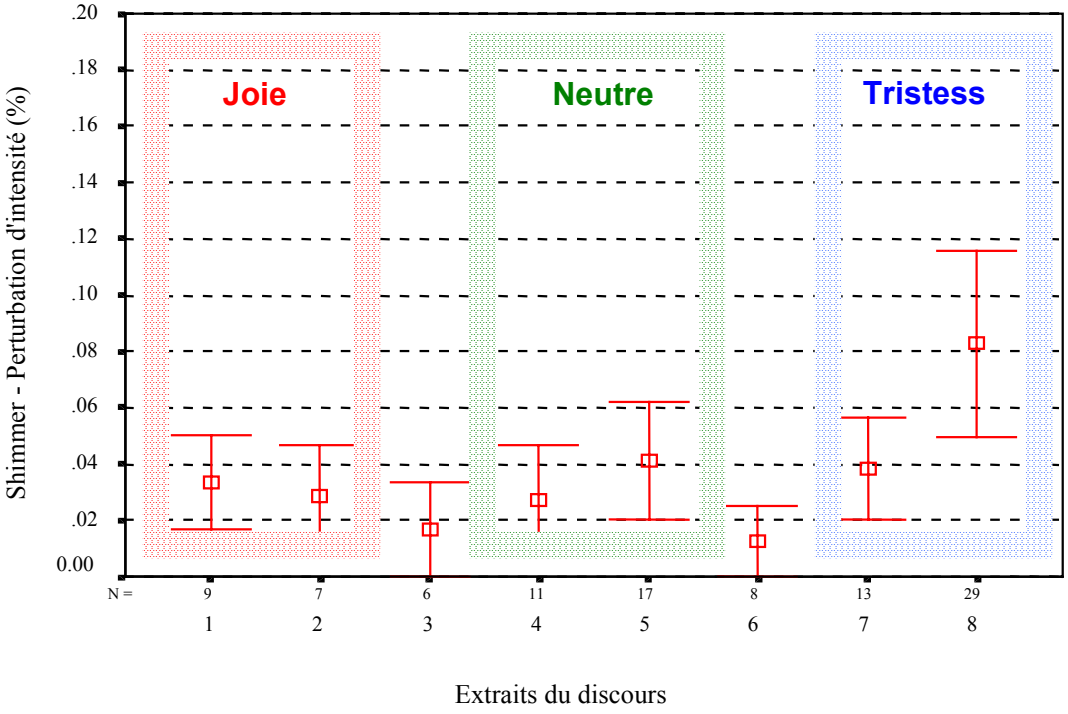


Figure 28. Moyennes et écart-types de la perturbation d'intensité (shimmer, %) des énoncés dans les huit extraits du discours exprimant la joie, la neutre ou la tristesse.

D'après l'observation des valeurs de shimmer individuelles des énoncés dans notre corpus (présentées dans la Figure 28), les valeurs de shimmer de la voix de notre locutrice WJ sont considérablement plus petites que celles trouvées dans les autres études (voir le Tableau 9). Ce fait explique l'absence de différence significative entre les valeurs dans notre résultat, à la différence du résultat de l'expérience de Bachorowski & Owren (1995) dans lequel le shimmer varie de manière significative en fonction de l'expression de l'émotion positive et de l'émotion négative.

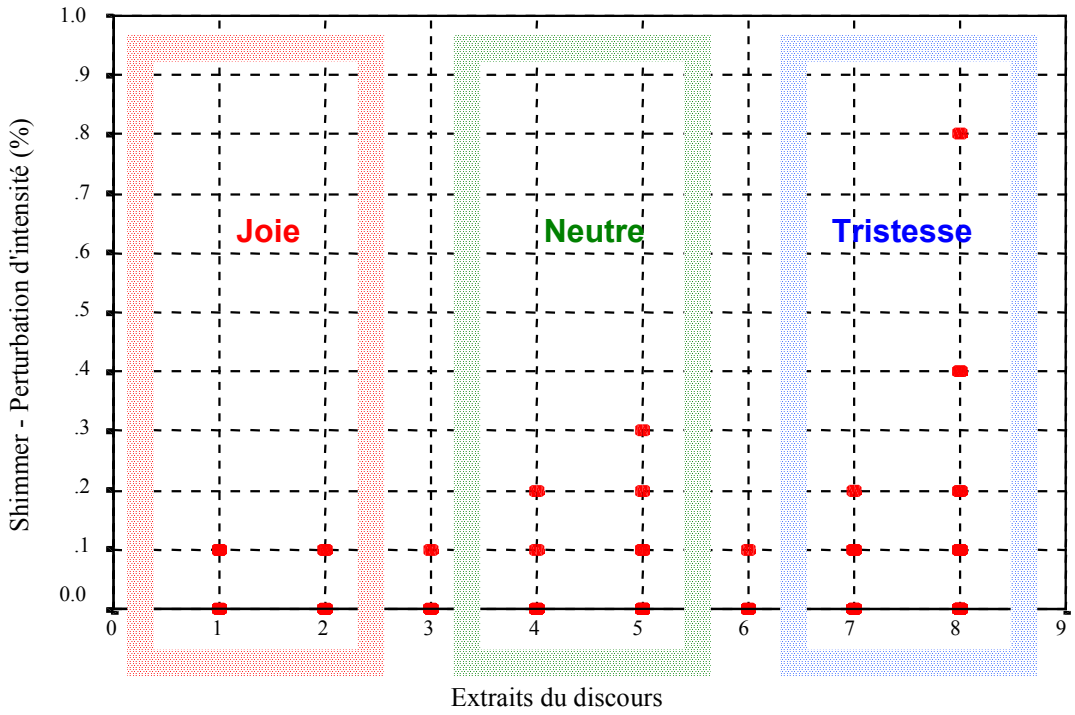


Figure 29. Valeurs de la perturbation d'intensité (shimmer, %) des énoncés dans les huit extraits du discours[58].

IV.4.1.4. Débit de parole

Le débit de parole est l'un des paramètres souvent notés, avec le Fo, dans la description de l'émotion vocale. Fairbanks & Hoaglin (1941) caractérisent la tristesse et le mépris par un débit lent en tant que trait commun mais ils distinguent les différentes sources du débit lent entre les deux émotions. D'après leur analyse, le ralentissement du débit dans l'expression de la tristesse résulte de l'allongement des pauses silencieuses, tandis que le ralentissement du débit dans l'expression du mépris est dû à l'allongement des pauses et à l'allongement des énoncés à la fois. Le débit lent de la tristesse est retrouvé dans la plupart des études suivantes (Davitz, 1964 ; William & Stevens, 1972 ; Fónagy, 1980). La joie est caractérisée par un débit rapide, comme le cas de la colère (Davitz, 1964 ; Fónagy, 1981). L'accélération et le ralentissement du débit en fonction de l'excitation émotionnelle sont démontrés par Arnfield *et al.* (1995). Pourtant, la variation du débit est moins pertinente que celle du Fo dans la perception de la parole en général (Swerts & Geluykens, 1993).

[58] Le point dans le graphe représente la concentration d'un certain nombre de valeurs identiques.

Dans notre analyse, le débit de parole est estimé par le nombre de syllabes par seconde[59]. La Figure 30 montre une tendance descendante du débit de l'extrait 1 vers l'extrait 8, mais la différence du débit entre les extraits n'est pas trouvée significative ($F(7,92)=1,76$, $p>0,05$). Les valeurs de débit des trois catégories émotionnelles - positive, neutre et négative - sont de respectivement 7,0, 6,4 et 6,3, dont la différence n'est pas statistiquement significative non plus ($F(2,97)=2,11$, $p>0,05$).

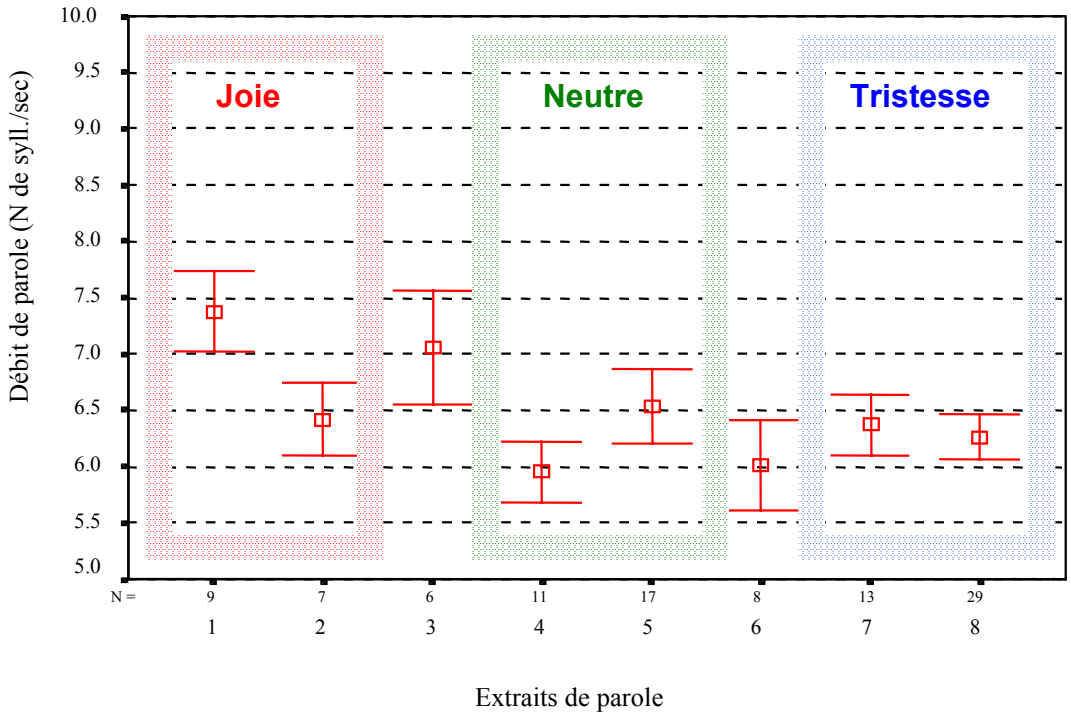


Figure 30. Moyennes et écart-types du débit de parole (nombre de syllabes par seconde) des énoncés dans les huit extraits exprimant la joie, la neutre ou la tristesse.

Malgré l'absence de significativité dans les analyses statistiques, les valeurs de débit de notre corpus confirment les découvertes précédentes de manière globale, à savoir un débit rapide pour l'émotion de la joie et un débit lent pour l'émotion de la tristesse. Le ralentissement du débit de la parole de la locutrice WJ d'un état joyeux vers un état triste devient plus visible dans la variation du débit minimum de l'extrait (voir la Figure 31) que dans la variation du débit moyen de l'extrait (voir la Figure 30). Or, le débit de l'extrait 2 n'est pas aussi rapide que celui de l'extrait 1, même si la même émotion de joie est

[59] Le débit est estimé par le calcul suivant : débit = (nombre de syllabes dans l'énoncé x 1000) / durée de l'énoncé. Par exemple, soit 10 syllabes dans un énoncé d'une durée de 1522ms. Le débit de l'énoncé est de 6,6 (= (10 x 1000) / 1522).

exprimée dans les deux extraits. La Figure 31 montre que la raison de la valeur basse du débit moyen de l'extrait 2 est l'absence de hautes valeurs de débit pour les énoncés de cet extrait, mais pas le rabaissement global du débit des énoncés de cet extrait. L'observation minutieuse des durées syllabiques de l'énoncé nous indique que le débit lent des énoncés de l'extrait 2 est dû particulièrement à l'allongement de la syllabe finale de l'énoncé, plutôt qu'à l'allongement général des syllabes et des pauses de l'énoncé. Cet allongement final de l'énoncé est souvent accompagné d'un contour de Fo montant-descendant, ce qui semble exprimer un sentiment agréable de la locutrice et donner une impression agréable à l'interlocuteur. Le rôle de la durée et du contour de Fo dans la communication émotionnelle est examiné dans notre Expérience 7, ce qui est présenté dans le chapitre VI.

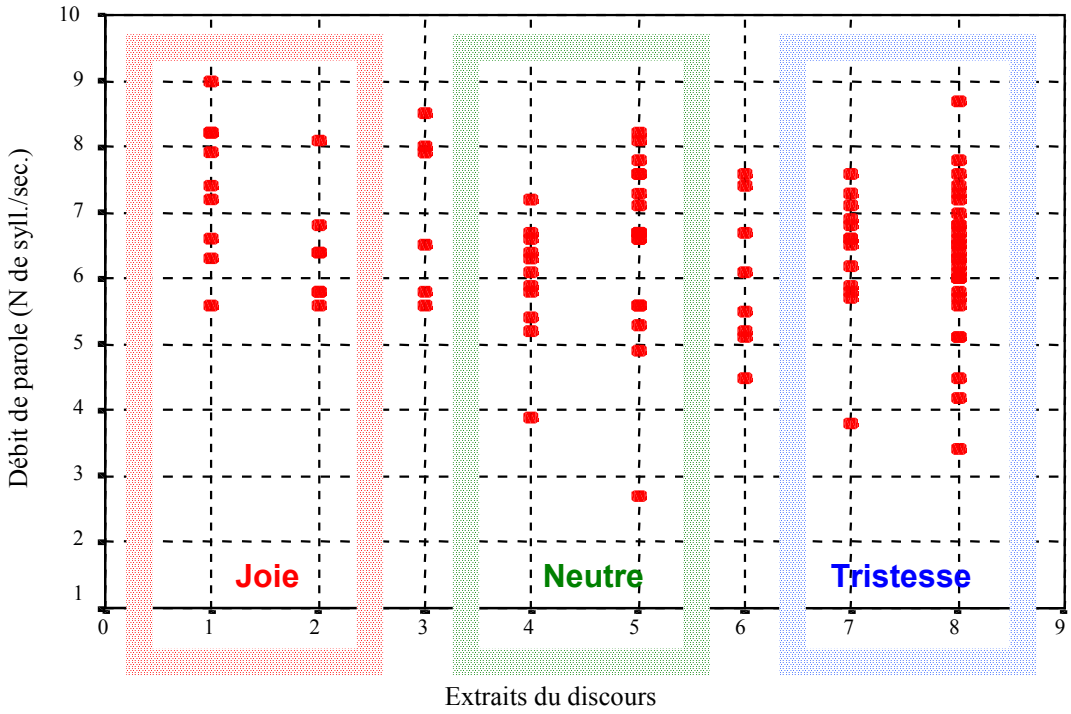


Figure 31. Valeurs du débit de parole (nombre de syllabes par seconde) des énoncés dans les huit extraits du discours[60].

[60] Le point dans le graphe représente la concentration d'un certain nombre de valeurs identiques.

IV.4.1.5. Qualité de la voix (timbre)

La qualité de la voix (‘*timbre*’) des expression émotionnelles est considérée comme essentielle dans l’expression et la perception de l’émotion mais sa mesure quantitative est largement limitée à l’heure actuelle. La description physiologique et acoustique des différentes phonations vocales est proposée par Laver (1980) mais cette description ne porte que sur la voix émotionnellement neutre. Les paramètres de jitter et de shimmer fournissent des valeurs numériques de l’état phonatoire de l’émotion vocale, mais la pertinence de ces paramètres dans l’expression et la perception de l’émotion n’a pas atteint la significativité statistique dans notre analyse. L’analyse spectrale peut servir à l’estimation de la différente qualité vocale entre les différentes expressions émotionnelles (Kaiser,1962 ; Hecker *et al.*, 1968). A partir de l’analyse spectrale des trois paires d’émotions (positive-négative)[61], Kaiser (*ibid.*) caractérise l’émotion négative (comme la tristesse) par un rétrécissement du gosier et un timbre pauvre, et l’émotion positive (comme la tendresse) par un élargissement du gosier et un timbre riche. Cette caractérisation du timbre se base sur l’observation de la quantité d’énergie dans les différentes gammes fréquentielles. D’après Trojan (1948, cité par Kaiser, 1962, p304), le rétrécissement du gosier fait augmenter la fréquence fondamentale et renforce l’énergie dans les hautes fréquences, ce qui rend la voix pauvre en harmoniques. L’un des effets du stress émotionnel (causé par la difficulté de la tâche à accomplir) sur la production vocale est le renforcement de l’irrégularité de la période glottale à la fin de l’énoncé (Hecker *et al.*, 1968, p999). Cet effet est retrouvé dans la voix de notre locutrice WJ quand elle était en détresse[62] à cause de sa situation de trouble (voir la Figure 34).

[61] Les trois paires d’émotion de l’expérience de Kaiser (1962) sont décrites par les adjectifs suivants : gai (‘*cheerful*’) – triste (‘*sad*’), tendre (‘*kind*’) – dur (‘*grim*’) et enthousiate (‘*enthusiastic*’) – dégoûtant (‘*disgusting*’).

[62] La tristesse de la locutrice WJ, exprimées à la fin de l’entretien, est considérée comme de la *détresse*, qui se compose de deux aspects émotionnels par définition, la *tristesse* et le *stress*. La locutrice WJ était triste, parce qu’elle se trouvait dans une situation malheureuse, et en même temps, elle était stressée, frustrée, parce qu’elle se considérait incapable de résoudre le problème elle-même.

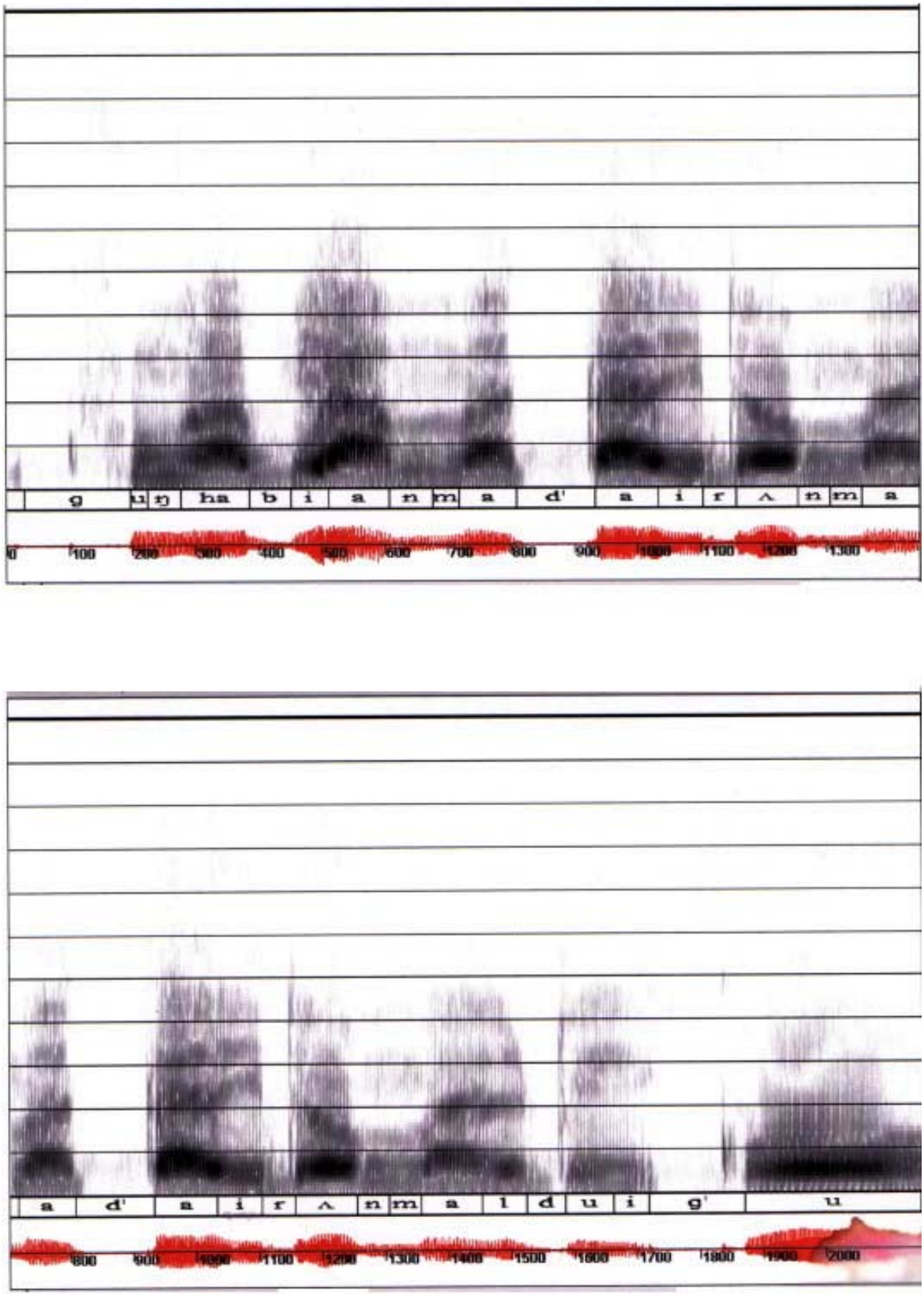


Figure 32. Spectrogramme de l'énoncé de la locutrice WJ, émotionnellement neutre.

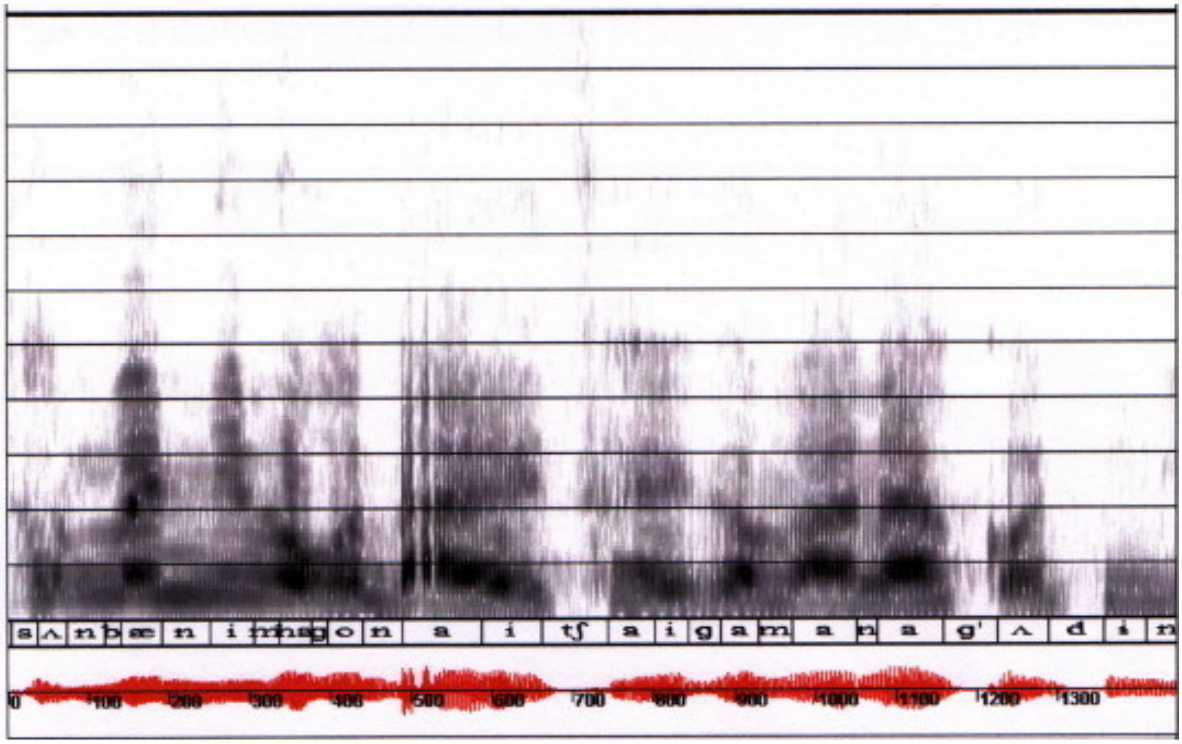


Figure 33. Spectrogramme de l'énoncé de la locutrice WJ, émis dans un état émotionnel positif (joie).

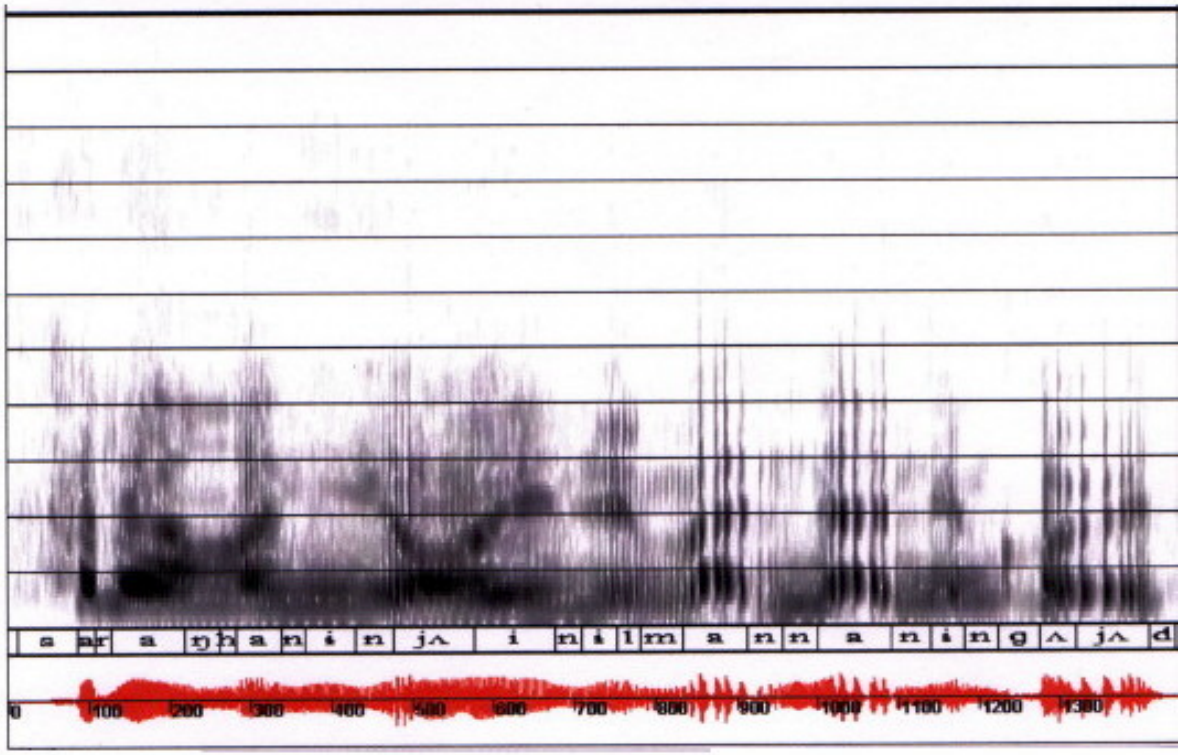


Figure 34. Spectrogramme de l'énoncé de la locutrice WJ, émis dans un état émotionnel négatif (détesse).

IV.4.2. Corrélats acoustiques de la joie et de la tristesse.

Ici, nous discutons les résultats de notre analyse acoustique en termes de corrélats acoustiques de l'émotion positive et de l'émotion négative. Les deux sortes d'émotion sont exprimées dans notre corpus sous les formes fortes de la joie et de la tristesse. Ces formes émotionnelles peuvent être décrites comme respectivement « *un sentiment de grande satisfaction, vif sentiment de plaisir* » et « *un sentiment d'abandon, de solitude, d'impuissance, que l'on éprouve dans une situation poignante* »[63]. Les expressions de ces émotions entraînent les changements des traits prosodiques, ce qui est estimé par nos mesures acoustiques aux niveaux de la fréquence, du temps et de l'intensité. La différence des valeurs acoustiques entre les trois sortes d'émotions, positive, neutre et négative, est considérée comme la différence des propriétés acoustiques de ces émotions, ce qui est montré dans le Tableau 10. Ce dernier consiste en le résumé des résultats de nos mesures acoustiques, exposés dans les parties précédentes (IV.4.1.1 ~ IV.4.1.5). Il montre comment les trois catégories d'émotion, déterminées dans notre expérience 1 (voir la page 102), sont distinguées par la différence des valeurs des paramètres acoustiques.

	Catégories d'émotions		
Paramètres acoustiques	**Positive**	**Neutre**	**Négative**
Fo moyen (Hz)	249,75 *	222,07	229,81
Fo maximum (Hz)	348,94 *	302,60	347,17 *
Fo minimum (Hz)	124,69	109,71	76,55 *
Fo Moy 20% Bas (Hz)	186,69	169,02	127,83 *
Plage de Fo (Hz)	224,25	192,88	270,62 *
Jitter (%)	0,60	0,80	0,90
Shimmer (%)	0,03	0,03	0,07
Débit de parole (N. de syll./sec)	7,00	6,40	6,30

Tableau 10. Moyenne des valeurs acoustiques des énoncés pour les trois catégories d'émotion, positive, neutre et négative.

N.B. Le signe '*' indique la différence significative de la valeur par rapport aux autres, (ANOVA, α=0,05).

[63] Ce sont les définitions de la *joie* et de la *détresse*, présentées dans le dictionnaire 'Le Petit Robert' (1982).

D'après le Tableau 10, l'excitation émotionnelle de l'émotion positive et de l'émotion négative est distinguée de l'état émotionnellement neutre par l'élévation du Fo maximum. L'élévation de la plage de Fo semble être plus ou moins pertinente pour indiquer l'apparition de l'émotion positive ou négative dans la voix. L'émotion positive est distinguée des émotions neutre et négative par l'élévation du Fo moyen, tandis que l'émotion négative est distinguée des émotions neutre et positive par l'abaissement du Fo minimum et du 'Fo Moy Bas'. Les valeurs de jitter ne varient guère entre les émotions positive, neutre et négative. Les valeurs de shimmer de l'émotion négative sont plus élevées que celles des émotions positive et neutre, mais cette élévation n'est pas statistiquement significative. Le débit de parole est plus rapide dans l'expression de l'émotion positive que dans celles des émotions négative et neutre, mais cette différence n'atteint pas la significativité statistique (α=0,05).

IV.5. Analyse perceptive

La partie IV.5 examine la perception de l'émotion. Deux études sont présentées dans cette partie : une analyse statistique sur la relation entre les valeurs acoustiques et les valeurs perceptuelles et une analyse expérimentale sur l'identification de l'émotion par les auditeurs coréens, français et américains. La première étude montre comment la perception de l'auditeur est liée à la variation des valeurs acoustiques. La deuxième étude montre comment la perception de l'émotion varie selon les différentes connaissances linguistiques culturelles des auditeurs.

IV.5.1. Expérience 3 : Relation entre les valeurs acoustiques et les degrés d'émotion perçue

Il est bien connu que l'auditeur perçoit l'émotion du locuteur en se basant sur les indices acoustiques de sa voix. Scherer & Oshinsky (1977), dans l'article intitulé *l'utilisation des indices dans l'attribution de l'émotion à partir des stimuli auditifs*, montrent comment différentes valeurs acoustiques influencent l'évaluation émotionnelle de l'auditeur. Dans leur expérience, ils ont manipulé d'une façon systématique le niveau de Fo (haut-bas), le contour de Fo (montant-descendant), la variation de Fo (grand-petit), la variation d'amplitude (grand-petit) et la variation temporelle (rapide-lent) dans des séquences de tons synthétiques. Puis ils ont demandé aux auditeurs d'évaluer l'émotion de la séquence de ton sur l'échelle d'activation, l'échelle de valence et l'échelle de puissance. D'après leurs résultats, les valeurs d'activation et de puissance augmentent quand la séquence de tons a des propriétés acoustiques comme le haut niveau de Fo, la grande variation de Fo et le débit rapide. En général, plus le Fo augmente, plus le stress émotionnel est perçu par l'auditeur (Streeter *et al.*, 1983 ; Protopapas & Lieberman, 1995 ; Protopapas & Eimas, 1997). Au niveau de la valeur de valence, l'émotion de la séquence de tons est perçue comme plus agréable quand la variation est rapide.

IV.5.1.1. Analyse des données

La contribution des valeurs acoustiques à la perception de l'émotion est examinée par l'estimation de la corrélation entre les valeurs acoustiques et les degrés d'émotion perçue. Il s'agit de la réinterprétation des résultats acoustiques, présentés dans la partie précédente, du point de vue perceptuel. Une série d'analyse de corrélation de Pearson-(α=0,05) est effectuée avec les valeurs acoustiques issues de nos mesures acoustiques et les valeurs perceptuelles, comme les valeurs d'activation et les valeurs de valence, évaluées par les Coréens dans l'expérience 1.

IV.5.1.2. Résultats

Le Tableau 11 montre comment l'auditeur perçoit l'intensité émotionnelle et la positivité émotionnelle sur base des indices acoustiques de la voix du sujet parlant. La corrélation positive entre le Fo moyen et la valeur d'activation indique que plus le Fo moyen de la locutrice est élevé, plus l'intensité émotionnelle[64] est perçue. La corrélation positive entre le Fo maximum et la valeur d'activation est plus forte que celle entre le Fo moyen et la valeur d'activation. C'est parce que la mesure du Fo moyen prend en compte les Fo bas, qui sont négativement corrélés aux valeurs d'activation, tandis que la mesure du Fo maximum n'est pas influencée par ces Fo bas. Ce raisonnement est confirmé par la corrélation significative entre le Fo minimum et la valeur d'activation dans la fonction négative. Cela signifie que plus le Fo minimum est bas, plus l'intensité émotionnelle (la tristesse) est perçue par l'auditeur. L'explication de la corrélation négative entre le 'Fo Moy Bas' et la valeur d'activation est la même que celle de la corrélation entre le Fo minimum et la valeur d'activation. La corrélation positive entre la plage de Fo et la valeur d'activation indique que plus la plage de Fo est grande, plus l'intensité émotionnelle est perçue. Les valeurs de jitter semblent être liées aux valeurs d'activation dans la fonction positive, mais leur lien est trop faible pour affirmer qu'elles sont corrélées. Il est en de même pour la relation entre les valeurs de shimmer et les valeurs d'activation. La corrélation entre le débit et la valeur d'activation est trouvée nulle parce que l'évaluation de la valeur d'activation de notre corpus concerne les deux sortes d'émotion, qui montrent les

[64] Le terme d'intensité émotionnelle est l'équivalent du terme d'excitation émotionnelle dans la présente étude.

débits différents. Le débit est rapide pour l'émotion positive (la joie), dont la valeur d'activation est plus élevée que celle de l'émotion neutre, alors que le débit est lent pour l'émotion négative (la tristesse), dont la valeur d'activation est aussi plus élevée que celle de l'émotion neutre. La combinaison de ces deux corrélation résulte donc en l'absence de corrélation entre le débit et la valeur d'activation.

Paramètres acoustiques	**Valeurs d'activation**	**Valeurs de valence**
Fo moyenne (Hz)	0,22 *	0,04
Fo maximum (Hz)	0,52 *	-0,20 *
Fo minimum (Hz)	-0,58 *	0,50 *
Fo Moy 20% Bas (Hz)	-0,65 *	0,61 *
Plage de Fo (Hz)	0,68 *	-0,40 *
Jitter (%)	0,14	-0,18
Shimmer (%)	0,13	-0,14
Débit de parole (N. de syll./sec)	0,02	0,03

Tableau 11. Coefficient de corrélation de Pearson ('r') des valeurs acoustiques et des valeurs d'activation et de valence, perçues dans l'expérience 1.
N.B. Le signe '*' indique la différence significative de la valeur par rapport aux autres, (ANOVA, α=0,05).

La corrélation entre les mesures acoustiques et la valeur de valence est interprétée principalement comme la relation entre la variation acoustique et la perception de l'émotion négative, en considération du fait suivant. Les énoncés de l'émotion négative (la tristesse) sont plus nombreux que ceux de l'émotion positive (la joie) dans notre corpus coréen, et l'ampleur de la valeur de valence de l'émotion négative est plus grande que celle de l'émotion positive dans les résultats de l'expérience 1 (voir la Figure 18). La corrélation négative entre le Fo maximum et la valeur de valence signifie que plus le Fo maximum de l'énoncé est élevé, plus la valeur de valence de l'énoncé est proche de '-1' (étiquetée comme l'émotion négative) [65]. La corrélation positive entre le Fo minimum et la valeur de valence signifie que plus le Fo minimum est rabaissé, plus la valeur de valence de l'énoncé est proche de '-1' (étiquetée comme l'émotion négative). Ces corrélations sont en accord avec les résultats de notre analyse acoustique, à savoir que l'émotion négative (la tristesse)

[65] Etant donné que les valeurs de valence des émotions, positive et négative, sont de respectivement '+1' et '-1' dans l'échelle de la positivité émotionnelle, la meilleur perception de l'émotion positive est indiquée par l'élévation de la valeur de valence (près de la valeur de '+1') tandis que la meilleur perception de l'émotion négative est indiquée par le rabaissement de la valeur de valence (près de la valeur de '-1').

est exprimée dans la voix de notre locutrice WJ par l'élévation des hautes valeurs de Fo et le rabaissement des basses valeures de Fo. La corrélation positive entre le 'Fo Moy Bas' et la valeur de valence est similaire à la corrélation entre le Fo minimum et la valeur de valence : c'est-à-dire que lorsque le 'Fo Moy Bas' est rabaissé, l'émotion de l'énoncé est perçue comme plus négative. La corrélation négative entre la plage de Fo et la valeur de valence veut dire que plus la plage de Fo est grande, plus la valeur de valence est basse, proche de '-1'(étiquetée comme l'émotion négative)[66]. Le lien possible entre l'élévation du jitter et la perception de l'émotion négative est signalé par le coefficient de corrélation négatif entre les valeurs de jitter et les valeurs de valence, mais ce coefficient est trop faible pour être considéré. Il en est de même pour le lien entre l'élévation du shimmer et la perception de l'émotion négative. Le débit de parole est censé être lié à la valeur de valence d'une fonction positive - l'accélération du débit pour l'émotion positive (dont la valeur de valence est élevée) et le ralentissement du débit pour l'émotion négative (dont la valeur de valence est basse) - mais cette relation n'est guère confirmée dans notre résultat. Cela semble être dû au fait que la variation du débit en fonction de l'émotion est trop faible pour être utilisée en tant qu'indice de l'émotion positive ou négative dans le test de perception (voir IV.4.1.4). L'absence de corrélation entre le Fo moyen et la valeur de valence vient de l'existence de deux corrélations opposées : une corrélation positive entre le Fo moyen (élevé) et la perception de l'émotion positive (dont la valeur de valence est élevée) et une corrélation négative entre le Fo moyen (élevé) et la perception de l'émotion négative (dont la valeur de valence est basse).

IV.5.1.3. Discussion

Les résultats de l'Expérience 3 montrent les fortes corrélations entre les valeurs de Fo et les valeurs d'activation. L'auditeur perçoit plus d'excitation émotionnelle quand le Fo de la voix du locuteur est plus élevé, ce qui est en accord avec les résultats de Streeter *et al.* (1983), de Protopapas & Lieberman (1995) et de Protopapas & Eimas (1997). Le grand coefficient de corrélation entre la plage de Fo et la valeur d'activation confirme la constatation de Hutter (1968) que le degré d'émotion perçue est fortement et positivement corrélé aux degrés de la variation de Fo et de la variation d'intensité. La correspondance entre les valeurs de la plage de Fo et les valeurs d'activation pour les huit extraits de notre

[66] La plage de Fo de l'émotion positive est aussi grande mais cette corrélation est masquée par la majorité des énoncés de l'émotion négative dans notre corpus coréen.

corpus est constatée dans la comparaison entre la Figure 25 et la Figure 17. Les corrélations fortes entre les paramètres concernant les Fo bas (comme le Fo minimum et le 'Fo Moy Bas') et la perception de l'émotion négative (la tristesse dans notre étude) montrent l'importance de la variation des Fo bas dans la perception de la voix pleurante. Les résultats des corrélations entre les valeurs acoustiques et les valeurs perceptuelles reflètent la correspondance entre les changements acoustiques de la voix émotionnelle et la stratégie de l'auditeur dans l'identification de l'émotion vocale.

IV.5.2. Expérience 4 : Identification de l'émotion par les Coréens, les Français et les Américains

L'Expérience 4 examine comment les émotions de joie et de tristesse exprimées par la locutrice WJ sont identifiées par des auditeurs qui ont différentes connaissances linguistiques et culturelles. Cette problématique est liée à la question de la perception de l'émotion, soit universelle, soit spécifique à la culture. Dans cette expérience, un test de l'identification de l'émotion est effectué avec trois groupes d'auditeurs : Coréens, Français et Américains. Les réponses des auditeurs sont comparées au moyen d'un traitement statistique en ce qui concerne les différentes cultures des auditeurs (représentées par leurs différences langues maternelles) et les différentes émotions.

La joie et la tristesse sont considérées comme les émotions primaires[67], et leur expression et leur perception sont considérées comme plus ou moins universelles (Tomkins, 1962, 1963 ; Plutchik, 1980b ; Ortony & Turner, 1990)[68]. En ce qui concerne l'expression émotionnelle dans les différentes langues, Kim (1978) a constaté la correspondance sémantique entre les termes émotionnels coréens et les termes émotionnels anglais[69], et a insisté sur l'aspect multiculturel de l'émotion, appelé l'*universalité* de l'émotion. Van Bezooyen (1984) a étudié l'expression de dix émotions dans les différentes langues européennes[70] et la perception de ces émotions par les Hollandais, les Japonais et les Taiwanais. D'après elle, la communication de l'émotion est universelle, vu que les caractéristiques acoustiques de l'expression émotionnelle sont comparables entre les différentes langues et la plupart des émotions sont identifiées par les auditeurs avec une précision supérieure à celle qui pourrait être obtenue au hasard. En même temps, la communication de l'émotion a aussi un aspect spécifique à la culture, c'est-à-dire que l'identification de certaines émotions varie en fonction de l'origine culturelle de l'auditeur

[67] Ortony & Turner (1990) considèrent la *joie*, la *tristesse*, la *colère* et la *peur* comme les émotions de base qui sont psychologiquement primaires et sociologiquement fréquentes dans les cultures occidentales.
[68] Voir II.3.2 et II.3.3.
[69] Kim (1978) a recueilli une base de données qui se compose de plus de 500 mots émotionnels en coréen et a effectué un test de catégorisation sémantique de ces mots avec les Coréens et les Américains. Dans le résultat, 20 catégories émotionnelles ont été identifiées par le regroupement des mots émotionnels, dont les 13 catégories étaient identiques entre les Coréens et les Américains et les autres catégories étaient aussi similaires entre les deux groupes de sujets.
[70] Les dix émotions étudiées dans les expériences de Bezooyen (1984) sont la *joie*, la *tristesse*, la *colère*, la *peur*, le *dégoût*, la *surprise*, le *mépris*, la *honte*, l'*intérêt* et le *neutre*, exprimées dans les langues comme l'anglais, le français, l'allemand, le hollandais, le hongrois, et le tchèque.

(par exemple, la joie exprimée par les Hollandais a été correctement identifiée par les Hollandais avec un taux de 76%, tandis que les taux d'identification correcte de cette émotion par les Taiwanais et par les Japonais n'étaient que de 24% et 20%, respectivement). A propos de ces résultats apparemment contradictoires[71], Van Bezooyen les explique par deux sortes de raisonnement. D'un côté, les expressions émotionnelles ne sont comparables qu'entre les langues occidentales, donc les caractéristiques acoustiques des émotions, décrites sur base des langues occidentales, ne prédisent pas la perception des émotions par les auditeurs ayant une culture orientale. D'un autre côté, la comparabilité des caractéristiques des émotions vient du fait que ces mesures acoustiques (du Fo moyen, de la plage de Fo, de l'intensité et du débit) sont trop globales pour capter la différence subtile des émotions dans les différentes langues, l'expression et la perception de l'émotion variant entre les différentes cultures. Scherer *et al.* (1988) ont demandé à des Japonais, à des Américains et à des Européens de décrire les expériences émotionnelles (comme la joie, la tristesse, la colère et la peur) : leur cause et leur durée, leurs symptômes physiologiques ressentis et leurs réactions (expressions vocale, faciale et corporelle). Leurs résultats montrent que les expériences émotionnelles des trois groupes sont similaires ; par exemple, la durée des émotions est de l'ordre 'tristesse > joie > colère > peur' ; l'expression émotionnelle de la joie et de la colère sont plus socialisées que celles de la tristesse et de la peur. Pourtant, ils ont aussi trouvé une différence entre les trois groupes d'auditeurs en ce qui concerne la cause et la réaction émotionnelles ; par exemple, les Japonais montrent moins de réactions vocales[72] que les Américains et les Européens ; les Américains sont les plus expressifs parmi les sujets de l'expérience. Ces résultats sont confirmés dans l'étude de Wallbott & Scherer (1988), où des gens de 27 pays de cinq continents ont participé à l'enquête sur des expériences émotionnelles de la joie, de la tristesse, de la colère et de la peur.

Vu ces problématiques de l'universalité de l'émotion au niveau expressif et au niveau perceptuel, nous examinons dans l'Expérience 4 si les émotions de notre corpus, exprimées par la locutrice Coréenne, peuvent être correctement identifiées par les auditeurs étrangers qui ne connaissent pas la langue coréenne. Vu l'expression explicite des émotions de la locutrice WJ et la simplicité de la tâche des auditeurs, nous nous attendons à

[71] Le résultat descriptif des mesures acoustiques des expressions émotionnelles montre un aspect universel ou multiculturel de l'émotion, tandis que le résultat perceptuel de l'identification des émotions par les auditeurs provenant de différentes cultures montre un aspect spécifique culturel de l'émotion.

[72] La réaction faciale de l'émotion est pourtant similaire entre les Japonais, les Américains, et les Européens.

l'identification correcte des émotions par les trois groupes d'auditeurs, Coréens, Français et Américains. Pourtant, l'instabilité de la reconnaissance de la joie par les différents auditeurs, reportée dans les résultats précédents (Van Bezooyen, 1984 ; Chung, 1995, 1999), nous signale une possibilité de divergence de l'identification de l'émotion par les trois groupes d'auditeurs, qui sont culturellement différents. Dans le test de perception, les trois émotions - la joie, la tristesse et le neutre - sont désignées par les termes généraux *émotions positive, neutre et négative* afin d'éliminer la confusion des termes émotionnels spécifiques entre les auditeurs, due à leur différence culturelle (voir la page 96).

IV.5.2.1. Préparation des stimuli

Nous avons sélectionné 15 énoncés de la locutrices WJ, cinq de l'émotion positive, cinq de l'émotion neutre et cinq de l'émotion négative, pour les stimuli du test de perception. La détermination des trois catégories émotionnelles est basée sur le résultats de l'expérience 1 (voir le Tableau 7). Afin d'avoir les stimuli dont l'émotion est exprimée au niveau prosodique (plutôt qu'au niveau lexical), nous avons choisi les énoncés dont la sémantique est relativement neutre, pour les trois catégories émotionnelles. La durée est comparable entre les énoncés, dont la moyenne (écart-type) est de 1948 (428,6)ms[73]. La structure syntaxique[74] n'est pas identique entre les énoncés. Etant donné que les énoncés sont prononcés dans une prise de souffle, il n'y a pas de silence perceptible à l'intérieur de l'énoncé.

(1)	AdvP	NP	VP	
(2)	[gIrEke	mals'lmll	hasigile]	: WJ14 (1591ms)
(3)	ainsi	la parole	a proféré	('puisqu'elle l'a ainsi dit')

(1)	AdP	NP	NP	VP	
(2)	[uri	namdoNseNi	gosami	gEdlnjo]	: WJ817 (1942ms)
(3)	notre	frère	un lycéen	est	('notre frère est un lycéen')

Figure 35. Exemples des stimuli de l'Expérience 4, WJ14 (1591ms) et WJ817 (1942ms).
(1) Etiquetage syntaxique, (2) transcription phonétique de l'énoncé
(3) Equivalent des mots français (Traduction française complète).

[73] La transcription et la durée des stimuli de l'expérience 3 sont présentées dans l'annexe.
[74] La structure syntaxique de la phrase coréenne est en forme de '(sujet) + (object) + (adverbe) + verbe'.
N.B. L'élément entre parenthèses '( )' peut être omis.

IV.5.2.2. Test de perception

Nous avons fait passer le test de perception à trois groupes d'auditeurs : dix Coréens, dix Français et dix Américains. La moitié des auditeurs était des hommes et l'autre moitié des femmes. Ce sont des étudiants, des chercheurs, et des employés d'entreprise des trois pays, la Corée, la France et les Etats-Unis. Les langues maternelles des auditeurs sont le coréen, le français et l'anglais. Leur âge varie entre 22 et 35 ans. Ils se sont portés volontaires pour le test. Les auditeurs ont passé le test de perception individuellement dans différents endroits calmes[75]. La procédure était la même pour tous les auditeurs. La tâche de l'auditeur était de décider quelle sorte d'émotion est exprimée dans le stimulus, en choisissant l'une des trois cases, étiquetées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'[76]. Les instructions étaient écrites en anglais au début du questionnaire comme suit :

« Vous allez écouter des stimuli vocaux qui consistent en segments de la parole d'une locutrices Coréenne. Les stimuli sont des phrases, qui ne sont pas nécessairement d'une structure syntaxique complète. Votre tâche est, après avoir écouté un stimulus, d'identifier quelle sorte d'émotion (positive, neutre ou négative) est exprimée dans ce stimulus. Ne cherchez pas à comprendre le sens du stimulus. Votre décision doit être basée sur votre impression subjective du stimulus sonore, mais pas sur votre compréhension lexicale du stimulus. Dès que votre décision sera prise, mettez une croix ('x') sur l'une des trois cases, marquées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'. »

Les auditeurs ont écouté les stimuli par l'intermédiaire d'un haut-parleur, et cela une fois, dans un ordre aléatoire. Ils ont écrit leur réponse sur un questionnaire pendant trois secondes. Le test a duré environ 15 minutes.

[75] Etant donné qu'il s'agit des stimuli sonores, le test a été effectué la plupart du temps dans une chambre insonorisée.

[76] Le mode de la réponse de cette tâche est le même que celui de la deuxième tâche de nos expériences précédentes (expérience 1 et expérience 2) mais le traitement des réponses de cette expérience est différent de celui des deux expériences précédentes (voir plus loin).

IV.5.2.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 150 réponses des trois groupes d'auditeurs, Coréens, Français et Américains, 450 réponses au total (15 stimuli x 10 sujets x 3 groupes d'auditeurs). Les réponses dont le choix d'émotion correspond à la catégorie émotionnelle de l'énoncé, déterminée au niveau de la sélection des stimuli, sont considérées comme les réponses correctes. Le nombre de réponses correctes d'un sujet donné pour les trois catégories d'émotion est compté individuellement. Ainsi sont construites les données de l'analyse statistique en terme de la **fréquence des réponses correctes**. Ce genre de traitement des données diffère de celui de l'expérience 1, même si les types de réponses (tels l'émotion positive, le neutre et l'émotion négative) sont les mêmes dans les deux expériences. Les réponses des auditeurs dans l'Expérience 4 sont traitées comme les valeurs nominales (donc, il s'agit de la fréquence d'une telle valeur), tandis que les réponses des auditeurs dans l'expérience 1 étaient traitées comme les valeurs continues (donc, il s'agit de la moyenne des valeurs). Cette différence du traitement des données peut être expliquée par la différence du but de la recherche entre les deux expériences. Le but de l'Expérience 4 est de vérifier si l'émotion de l'énoncé, qui est déterminée à l'avance au moment de la sélection des stimuli, est correctement identifiée par les différents groupes d'auditeurs, tandis que le but de l'expérience 1 était de décrire comment l'émotion de l'énoncé est perçue par les auditeurs, sans définition préalable de l'émotion de l'énoncé. En un mot, l'Expérience 4 examine la différence des auditeurs dans l'identification de l'émotion, tandis que l'expérience 1 a examiné la différence des énoncés en ce qui concerne le degré d'émotion perçue.

Les nombres des réponses correctes des trois groupes d'auditeurs sont comparés en fonction des trois catégories émotionnelles, par deux analyses de variance (ANOVAs). Deux facteurs, LANGUE et EMOTION, sont examinés dans les ANOVAs, avec les deux questions suivantes. Est-ce que le nombre de réponses correctes est différent entre les trois groupes d'auditeurs (Coréens, Français et Américains) ? Est-ce que le nombre de réponses correctes est différent entre les trois catégories d'émotion (positive, neutre et négative) ? L'effet du facteur LANGUE sur le nombre de réponses correctes est examiné par une ANOVA inter-sujet (l'ANOVA à un facteur à trois niveaux), parce qu'il s'agit des différents sujets pour les trois niveaux du facteur LANGUE, les Coréens, les Français et les

Américains. L'effet du facteur EMOTION sur le nombre de réponses correctes est examiné par une ANOVA intra-sujets (l'ANOVA à un facteur avec mesures répétées par sujets), parce que chaque auditeur a évalué tous les trois niveaux du facteur EMOTION, positive, neutre et négative[77]. Les deux ANOVAs sont effectuées en même temps, de façon indépendante.

IV.5.2.4. Résultats

L'ANOVA inter-sujets montre que l'effet du facteur LANGUE sur le nombre de réponses correctes est significatif ($F(2,27)=7,93$, $p<0,01$). Ce résultat indique que le nombre de réponses correctes est différent entre les trois groupes d'auditeurs, Coréens, Français et Américains. Les moyennes de leurs réponses correctes sont de respectivement 80,6, 58,6 et 63,4%[78]. D'après le test post-hoc (scheffe, $\alpha=0,05$), les Coréens ont significativement mieux identifié les émotions que les Français et les Américains.

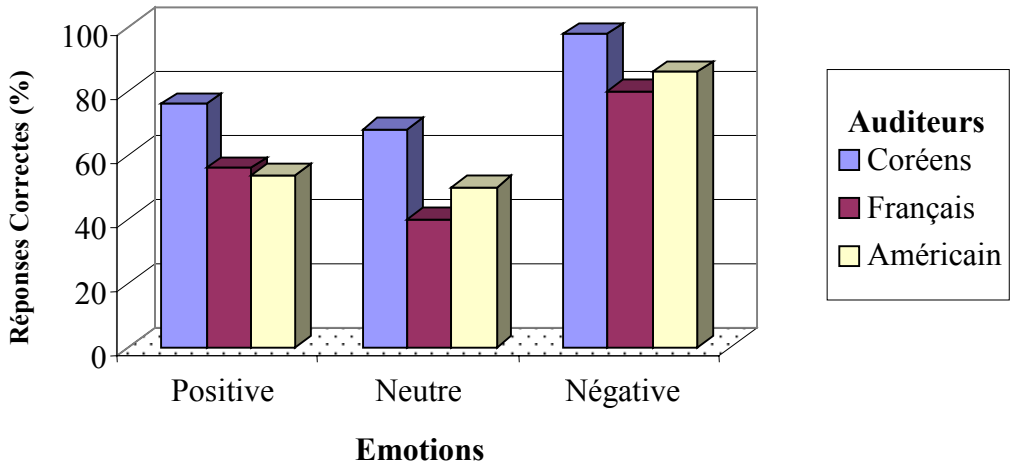


Figure 36. Moyennes des nombres de réponses correctes des trois groupes d'auditeurs, Coréens, Français et Américains, en ce qui concerne les trois catégories émotionnelles, positive, neutre et négative.

L'effet du facteur EMOTION est aussi trouvé significatif dans l'ANOVA intra-sujets ($F(2,54)=16,79$, $p<0,01$), ce qui signifie que le nombre de réponses correctes est différent pour les trois catégories. Les moyennes des réponses correctes pour les trois

[77] Autrement dit, l'auditeur a répété son évalution émotionnelle des stimuli pour les trois niveaux du facteur EMOTION.
[78] Etant donné que 10 auditeurs ont identifié l'émotion de 5 stimuli dans une catégorie donnée, si l'émotion de tous les stimuli était correctement identifiée, la moyenne des réponses correctes du groupe d'auditeurs serait de 50.

catégories d'émotion - positive, neutre et négative - sont de respectivement 62,0, 52,6, 88,0%. Le résultat significatif de l'ANOVA indique que l'émotion négative est significativement mieux identifiée que l'émotion positive et que l'émotion neutre. Pourtant, il ne confirme ni la différence entre l'émotion positive et l'émotion négative ni la différence entre l'émotion positive et l'émotion neutre, parce que la comparaison à posteriori (test post-hoc) ne peut pas être effectuée dans l'ANOVA d'intra-sujets. La moindre valeur de la moyenne de l'émotion neutre signale que cette émotion a été souvent confondue comme l'émotion positive ou l'émotion négative par les auditeurs. L'interaction entre les deux facteurs, LANGUE et EMOTION n'est pas trouvée significative ($F(4,54)=0,25$, $p>0,05$). Cela veut dire que les émotions sont identifiées par les trois groupes d'auditeurs de la même manière. Par exemple, l'émotion négative est toujours la mieux identifiée tandis que l'émotion neutre est toujours la moins bien identifiée.

Afin de voir la distribution des réponses des auditeurs, nous avons construit les matrices de confusion, ce qui est présenté dans le Tableau 12. Nous constatons que les auditeurs étrangers, Français et Américains, ont plus confondu les émotions que les Coréens, surtout en ce qui concerne l'émotion positive et l'émotion neutre. La différence entre les Coréens et les auditeurs étrangers semble être largement due à la divergence de leur identification de l'émotion positive et de l'émotion neutre[79].

		Réponses coréennes		
		P	∅	N
Emotions	P	**38**	10	0
	∅	9	**34**	1
	N	3	6	**49**

		Réponses françaises		
		P	∅	N
Emotions	P	**28**	17	9
	∅	14	**20**	1
	N	8	13	**40**

		Réponses américaines		
		P	∅	N
Emotions	P	**27**	16	3
	∅	13	**25**	4
	N	10	9	**43**

Tableau 12. Matrices de confusion, construites à partir des réponses des auditeurs coréens, français et américains, dans l'identification des émotions positive ('P'), neutre ('∅') et négative ('N').
N.B. La valeur de 50 de la moyenne des réponses correctes représente l'identification de l'émotion parfaite.

[79] Pourtant, la différence entre les Coréens et les Français en ce qui concerne l'émotion négative n'est pas négligeable.

IV.5.2.5. Test supplémentaire

Etant donné que la différence entre les Coréens et les auditeurs étrangers, Français et Américains, est considérablement plus grande que celle entre les deux groupes d'auditeurs étrangers, nous nous sommes demandée si la différence majeure entre les groupes d'auditeurs vient du fait que les Coréens ont compris le sens des énoncés tandis que les Français et les Américains ne l'ont pas compris. Donc, nous avons effectué un test supplémentaire afin de vérifier l'influence lexicale sur l'identification de l'émotion. Nous avons transcrit les stimuli de l'Expérience 4 en coréen et avons demandé à dix nouveaux Coréens (qui n'avaient pas participé à notre test précédent) d'identifier l'émotion des stimuli écrits. La tâche et la procédure de ce test de perception étaient identiques à celui de l'Expérience 4 (voir IV.5.2.2), sauf que les stimuli ont été présentés à l'écrit, au lieu d'à l'audio. Le résultat du test supplémentaire, présenté dans le Tableau 13, montre peu d'influence lexicale sur l'identification des trois émotions. C'est-à-dire que les trois émotions - positive, neutre et négative - ne sont guère distinguées au niveau lexical. Les réponses correctes sont à peine plus nombreuses que les réponses incorrectes en ce qui concerne l'émotion positive et l'émotion négative. Les réponses incorrectes sont même plus nombreuses que les réponses correctes en ce qui concerne l'émotion neutre. La matrice de confusion montre que la sémantique des énoncés sélectionnés pour la catégorie d'émotion neutre est perçue comme émotionnellement plus négative que celle des énoncés sélectionnés pour la catégorie d'émotion négative.

		Réponses coréennes		
		P	∅	N
Emotions	P	**18**	13	15
	∅	17	**11**	16
	N	15	26	**19**

Tableau 13. Matrice de confusion des réponses des Coréens dans l'identification des émotions positive ('P'), neutre ('∅') et négative ('N') avec les stimuli écrits.

N.B. La valeur de 50 de la moyenne des réponses correctes représente l'identification de l'émotion parfaite.

IV.5.2.6. Discussion

D'après l'Expérience 4, le **taux d'identification émotionnelle** (estimé par le nombre de réponses correctes) est différent entre les trois groupes d'auditeurs, Coréens, Français et Américains. Ce résultat est surprenant, vu qu'il s'agit des émotions primaires, comme la joie et la tristesse dont la perception est souvent citée comme universelle, et que la tâche de l'auditeur était assez simple comme d'identifier les émotions dans les catégories générales (les émotions positive, neutre et négative). Ce résultat montre que la connaissance de la langue dans laquelle l'émotion est exprimée joue un rôle essentiel dans l'identification émotionnelle. La connaissance de la langue représente non seulement la compréhension lexicale de l'énoncé mais aussi la connaissance de la culture en général. Le taux d'identification émotionnelle des auditeurs coréens est significativement plus élevé que celui des auditeurs étrangers qui ne connaissent pas la langue coréenne. D'après les matrices de confusion, les deux groupes d'auditeurs étrangers (Français et Américains) ont confondu l'émotion positive et l'émotion neutre d'une manière similaire. D'après le résultat du test supplémentaire, la sémantique des énoncés est peu informative de l'émotion de l'énoncé. Ces considérations étant prises en compte, nous concluons que la meilleure identification des Coréens est principalement due à leur familiarité de la culture coréenne, plutôt qu'à leur compréhension lexicale des énoncés coréens.

Notre résultat confirme l'universalité de la perception de la joie et de la tristesse à travers les différentes cultures, dans le sens que tous les trois groupes ont identifié les émotions avec une précision supérieure à celle qui pourrait être faite au hasard (le nombre de réponses correctes est plus que la moitié des réponses totales). En même temps, notre résultat confirme la spécificité culturelle dans la perception de l'émotion, du fait que les Coréens, qui sont familiers de la culture dans laquelle l'émotion est exprimée, ont mieux identifié les émotions que les Français et les Américains, qui n'ont pas de connaissance de la culture coréenne. Ces résultats sont en accord avec les résultats de Kramer (1963a), de McCluskey *et al.* (1975) et de Van Bezooyen (1984), en ce qui concerne la différente perception de l'émotion parmi les sujets de différentes cultures.

Le résultat de l'émotion négative qui est mieux identifiée que l'émotion positive peut être expliqué par deux raisons. D'un côté, la tristesse est exprimée d'une manière plus

explicite que la joie dans la voix de la locutrice WJ, ce qui a été noté en tant que l'une des caractéristiques de notre corpus (voir IV.2.3 et I.1.1). De l'autre côté, en général, l'émotion négative est mieux identifiée que l'émotion positive. La supériorité de l'identification de l'émotion négative, par rapport à l'identification de l'émotion positive, a été remarquée par les différents chercheurs (voir II.5.2). Par exemple, dans une étude précédente (Chung, 1994, 1995a), nous avons constaté que la joie était souvent confondue avec la colère comme la tendresse était confondue avec la tristesse, tandis que la confusion de la colère pour la joie et de la tristesse pour la tendresse n'est guère arrivée. Hansen & Hansen (1988) et Pratto & John (1991) expliquent cette tendance perceptuelle (*primauté perceptuelle de l'émotion négative*[80]) du point de vue évolutionnaire, par le fait que l'être humain fait attention, de manière prioritaire, aux stimuli négatifs, pour une raison de protection, au cas où les stimuli mettraient sa vie ou son bien-être en danger. Cette tendance psychologique a évolué au cours de l'histoire humaine et se trouve au niveau inconscient de l'homme.

[80] Voir II.5.2.

IV.6. Analyse communicative

IV.6.1. Expérience 5 : Le rôle des différentes parties extraites de l'énoncé dans la communication de l'émotion

L'Expérience 5 s'adresse à la question de savoir comment le stress émotionnel est exprimé dans les différentes parties (initiale, médiane et finale) de l'énoncé. Cette problématique vient de notre observation des énoncés du corpus coréen et des remarques des études précédentes sur la particularité de la partie finale de l'énoncé dans la communication des informations linguistique et paralinguistique. Dans notre observation spectrographique, nous avons constaté que la voix triste (en détresse) devient plus glottalisée vers la fin de l'énoncé (voir la Figure 34). Cette constatation est en accord avec ce que Hecker *et al.* (1968) ont noté à propos de la manifestation du stress émotionnel dans la voix du locuteur. D'après leur analyse spectrographique, le stress émotionnel du locuteur causé par une tâche difficile à accomplir fait ralentir la vibration des cordes vocales du locuteur vers la fin de l'énoncé[81]. La vibration glottale devient aussi plus irrégulière dans la partie finale de l'énoncé. Ces effets peuvent être clairement observés sur un spectrogramme en bande large, par la présence des grands intervalles entre les pulsations glottales et l'irrégularité de ces dernières à la fin de l'énoncé. Un autre effet du stress émotionnel sur la voix est le changement de la quantité d'énergie dans les hautes fréquences. Dans le spectrogramme, le troisième formant et le quatrième formant sont généralement affaiblis dans la voix produite dans une situation stressante par rapport à ceux de la voix neutre.

La partie finale de l'énoncé semble avoir un statut spécial aux niveaux linguistique et paralinguistique. Il est bien connu que la phrase affirmative et la phrase interrogative sont distinguées essentiellement par l'intonation descendante et l'intonation montante de la partie finale de la phrase. La frontière de phrase est signalée par l'allongement des syllabes finales de la phrase et la chute d'intensité en fin de phrase (Vaissière, 1983). Le cliché mélodique[82] est souvent marqué à la fin de l'énoncé (Fónagy, 1983a ; Léon, 1993).

[81] Hecker *et al.* (1968) utilisent deux termes, *énoncé* (‘*utterance*’) et *goupe de souffle* (‘*Breath Group*’), sans distinction.
[82] Le *cliché mélodique* est une mélodie stéréotypée, reproduite toujours de la même manière par la personne. Il peut montrer sa personnalité et sa profession selon le marquage des indices stéréotypiques.

L'acteur exprime les différentes émotions par la variation des traits prosodiques, surtout à la fin de l'énoncé (Chung, 1994, 1995b). Le système de la synthèse vocale peut produire différentes impressions émotionnelles de la voix synthétique, par la manipulation du contour de Fo en différentes formes, qui varient de manière distinctive principalement dans la partie finale de l'énoncé (Mozziconacci, 1998, p103).

Vu ces particularités de la partie finale de l'énoncé dans les communications linguistique et paralinguistique, nous allons examiner, dans l'Expérience 5, l'hypothèse que le bouleversement émotionnel (comme la détresse) de la locutrice WJ serait mieux exprimer et mieux identifié dans la partie finale de l'énoncé que les parties initiale et médiane de l'énoncé. Autrement dit, la détresse de la locutrice WJ serait renforcée vers la fin de l'énoncé et ce stress émotionnel serait plus exprimé et mieux identifié par l'auditeur dans la partie finale, extraite de l'énoncé et présentée en forme isolée, que dans les autres parties, initiale et médiane, extraites du même énoncé et présentée en forme isolée.

IV.6.1.1. La préparation des stimuli

Les stimuli de l'Expérience 5 sont construits à partir des stimuli de l'Expérience 4. Chaque énoncé est divisé, de manière proportionnelle, en trois parties, ce que nous appelons les parties initiale, médiane et finale de l'énoncé (voir la Figure 37). La frontière de la partie correspond à la frontière du mot dans la mesure du possible. Quand le mot est trop long et quand il doit être divisé en deux, en vue du découpage proportionnel de l'énoncé, cela est fait en considération de l'unité lexicale (lexème) et de l'unité phonétique (syllabe). Les morceaux du mot produits à la suite de cette division sont inclus dans les parties concernées, à condition qu'ils rendent une impression auditive naturelle. Si le morceau du mot se trouve isolé au début ou à la fin de la partie, et si sa présence affecte la perception naturelle de la partie, il est soit compris dans l'une des deux parties concernées, soit supprimé entièrement, en fonction de la comparabilité des durées des parties à l'intérieur de l'énoncé. Etant donné que l'Expérience 5 vise à comparer le degré d'expression émotionnelle entre les trois parties de l'énoncé, il est essentiel d'avoir des durées similaires entre les trois parties de l'énoncé. Ainsi, 45 stimuli (15 énoncés x 3 parties) sont préparés pour le test de perception. La durée moyenne (l'écart-type) des stimuli est de 538,2 (150,2) ms. La transcription phonétique et la description de la durée

des énoncés sont présentées dans l'annexe. Le logiciel 'Mev'[83] est utilisé pour toutes les manipulations acoustiques dans la préparation des stimuli de l'Expérience 5.

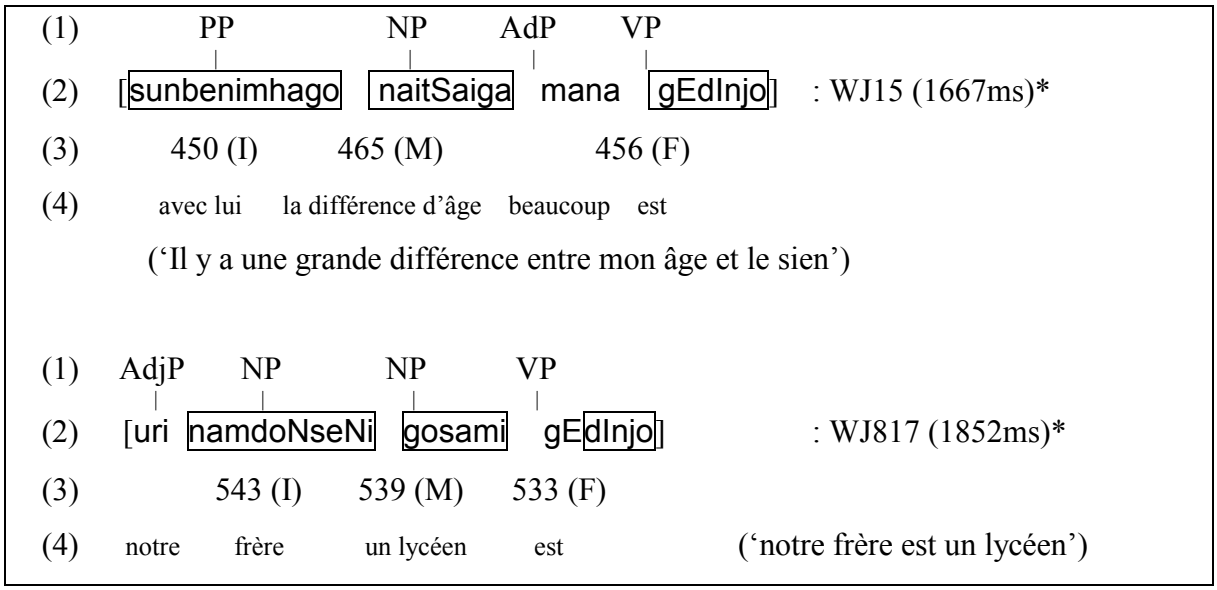


Figure 37. Exemples du découpage de l'énoncé en trois parties, Initiale, Médiane et Finale.
* Les énoncés sont prononcés dans une prise de souffle (voir IV.5.2.1).
(1) Etiquetage syntaxique, (2) Transcription phonétique de l'énoncé, (3) Durées des parties, (4) Equivalent des mots français (Traduction française complète).

IV.6.1.2. Test de perception

Nous avons effectué le test de perception avec trois groupes d'auditeurs : dix Coréens, dix Français et dix Américains, qui n'avaient pas participé à notre test précédent. La moitié des auditeurs était des hommes et l'autre moitié des femmes. Ils sont étudiants universitaires dans les trois pays, la Corée, la France et les Etats-Unis. Leurs langues maternelles sont respectivement le coréen, le français et l'anglais. Leur âge varie de vingt-cinq ans à trente-huit ans. Ils se sont portés volontaires pour le test. La tâche de l'auditeur consistait à identifier l'émotion de la partie. C'est-à-dire, l'auditeur devait choisir l'une des trois cases, étiquetées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'[84].

[83] La description du programme 'Mev' se trouve dans la note n°31.

« Vous allez écouter les stimuli vocaux, qui consistent en segments de la parole d'une locutrice Coréenne. Les stimuli se composent d'un ou deux mots, extraits des phrases individuelles. Ils sont relativement courts et ils ne sont pas nécessairement compréhensibles en forme isolée. Votre tâche est d'identifier, après avoir écouté un stimulus, quelle sorte d'émotion (positive, neutre ou négative) est exprimée dans ce stimulus. Ne cherchez pas à trouver le sens du stimulus. La décision doit être basée sur votre impression subjective du stimulus sonore, mais pas sur votre compréhension lexicale du stimulus. Dès que la décision sera prise, mettez une croix ('x') sur l'une des trois cases, marquées comme 'émotion positive,' 'neutre (non-émotionnel)' et 'émotion négative'. »

Les stimuli étaient présentés aux auditeurs par l'intermédiaire d'un haut-parleur, et cela une fois dans un ordre aléatoire. Les auditeurs ont répondu sur le questionnaire pendant une seconde. Ils ont passé le test individuellement dans des endroits insonorisés. Chaque test a duré environ 15 minutes.

IV.6.1.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 450 réponses des trois groupes d'auditeurs, Coréens, Français et Américains, 1350 réponses au total (45 stimuli x 10 sujets x 3 groupes d'auditeurs). Les réponses dont le choix d'émotion correspond à la catégorie émotionnelle déterminée au moment de la sélection des stimuli sont considérées comme les réponses correctes[85]. Les nombres de réponses correctes des trois groupes de stimuli, partie initiale, partie médiane et partie finale, sont pris en tant que données à analyser. Les données sont subdivisées en trois sous-ensembles selon les catégories d'émotion - positive, neutre et négative - et les ANOVA sont effectuées pour chaque catégorie d'émotion, de façon indépendante.

La variation du nombre de réponses correctes en fonction des trois parties de l'énoncé est examinée par ANOVA à un facteur avec mesures répétées par sujets. La dernière analyse, appelée ANOVA intra-sujets, est choisie parce que les auditeurs ont répété la tâche d'identification émotionnelle pour tous les trois niveaux du facteur PARTIE

[84] La tâche de l'auditeur et la procédure du test de perception dans l'Expérience 5 étaient les mêmes que celles appliquées à l'Expérience 4.
[85] Le traitement des réponses de l'Expérience 5 est identique à celui de l'Expérience 4.

(la partie initiale, la partie médiane et la partie finale). Vu le résultat significatif du facteur LANGUE dans l'Expérience 4, la variation du nombres de réponses correctes en fonction des trois groupes d'auditeurs est examinée de nouveau dans l'Expérience 5, au moyen de l'ANOVA à un facteur (LANGUE) à trois niveaux (Coréens, Français et Américains). La dernière, étant l'une des ANOVA inter-sujets, est choisie parce qu'il s'agit des différents sujets pour les trois niveaux du facteur LANGUE. Les deux ANOVAs sont effectuées à la fois pour une catégorie d'émotion donnée, et l'interaction des deux facteurs, PARTIE et LANGUE, est estimée à l'intérieur de la catégorie. Au total, six ANOVAs sont menées pour les trois sous-ensembles de données (les trois catégories d'émotions, positive, neutre et négative). Les résultats des ANOVAs sont comparés au fur et à mesure dans notre présentation.

IV.6.1.4. Résultat

En ce qui concerne la catégorie d'émotion négative, l'ANOVA intra-sujets montre un effet significatif du facteur PARTIE sur le nombre de réponses correctes ($F(2,54)=30,11$, $p<0,01$). Vu les moyennes des réponses correctes des trois parties, initiale, médiane et finale, qui sont de respectivement 70,0, 72,6 et 92,6%, le résultat significatif de l'ANOVA indique que le nombre de réponses correctes est significativement plus élevé quand l'émotion est identifiée avec les stimuli extraits de la partie finale de l'énoncé qu'avec les stimuli extraits de la partie initiale ou de la partie médiane de l'énoncé.

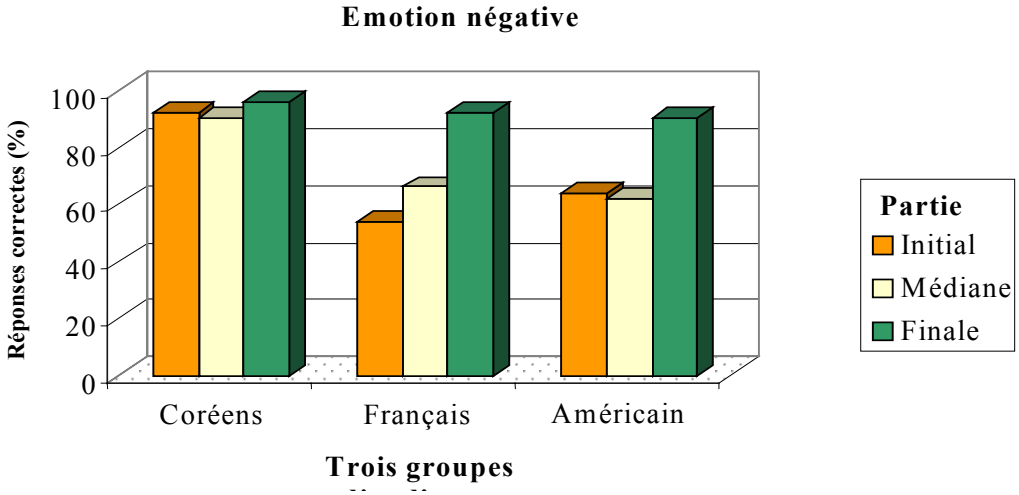


Figure 38. Moyennes des nombres de réponses correctes des trois groupes d'auditeurs (Coréens, Français et Américains) pour les trois groupes de stimuli de parties (initiale, médiane et finale), en ce qui concerne l'émotion négative.

L'effet du facteur LANGUE est aussi trouvé significatif dans l'ANOVA inter-sujets ($F(2,27)=9,54$, $p<0,01$). Les moyennes des réponses correctes des trois groupes d'auditeurs, Coréens, Français et Américains, qui sont de 92,6, 70,6 et 72,0% indiquent que les Coréens ont significativement mieux identifié l'émotion des stimuli que les Français et les Américains. Ce résultat de l'Expérience 5 avec les stimuli de parties de la phrase correspond à celui de l'Expérience 4 avec les stimuli de phrases (voir IV.5.2.4).

L'interaction entre les deux facteurs, PARTIE et LANGUE, est significative ($F(4,54)=4,83$, $p<0,01$), à cause du fait suivant. Le nombre de réponses correctes est significativement plus élevé dans la partie finale que dans les parties initiale et médiane, en ce qui concerne les groupes d'auditeurs français et américains, tandis qu'il n'en est pas ainsi pour le groupe d'auditeurs coréens. C'est à cause de l'effet plafond dans ce dernier cas[86].

En ce qui concerne la catégorie d'émotion neutre, l'ANOVA intra-sujets montre un effet significatif du facteur PARTIE sur le nombre de réponses correctes ($F(2,54)=15,91$, $p<0,01$). Ce résultat semble être le même que celui du facteur PARTIE dans la catégorie d'émotion négative, mais en fait il est l'inverse du résultats de l'effet du facteur PARTIE précédent. Les moyennes des réponses correctes des trois parties, initiale, médiane et finale, de respectivement 69,4, 70,6 et 44,0%, indiquent que le nombre de réponses correctes est significativement plus bas quand l'émotion est identifiée avec les stimuli, extraits de la partie finale de l'énoncé, qu'avec les stimuli extraits de la partie initiale ou de la partie médiane de l'énoncé. L'observation des réponses individuelles nous indique que l'émotion des stimuli extraits de la partie finale des énoncés de la catégorie d'émotion neutre est souvent identifiée (de façon confuse) comme positive, au lieu de neutre. Les moyennes des réponses correctes des trois groupes d'auditeurs, Coréens, Français et Américains, sont différentes (67,4, 49,4 et 67,4%) mais leur différence n'est pas statistiquement significative d'après l'ANOVA inter-sujets ($F(2,27)=2,64$, $p>0,05$). L'interaction entre les deux facteurs PARTIE et LANGUE n'est pas trouvée significative ($F(4,54)=0,31$, $p>0,05$).

[86] Les nombres de réponses correctes des auditeurs coréens sont tous élevés pour les trois parties, initiale, médiane et finale, donc la différence entre les trois parties n'est pas visible.

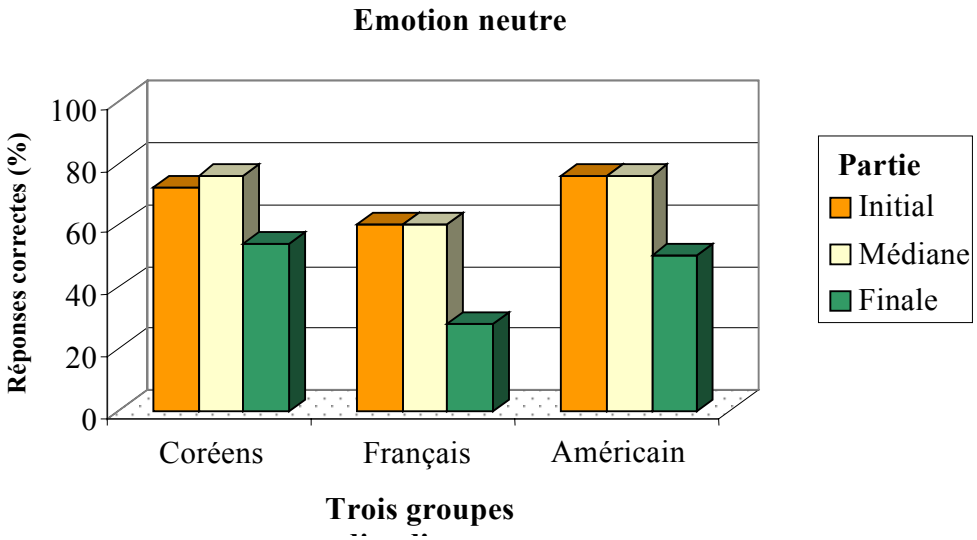


Figure 39. Moyennes des nombres de réponses correctes des trois groupes d'auditeurs (Coréens, Français et Américains) pour les trois groupes de stimuli de parties (initiale, médiane et finale), en ce qui concerne l'émotion neutre.

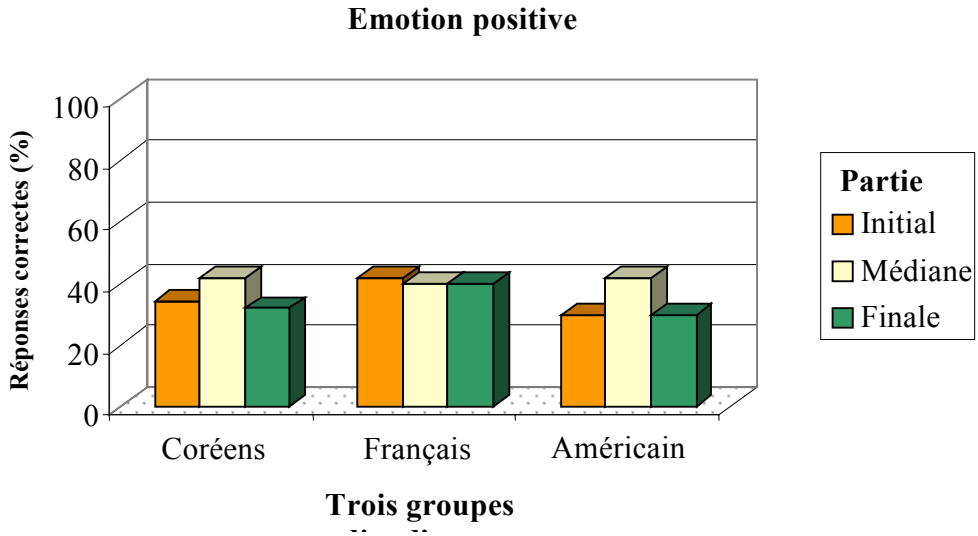


Figure 40. Moyennes des nombres de réponses correctes des trois groupes d'auditeurs (Coréens, Français et Américains) pour les trois groupes de stimuli de parties (initiale, médiane et finale), en ce qui concerne l'émotion positive.

En ce qui concerne la catégorie d'émotion positive, les ANOVAs ne montrent ni effet du facteur PARTIE ($F(2,54)=1,42$, $p>0,05$), ni effet du facteur LANGUE ($F(2,27)=0,34$, $p>0,05$), ni interaction entre les deux facteurs ($F(4,54)=0,49$, $p>0,05$). Les moyennes des réponses correctes sont de 35,4, 41,4 et 43,0%, respectivement pour les trois parties, initiale, médiane et finale, et de 36,0, 40,6, 43,0%, respectivement pour les trois groupes d'auditeurs, Coréens, Français et Américains. Aucune différence n'est trouvée significative en termes statistiques.

IV.6.1.5. Discussion

D'après l'Expérience 5, le taux d'identification correcte (estimé par le nombre de réponses correctes) de l'émotion négative est significativement plus élevé quand les stimuli sont extraits de la partie finale de l'énoncé que quand ils sont extraits de la partie initiale ou de la partie médiane de l'énoncé. Ce résultat confirme notre hypothèse, en indiquant que l'émotion de la détresse est exprimée de manière plus explicite dans la partie finale de l'énoncé que dans les autres parties, initiale et médiane. Cette meilleure expression de la détresse dans la partie finale est liée à la manifestation caractéristique du stress émotionnel (des pleurs en particulier dans cette expérience) à la fin de l'énoncé, ce qui a été constaté par Hecker *et al.* (1968). Malgré de nombreuses remarques sur l'importance de la partie finale dans l'expression et la perception des informations linguistiques et paralinguistiques, aucune étude, à notre connaissance, n'a expérimenté, de manière systématique, la primauté de la partie finale dans l'expression de l'émotion de détresse. Le résultat de notre Expérience 5 est donc un apport dans ce domaine de travail.

Quand on compare les résultats de l'Expérience 5 avec les résultats de l'Expérience 4, on constate que le taux d'identification correcte est beaucoup plus bas avec les stimuli de parties qu'avec les stimuli de phrases, surtout en ce qui concerne la catégorie d'émotion positive. Cette diminution du taux d'identification correcte signale que le découpage de l'énoncé en trois parties, initiale, médiane et finale, aurait détruit la structure informationnelle de l'émotion positive. L'absence de différence significative entre les trois parties indique que l'émotion positive de la locutrice WJ est exprimée par la configuration globale des traits prosodiques au cours de l'énoncé, plutôt que par le changement vocal dans un endroit spécifique de l'énoncé. Or, l'émotion négative est aussi bien identifiée avec

les stimuli de parties finales qu'avec les stimuli d'énoncés, parce que la vibration des cordes vocales, rallongée et irrégulière, apparaît surtout à la fin de l'énoncé.

En ce qui concerne les stimuli de la catégorie d'émotion neutre, il est intéressant de remarquer que l'émotion neutre est correctement identifiée comme neutre quand il s'agit des stimuli extraits de la partie initiale et de la partie médiane des énoncés, tandis qu'elle est perçue comme positive, plutôt que comme neutre, quand il s'agit des stimuli extraits de la partie finale des énoncés. D'après notre observation acoustique des énoncés de la catégorie d'émotion neutre, la syllabe finale de l'énoncé est souvent rallongée (sans glottalisation) avec une intonation montante suivie d'une intonation descendante (le contour de Fo *montant-descendant*), ce qui est typique dans l'expression de la politesse par les jeunes Coréennes, surtout dans la région Séoulienne. Ces traits prosodiques semblent être responsables de la perception de la positivité dans la partie finale de l'énoncé, bien que l'énoncé entier soit identifié comme émotionnellement neutre. La contribution respective de ces traits prosodiques à la perception de l'émotion positive et de l'émotion négative est mise en lumière par une expérimentation avec la synthèse de parole (voir le chapitre VI).

Ici, deux sortes de changement acoustique dû à l'effet émotionnel peuvent être distinguées. Le changement acoustique dans les énoncés de la catégorie d'émotion neutre semble être volontaire, motivé par l'intention de sujet parlant d'exprimer une telle attitude (l'émoion du locuteur vis-à-vis de l'interlocuteur ; la politesse dans ce cas), tandis que le changement acoustique dans les énoncés de la catégorie d'émotion négative (la détresse) semble être involontaire, se manifestant dans la voix malgré l'intention du sujet parlant de refouler son bouleversement émotionnel. Ce gendre de différence d'expression peut être lié au fait que les gens sentent plus ou moins à l'aise d'exprimer leur émotion positive tandis qu'ils essaient de ne pas exprimer leur émotion négative dans l'interaction sociale. Dans le même fil d'idées, l'expression de l'émotion positive vis-à-vis l'interlocuteur peut être considérée comme une configuration des traits prosodiques par la conscience du locuteur pour exprimer son intention positive, tandis que les traits prosodiques de l'expression de la détresse n'est qu'un des résultats de la perte de contrôle du sujet parlant sur sa production vocale. Vu que ces changements acoustiques volontaires et involontaires sont accentués à la fin de l'énoncé et que ces phénomènes influencent la perception de l'émotion, nous concluons que la partie finale de l'énoncé a un statut spécial dans la communication émotionnelle au niveau biologique et au niveau social.

IV.7. Conclusion du chapitre IV.

Dans le chapitre IV, nous avons montré comment l'émotion est exprimée dans la parole spontanée en coréen et comment elle est perçue par différents groupes d'auditeurs ayant différentes connaissances linguistiques et culturelles. Nous avons aussi démontré que l'émotion de détresse est mieux exprimée dans la partie finale de l'énoncé que dans les autres parties, initiale et médiane, ce qui constitue un des aspects originaux du présent travail. Ici, nous nous proposons de résumer les résultats des expériences présentés dans le chapitre IV, et de mettre en valeur les apports de notre étude.

Le chapitre IV s'est déroulé en trois parties majeurs, concernant trois tâches d'investigation majeures ; (1) identifier les caractéristiques acoustiques de l'émotion de la joie et de la tristesse, exprimées dans la parole spontanée ; (2) examiner la perception de l'émotion par les auditeurs dont la langue maternelle est différente; et (3) examiner la contribution des trois parties de l'énoncé, initiale, médiane et finale, à la communication de l'émotion. D'abord, le corpus coréen est décrit dans la partie IV.2 en tant que données d'émotions vécues exprimées dans le discours spontané de la locutrice WJ pendant un entretien télévisé. Le choix de ce type de corpus est expliqué par notre intention d'analyser des données authentiques qui représentent le mieux la réalité de la parole émotionnelle.

Dans la partie IV.3, les expressions émotionnelles du corpus coréen sont décrites en terme d'intensité et de positivité émotionnelles avec des valeurs perceptuelles. Ces valeurs sont issues de deux tests de perception ; l'un dans lequel les auditeurs coréens ont évalué le degré d'émotion (valeur d'activation) et la positivité d'émotion (valeur de valence) des énoncés, en écoutant des stimuli vocaux (Expérience 1), l'autre dans lequel les auditeurs coréens ont accompli les mêmes tâches, en lisant des stimuli écrits (Expérience 2). D'après ces expériences, les émotions de la joie et de la tristesse de la locutrice WJ sont explicitement exprimées dans la voix, mais elles sont peu exprimées dans la lexique des énoncés. Cette étape de recherche a été conçue dans le but de fournir une description objective des caractéristiques du corpus coréen, basée sur des jugements multiples d'un certain nombre de sujets au lieu d'un jugement singulier de l'expérimentatrice.

La partie IV.4 consiste en des mesures acoustiques des énoncés, comme le Fo moyen, le Fo maximum, le Fo minimum, le 'Fo Moy Bas,' la plage de Fo, le jitter, le shimmer, le débit de parole. La propriété spectrale des énoncés n'est pas mesurée en valeurs numériques, elle est estimée par l'observation des spectrogrammes des énoncés. D'après notre analyse acoustique, la joie se manifeste par l'augmentation des valeurs de Fo en général, tandis que la tristesse (exprimée sous forme de détresse) se manifeste par l'augmentation des hautes valeurs de Fo et la diminution des basses valeurs de Fo en même temps. Cette forme excitée de la tristesse se distingue d'une autre forme calme de la tristesse, qui se caractérise par la diminution du Fo en général (ce qui a été noté dans la plupart des études précédentes). D'après l'observation des spectrogrammes des énoncés, la vibration glottale ralenti et devient irrégulière à la fin de l'énoncé dans le cas de l'émotion de détresse. En termes des corrélats acoustiques des émotions, l'augmentation du Fo moyen est significative pour repérer l'apparition de l'émotion positive (la joie), alors que celle du Fo minimum et du 'Fo Moy Bas' est significative pour repérer l'apparition de l'émotion négative (la tristesse). Le Fo maximum et la plage de Fo sont des bons indices de l'excitation émotionnelle, soit pour l'émotion positive soit pour l'émotion négative. Les paramètres comme le jitter, le shimmer et le débit varient en fonction des émotions, mais leur variation n'est pas statistiquement significative.

La partie IV.5 présente deux analyses concernant la perception de l'émotion. Dans la première analyse (Expérience 3), le degré du lien entre les valeurs acoustiques et les valeurs perceptuelles (valeurs d'activation et valeurs de valence) est estimé par une analyse de corrélation. Les valeurs fréquentielles (du Fo) sont fortement corrélées aux degrés d'intensité émotionnelle perçue, tandis qu'il n'en est pas ainsi pour les valeurs temporelles ou les valeurs d'intensité. Dans la deuxième analyse (Expérience 4), l'universalité de la perception de l'émotion par les auditeurs de différentes cultures est examinée par un test de perception. Dans le test, trois groupes d'auditeurs, Coréens, Français et Américains, ont identifié les émotions produites par une Coréenne (la joie, la tristesse, et le neutre) en catégories générales (les émotions positive, neutre et négative). Les résultats du test de perception montrent que le taux d'identification correct des Coréens est significativement plus élevé que les taux d'identification correct des auditeurs étrangers (Français et Américains), tandis que les derniers ne diffèrent pas de manière significative. Ces résultats ne sont pas surprenants, vu que les stimuli d'émotions sont les extraits de la parole de la locutrice coréenne (donc, la familiarité culturelle des auditeurs coréens donne une

meilleure identification des émotions). Pourtant, ces résultats sont inattendus, étant donné qu'il s'agit des émotions primaires dont la perception est censée être universelle, et que la tâche de l'auditeur dans le test de perception est tellement simple et générale qu'on s'attendrait à ce que les taux d'identification soient similaires entre les trois groupes d'auditeurs. En conclusion, les résultats de l'Expérience 4 montrent que les émotions de la joie et de la tristesse peuvent être identifiées de façon globale par les auditeurs de différentes cultures mais que la connaissance culturelle de l'auditeur joue un rôle crucial dans l'identification de l'émotion en ce qui concerne le degré de précision.

Dans la partie IV.6, la particularité de la partie finale dans l'expression et la perception de l'émotion est démontrée par la combinaison de l'analyse acoustique et du test de perception (Expérience 5). L'observation spectrographique montre que la vibration glottale ralentit et devient irrégulière en fin d'énoncé produit au moment où la locutrice WJ est en détresse. Le test de perception montre que l'émotion de détresse est mieux identifiée avec un morceau de parole extrait de la partie finale de l'énoncé qu'avec un morceau de parole extrait de la partie initiale ou de la partie médiane de l'énoncé. Or, il n'en est pas ainsi pour son émotion de joie qui ne montre pas de tel changement acoustique en fin d'énoncé et dont l'identification ne dépend pas des différentes parties de l'énoncé. Quant aux énoncés de la catégorie d'émotion neutre, la syllabe finale est allongée (sans glottalisation) avec une intonation montante suivie d'une intonation descendante. La partie finale de ce genre d'énoncés est souvent perçue comme émotionnellement positive, tandis que leurs parties initiale et médiane sont perçues comme émotionnellement neutres, ce qui est considéré comme l'identification correcte de ces parties, puisqu'elles sont extraits des énoncés de la catégorie d'émotion neutre. L'allongement final combiné avec un contour de Fo montant-descendant connote une notion de politesse dans la jeune génération coréenne (surtout chez les Séouliennes), ce qui explique la perception de la positivité dans la partie finale de l'énoncé en ce qui concerne les stimuli de la catégorie d'émotion neutre. Vu que les changements acoustiques volontaires et involontaires sont mis en accent à la fin de l'énoncé et que ces phénomènes influencent la perception de l'émotion, nous concluons que la partie finale de l'énoncé a un statut spécial dans la communication émotionnelle, au niveau biologique et au niveau social.

Les caractéristique marquantes de notre étude, présentées dans le chapitre IV, sont les suivantes. Premièrement, notre étude se base sur des données qui se composent

d'émotions vécues (exprimées à partir de l'expérience réelle), tandis que la plupart des autres études de l'émotion se basent sur des données qui consistent en des émotions stylisées (imitées par un acteur professionnel ou non-professionnel). Deuxièmement, il est démontré dans notre étude que le taux de précision de l'identification émotionnelle varie en fonction de la connaissance culturelle de l'auditeur, même dans le cas des émotions primaires (comme la joie et la tristesse). Cet aspect a été aussi remarqué dans quelques autres études précédentes, mais il est souvent négligé du fait que les émotions comme la joie, la tristesse, la colère et la peur, peuvent être identifiées par les sujets de différentes cultures de manière globale (avec un taux de précision plus haut que celui qui serait dû au hasard). En ne niant pas ce dernier fait, nous insistons, d'après le résultat de notre expérience, sur le fait que le taux de réponses correctes est significativement différent entre les auditeurs natifs (Coréens) et les auditeurs étrangers (Français et Américains), surtout en ce qui concerne l'identification de la joie. Troisièmement, nous proposons la *primauté de la partie finale de l'énoncé dans la communication émotionnelle*. L'émotion de détresse du sujet parlant est exprimée d'une manière plus explicite dans la partie finale de l'énoncé que dans les autres parties initiale et médiane de l'énoncé. En termes des évidences acoustiques et perceptuelles, la vibration des cordes vocales s'est ralentie et est devenue irrégulière à la fin de l'énoncé dans l'expression de la détresse, et cette dernière émotion a été mieux identifiée avec les stimuli extraits de la partie finale de l'énoncé qu'avec les stimuli extraits de la partie initiale ou de la partie médiane de l'énoncé. Il semble que l'émotion positive du locuteur vis-à-vis de l'interlocuteur est aussi mieux communiquée dans la partie finale de l'énoncé que dans les autres parties initiale et médiane de l'énoncé. Cette émotion positive a été marquée par un allongement de la syllabe finale de l'énoncé et une intonation *montant-descendante* dans le cas de notre locutrice coréenne, et les auditeurs ont reconnu cette émotion dans la partie finale de l'énoncé lors du test de perception. En termes de Bühler (voir II.2.1), ces résultats indiquent que la partie finale de l'énoncé joue un rôle essentiel en tant que *symptôme* et en tant que *signal* dans la communication émotionnelle.

Chapitre V

Etude comparative avec un corpus anglais

Résumé

Ce chapitre présente une étude comparative avec un corpus anglais. Cette étude s'adresse spécifiquement à la manifestation de la détresse dans la voix et à la reconnaissance de cette émotion en fonction des parties initiale, médiane et finale de l'énoncé. Dans la partie V.2, nous avons décrit l'acquisition et les caractéristiques des données à analyser. Le corpus anglais consiste en des extraits des discours spontanés de cinq locutrices Américaines dans une série d'entretiens télévisés. Le contenu du discours et la nature des entretiens sont comparables à ceux du corpus coréen. L'entretien s'est fait avec une locutrice à la fois, et chaque enregistrement de l'entretien contient des moments où la locutrice parlait calmement (émotion neutre) et des moments où elle était tellement bouleversée qu'elle a fini par pleurer (voix larmoyante). Dans la partie V.3, une analyse acoustique a été effectuée sur 92 énoncés du corpus anglais. La Fo moyenne augmente et le débit de parole se ralentit quand le sujet parle en détresse, ce qui est similaire aux résultats de l'analyse acoustique du corpus coréen. La détresse des locutrices américaines a aussi causé une vibration glottale irrégulière ralentie vers la fin de l'énoncé. D'après une expérience perceptive dans la partie V.4, cette vibration irrégulière et ralentie dans la partie finale de l'énoncé a rendu une meilleure identification de la détresse dans cette partie, par rapport aux parties initiale et médiane de l'énoncé. Le taux d'identification émotionnelle avec la partie finale de l'énoncé seule (présentée en forme isolée) était presque comparable à celui avec l'énoncé entier. Ainsi, la primauté de la partie finale de l'énoncé dans l'expression et la perception de l'émotion fut confirmée dans le cas du corpus anglais.

V. Etude comparative avec un corpus anglais

V.1. Préliminaire

Le chapitre V présente une mini étude, dont la nature des données et le sujet d'analyse sont comparables à ceux de l'étude présentée dans le chapitre précédent. Cette étude porte sur un corpus anglais de l'expression émotionnelle, acquis à partir de la parole spontanée de cinq locutrices Américaines. Deux questions sont examinées dans cette étude : comment le *bouleversement émotionnel* (‘*emotional upset*’ en anglais) du locuteur est-il exprimé dans la voix ? et comment l'auditeur reconnaît-il cet état émotionnel du locuteur avec les différents parties de l'énoncé ? Le bouleversement émotionnel ici désigne un état émotionnel où le locuteur est tellement stressé par la tristesse qu'il finit par pleurer[87]. Cette étude avec le corpus anglais s'intéresse spécifiquement à cette émotion négative, alors que l'étude avec le corpus coréen porte sur les deux pôles émotionnels, l'émotion positive (comme la joie) et l'émotion négative (comme la tristesse). Chronologiquement, cette étude du corpus anglais a été menée avant celle du corpus coréen, et elle a servi à la dernière en tant qu'étude préliminaire sur l'expression et la perception de l'émotion dans la parole spontanée.

Ce chapitre consiste en deux parties majeures, une partie descriptive et une partie expérimentale. Dans la première partie, la procédure de l'acquisition du corpus anglais et les caractéristiques acoustiques des énoncés du corpus sont décrites de manière brève. Dans la deuxième partie, une expérience est présentée sur la perception du bouleversement émotionnel dans les différentes parties de l'énoncé.

[87] C'est ce qui a été défini comme l'émotion de *détresse*, en tant que forme excitée de la tristesse (dans IV.4.1.1). Le terme général *bouleversement émotionnel* est employé dans la présentation de l'étude avec le corpus anglais, comme les termes généraux *émotion positive*, *émotion neutre* et *émotion négative* sont employés dans la présentation de l'étude avec le corpus coréen. Le terme *bouleversement émotionnel* est préféré au terme *émotion négative* dans la présentation de l'étude avec le corpus anglais, parce que ce dernier ne contient pas de données qui constituent la contrepartie de l'émotion négative, c'est-à-dire l'émotion positive.

V.2. Sur le corpus anglais

Le corpus anglais se compose des échantillons de parole de cinq locutrices américaines, extraits de l'enregistrement d'entretiens télévisés. L'entretien est tiré d'une émission de télévision américaine (*National Broadcasting Company*), intitulée '*Selly*'. Dans cette émission, le locuteur est invité à parler de ses problèmes personnels, dans le but de trouver une solution à ces problèmes à travers la discussion publique. Diverses émotions vocales et faciales sont exprimées dans la parole selon son état émotionnel. Etant donné que les locuteurs de cette émission ne sont ni acteurs professionnels ni personnes connues du public et qu'il ne montrent aucune tentative de cacher ou exagérer leur émotion, leur expression émotionnelle est considérée comme naturelle et authentique.

V.2.1. Acquisition des données

Les entretiens de cinq locutrices ont été choisis pour notre corpus anglais et ont été enregistrés sur des vidéocassettes VHS par nous-mêmes. Pour l'enregistrement, la sortie du signal de la télévision de la marque *SONY (Trinitron)* était branchée sur l'entrée du signal du magnétoscope de la marque *HITACHI (DA4, MA423)*. Chaque entretien consiste en 20 minutes de la conversation entre une locutrice et une présentatrice. Etant donné qu'il s'agit d'une émission directe, des bruits extérieurs comme bruits de mouvements corporels du sujet parlant et applaudissements de l'auditoire sont inclus dans l'enregistrement. Il existe un bruit de fond mais son niveau n'est pas élevé.

Le choix de la langue anglaise et le choix de locutrices, au lieu de locuteurs, s'explique par la disponibilité des données. L'étude étant menée pendant le séjour de l'expérimentatrice aux Etats-Unis, les données en anglais de l'expression émotionnelle étaient abondantes et faciles à acquérir et l'analyse sur les données facilitait la discussion du travail avec les chercheurs américains. Le choix des sujets féminins n'était pas fixé à priori mais il a été décidé après l'observation des données potentielles. L'expression émotionnelle des femmes était plus régulière et plus cohérente que celle des hommes dans les entretiens observés.

V.2.2. Locutrices

Le corpus anglais consiste en discours de cinq locutrices américaines, qui ont été enregistrées durant plusieurs jours en février 1997. Les critères de sélection de ces locutrices sont identiques à ceux explicités dans la partie IV.2.2. En ce qui concerne les facteurs de l'identité sociale, les locutrices sont des femmes au foyer, ayant la trentaine et parlant l'anglais de l'est[88]. En ce qui concerne les facteurs situationnels, les discours des cinq locutrices sont comparables, du fait qu'ils sont produits en mode spontanée, non lue et qu'ils sont médiatisés en direct au moyen de la télévision. Le sujet de leurs discours est informel, il concerne les problèmes avec des membres de la famille. Les locutrices se sont présentées à cette émission en vue d'obtenir un conseil de la part de quelqu'un qui connaît ces problèmes.

En ce qui concerne les facteurs psychologiques, la comparabilité de l'expression émotionnelle entre les locutrices est prise en compte en tant que critère principal. Toutes les cinq locutrices ont commencé la conversation dans un état émotionnellement neutre (sans émotion particulière) et elles ont fini par pleurer à un moment donné, en racontant des problèmes avec leur mari ou leur enfant. Le trait commun de leur histoire est que la locutrice souffre du mauvais comportement de son mari ou de son enfant et qu'elle veut s'en sortir avec une aide publique. Les pleurs[89] de la locutrice sont considérés comme l'indice de son état émotionnel. Ayant une conscience sociale, les locutrices semblaient se retenir et éviter de pleurer devant le public jusqu'à un certain point, puis elles fondaient en larmes au moment où elles ne pouvaient plus se retenir. Cet état émotionnel marqué est identifié comme l'émotion de la tristesse (exprimée en forme de détresse), alors que l'état émotionnel non-marqué au début de l'entretien est identifié comme l'émotion neutre[90]. L'attitude de la locutrice vis-à-vis du présentateur pendant l'entretien peut être caractérisée comme ouverte et communicative.

[88] Etant donné que l'émission *SELLY* ne fournit pas d'information détaillée sur l'identitié personnelle de l'invité, l'identification des facteurs sociaux des locutrices est basée notre observation de l'émission et le contenu du discours.

[89] Darwin (1872, 153) a considéré les pleurs comme l'expression primaire de la souffrance, soit du type de la douleur physique, soit du type de détresse mentale. Il a fait une remarque intéressante : les larmes n'apparaissent pas dans le cri du nouveau-né, tandis que plus tard dans la vie, les larmes sont l'expression la plus explicite de l'homme, en ce qui concerne son état psychologique ou son état biologique.

[90] Voir la note n° 27 pour la définition de l'émotion *neutre*.

V.2.3. Segmentation des données

Le corpus anglais fut analysé de manière sélective. A partir de l'entretien enregistré sur la vidéocassette, nous avons prélevé une minute de conversation, produite au moment où chaque locutrice parlait calmement sans émotion particulière, et une minute de conversation produite au moment où la locutrice parlait en larmes 'sous le coup de l'émotion de détresse'. Ainsi, dix extraits de conversation ont été sélectionnés en tant que données à analyser (2 extraits de conversation x 5 locutrices). Ces extraits ont été numérisés de manière directe, du magnétoscope à l'ordinateur (*PC Pentium 1*). La numérisation a été faite à l'échantillonnage de 22kHz, en utilisant le logiciel 'Mev'[91].

Ayant le signal acoustique de l'extrait de conversation, nous avons segmenté le signal de parole de la locutrice en énoncés, en excluant le signal de parole de l'interlocuteur. Les énoncés de la locutrice ont été identifiés sur base des mêmes principes que ceux appliqués à la segmentation des énoncés de notre corpus coréen (voir IV.2.5). D'abord, la frontière des énoncés s'est imposée au début et à la fin de la prise de parole de la locutrice. Puis la frontière de la phrase syntaxique est devenue la frontière de l'énoncé, à moins que la pause entre deux phrases soit de moins de 100ms. Une phrase a été divisée en deux si elle contenait une pause plus longue que le seuil de coupure (voir la Figure 41). Dans la segmentation des énoncés anglais, le seuil de coupure était fixé à 400ms. A la suite de cette segmentation, nous avons obtenu 92 énoncés de cinq locutrices dans le corpus anglais. Ces énoncés ont été segmentés en unités plus petites, comme les syllabes et les phonèmes. La durée moyenne (écart-type) des énoncés totaux est de 1281,6 (383,1) ms. La durée moyenne (écart-type) des énoncés pour chacune des cinq locutrices est présentée dans le Tableau 14, avec les valeurs du Fo moyen et du débit moyen dont l'estimation est décrite dans la partie suivante.

[91] Voir la note n° 31 pour la desription du logiciel 'Mev'.

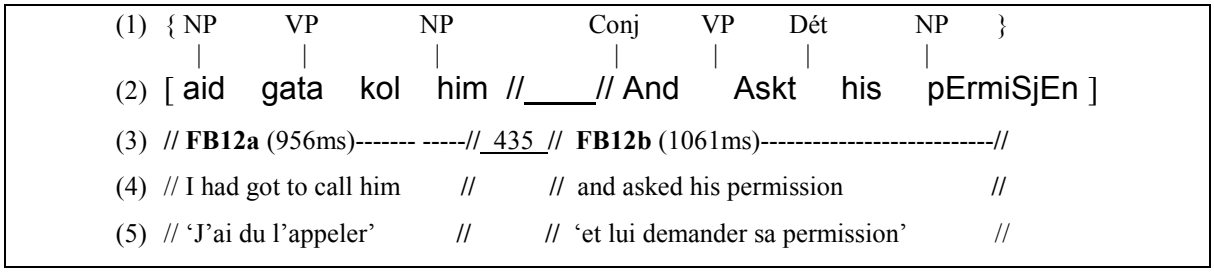


Figure 41. Exemple de la segmentation des énoncés par une pause plus longue que le seuil de coupure. ** '//' indique la frontière de l'énoncé.

(1) Etiquetage syntaxique : NP (syntagme nominal), VP (syntagme verbal), Conj (conjonction), Dét (déterminant). ** '{ }' indique l'unité de la phrase syntaxique.
(2) Transcription phonétique des énoncés anglais.
(3) Durées (ms) des énoncés **FB12a** et **FB12b** et de la pause silencieuse.
(4) Transcription des énoncés en alphabets anglais
** Les segments, omis dans l'énonciation, sont reconstruits dans cette transcription.
(5) Traduction française des énoncés.

Locutrice	**A ('FB')**		**B ('FE')**		**C ('FG')**		**D ('FH')**		**E ('FI')**	
Emotion	**E**	**N**	**E**	**N**	**E**	**N**	**E**	**N**	**E**	**N**
Durée Moy	1160,6	1250,1	1015,1	982,5	1369,9	1427,8	1481,4	1441,1	1519,8	1149,0
(E.T.)	(275,7)	(328,9)	(168,0)	(205,1)	(279,3)	(479,0)	(400,6)	(364,5)	(601,8)	(291,1)
Fo Moy.	298,1	252,0	230,6	201,3	231,3	203,4	233,4	191,1	293,5	201,4
(E.T.)	(36,2)	(34,3)	(11,0)	(22,3)	(19,2)	(25,0)	(39,4)	(12,9)	(50,1)	(16,7)
Débit Moy	5,0	5,4	4,4	5,0	5,6	6,6	5,2	6,2	4,3	5,8
(E.T.)	(0,4)	(1,2)	(1,2)	(1,6)	(0,7)	(1,3)	(1,3)	(0,7)	(1,3)	(0,3)
N	13	13	7	6	8	10	9	9	8	9

Tableau 14. Les durées moyennes (Ecart-Type), les Fo moyens (Ecart-Type) et les débits moyens (Ecart-Type), des énoncés neutres ('N') et des énoncés émotionnels de détresse ('E'), extraits de la parole de cinq locutrices Américaines, A, B, C, D, E.

V.3. Analyse acoustique

L'analyse acoustique du corpus anglais ne porte que sur deux mesures comme le Fo moyen et le débit, qui ont été prises en tant que paramètres représentatifs respectivement du domaine fréquentiel et du domaine temporel. Le changement spectral de la voix au cours du temps est examiné sur base du spectrogramme des énoncés. Le spectre du signal d'une partie de l'énoncé est aussi pris en compte dans cette analyse, afin d'examiner les propriétés spectrales de la voix à un moment particulier. Les valeurs acoustiques sont comparées entre la voix émotionnelle et la voix neutre, et leur différence est interprétée comme le changement de la voix dû à l'émotion de détresse.

V.3.1. Mesures acoustiques

Les valeurs du Fo moyen ont été prélevés à l'aide du logiciel 'Pitch'[92] dans le laboratoire phonétique à l'université de Brown aux Etats-Unis. La comparaison des Fo moyens entre les énoncés neutres et les énoncés émotionnels (détresse) montre que le Fo moyen est significativement plus élevée pour les énoncés émotionnels que pour les énoncés neutres ($F(1,82)=51,53$, $p<0,01$). Cette tendance est trouvée pour toutes les cinq locutrices américaines, même si leurs Fo moyens diffèrent considérablement les uns des autres ($F(4,82)=17,07$, $p<0,01$). Le niveau d'augmentation du Fo sous le coup d'émotion varie aussi en fonction des locutrices ($F(4,82)=3,01$, $p<0,05$) (voir la Figure 42).

La différence de débit entre les énoncés neutres et les énoncés émotionnels est aussi significative ($F(1,82)=17,22$, $p<0,01$). La détresse fait ralentir le débit de parole des cinq locutrices américaines. Le débit moyen est différent pour les cinq locutrices ($F(4,82)=4,56$, $p<0,01$) mais le ralentissement du débit par la détresse est comparable entre les locutrices ($F(4,82)=1,18$, $p>0,05$) (voir la Figure 43).

[92] Le logiciel 'Pitch' est dévéloppé par Mertus (1985) dans le department des sciences cognitives et lingustiques à l'université de Brown. Ce logiciel est une version originelle du logiciel 'Mev,' dévelopé par la même personne (1992). Pourtant, il est muni de plus de fonctions que ce dernier, en ce qui concerne l'analyse fréquentielle.

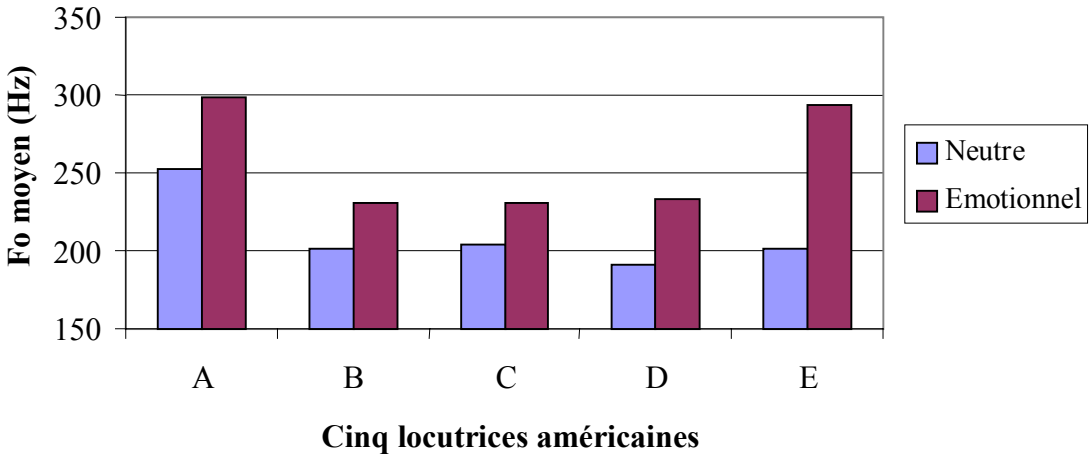


Figure 42. Fo moyens des énoncés neutres et des énoncés émotionnels (détresse) de cinq locutrices Américaines, A, B, C, D et E.

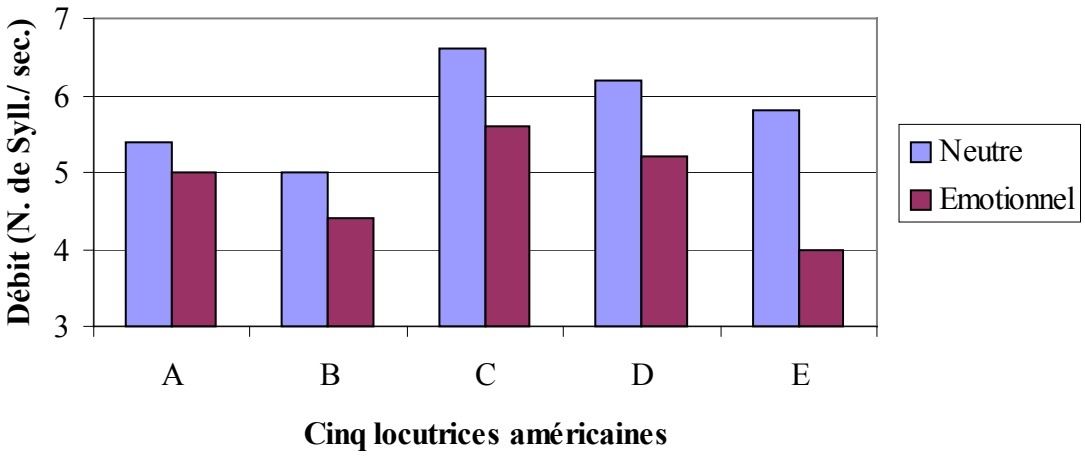


Figure 43. Débits moyens des énoncés neutres et des énoncés émotionnels (en détresse) de cinq locutrices Américaines, A, B, C, D et E.

Le spectrogramme des énoncés montre que les énoncés émotionnels, produits en cas de la détresse, se terminent souvent par une longue syllabe, glottalisée, à la différence des énoncés neutres (voir la Figure 44 et la Figure 45). Ces changements vocaux sont comparables à la voix de la locutrice WJ, produite en cas de détresse (voir la Figure 34). Ils ont été aussi observés par Hecker *et al.* (1968). D'après leur observation, la période glottale devient allongée et irrégulière vers la fin de l'énoncé, quand le sujet parle dans une situation stressante.

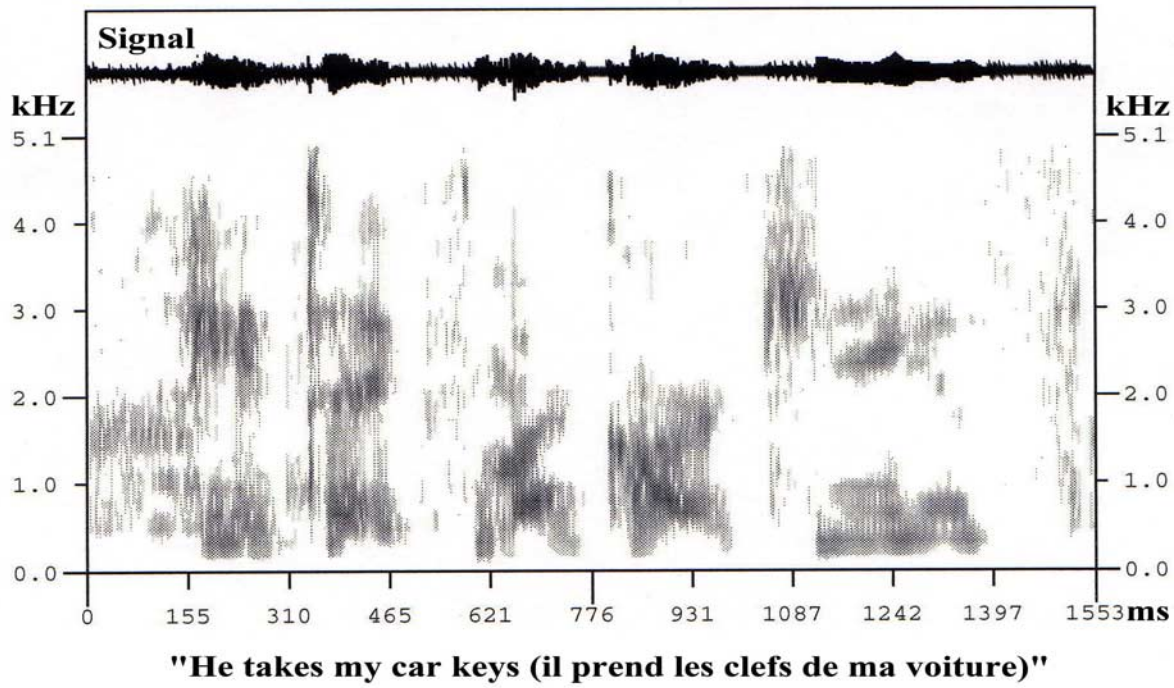


Figure 44. Spectrogramme de l'énoncé de la locutrice A, émis dans un état émotionnel neutre.

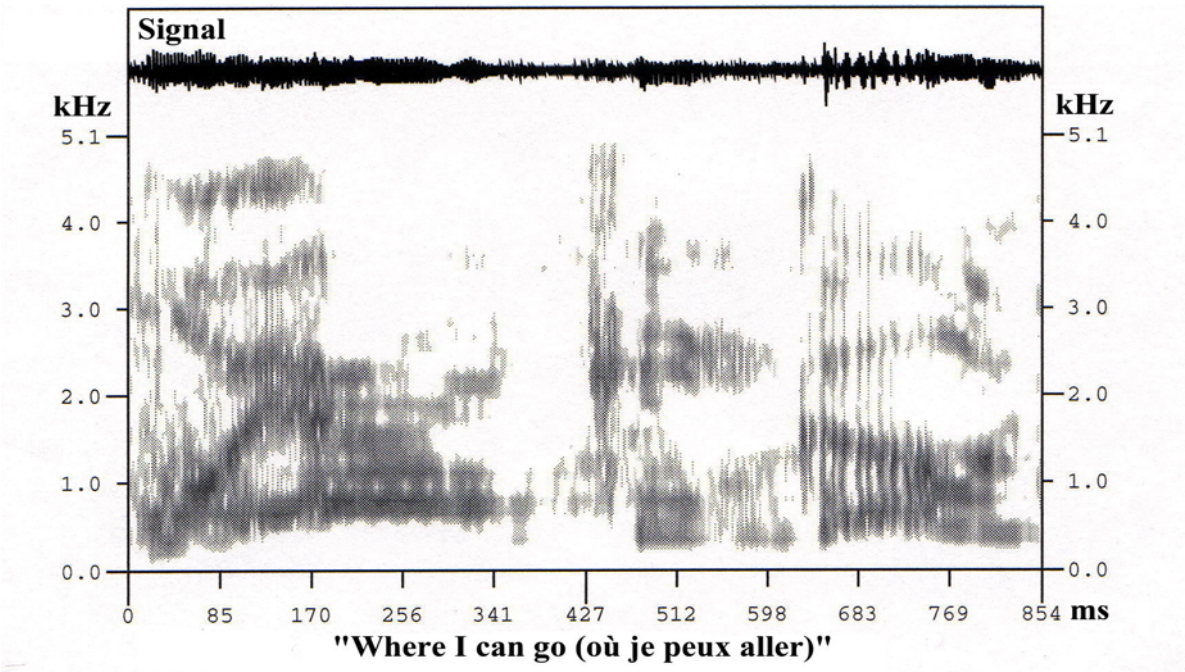


Figure 45. Spectrogramme de l'énoncé de la locutrice A, émis dans un état émotionnel négatif (en détresse).

La comparaison des spectres des mêmes voyelles, prononcées dans deux situations différentes, neutre et émotionnelle (en détresse), montre que la voix émotionnelle est plus pauvre en harmoniques que la voix neutre (voir la Figure 46 et la Figure 47). Cette différence rappelle ce que Kaiser (1962) a trouvé dans la comparaison du timbre vocal entre l'émotion positive et l'émotion négative. Selon lui, le timbre vocal de l'émotion négative est pauvre en harmoniques à cause du rétrécissement faucal, tandis que celui de l'émotion positive est riche en harmoniques à cause de l'élargissement faucal[93].

V.3.2. Corrélats acoustiques de l'émotion de détresse

Les résultats de l'analyse acoustique sur les données anglaises montrent que le Fo augmente et que le débit ralentit quand la locutrice est émotionnellement bouleversée par un mélange de stress et de tristesse. Ces changements vocaux dus à l'excitation émotionnelle de détresse peuvent être considérés comme les corrélats acoustiques de l'émotion de détresse, car ils sont observés ches toutes les cinq locutrices américaines dans notre corpus anglais.

L'augmentation du Fo par la détresse est conforme à notre attente, du fait qu'il s'agit d'une forme excitée de tristesse[94]. Pourtant, l'unanimité de l'augmentation du Fo dans la voix émotionnelle des cinq locutrices américaines[95] est remarquable, quand on prend en compte le fait qu'une grande variation individuelle est souvent notée dans les études de l'émotion, surtout en ce qui concerne la manifestation du stress dans la voix[96]. D'après l'expérience de Bonner (1943), 46 sur 65 sujets montrent une augmentation du Fo dans le cas du stress, tandis que les autres 19 sujets montrent une diminution du Fo dans le même contexte. L'étude de Hecker *et al.* (1968) montre que dans une situation stressante,

[93] Voir IV.4.1.5 pour le cas du corpus coréen.
[94] L'émotion du détress est aussi exprimée par l'augmentation du Fo dans notre corpus coréen (voir IV.4.1.1).
[95] Il faut souligner que le ch+oix des locutices Américainesétait aléatoire au moments de la construction du corpus anglais. C'est-à-dire que les locutrices ont été choisies selon le critère suivant : elles ont toutes parlé en larmes à un moment donné, bien qu'elles soient de personnalité différente.
[96] Le terme 'stress' ici désigne une excitation émotionnelle générale, non-spécifique, ce qui correspond à la notion générale du 'bouleversement émotionnel' dans la présente étude. Bonner (1943) a utilisé le terme 'tension émotionnelle' pour le même concept. Scherer (1979, p505) fait une revue des résultats des études sur le stress dans la voix et conclut que le Fo augmente dans le cas du stress mais la différence individuelle est tellement grande que le pourcentage du renversement de ce résultat (où le Fo diminue en cas de stress) atteint un niveau autour de 30%.

le Fo augmente pour deux locuteurs sur cinq, diminue pour les deux autres, et varie en fonction de la partie de l'énoncé pour le dernier locuteur[97].

En ce qui concerne le changement du débit par l'émotion de détresse, toutes les cinq locutrices ont ralenti leur débit de parole quand elles étaient en détresse, tandis que l'un des deux locuteurs dans l'étude de Streeter *et al*. (1983) a ralenti son débit de parole quand il était sous stress[98]. Cela semble être dû au fait que l'émotion de détresse vécue par nos locutrices américaines contient un aspect de tristesse (dont le débit est typiquement lent), tandis que l'émotion du stress vécue par les locuteurs dans l'expérience de Streeter *et al.* est une tension psychologique causée par une tâche de nature cognitive.

En résumé, l'analyse acoustique de notre corpus anglais montre une grande cohérence entre les cinq locutrices américaines en ce qui concerne les corrélats acoustiques de l'émotion de détresse, tandis que les autres analyses se heurtent souvent à une divergence considérable entre les individus dans la caractérisation des traits acoustiques du stress. L'une des raisons de la haute cohérence dans notre analyse acoustique peut être attribuée à la grande intensité émotionnelle vécue par nos locutrices américaines. Le stress vécu par les sujets dans les études citées au-dessus est produit avec une intensité émotionnelle d'un niveau relativement modéré par rapport à celui de notre corpus anglais. Une autre raison pour nos résultats unanimes peut être attribuée à la simplicité de nos mesures acoustiques puisqu'il ne s'agit que de deux paramètres (Fo moyen et débit moyen). Ces deux mesures peuvent être trop globales que pour discerner la différence subtile entre les locutrices. Malgré cette simplicité, les mesures du Fo moyen et du débit moyen semblent être utiles dans la caractérisation de la voix de détresse et dans la détection automatique du bouleversement émotionnel du sujet parlant. Le changement spectral des énoncés émotionnels, comme l'allongement et la glottalisation surtout vers la fin de l'énoncé, paraît être aussi un bon indice de l'émotion de détresse, ce que nous allons examiner au moyen d'une expérience perceptive dans la partie suivante.

[97] Dans la condition stressante, le Fo de la voix du dernier locuteur est rabaissé au début de l'énoncé mais il est élevé vers la fin de l'énoncé. Vu ce résultat, Hecker *et al.* (1968) suggèrent un contour différent de Fo pour la voix sous stress.
[98] Le débit de l'autre locuteur n'est pas changé dans la même situation stressante.

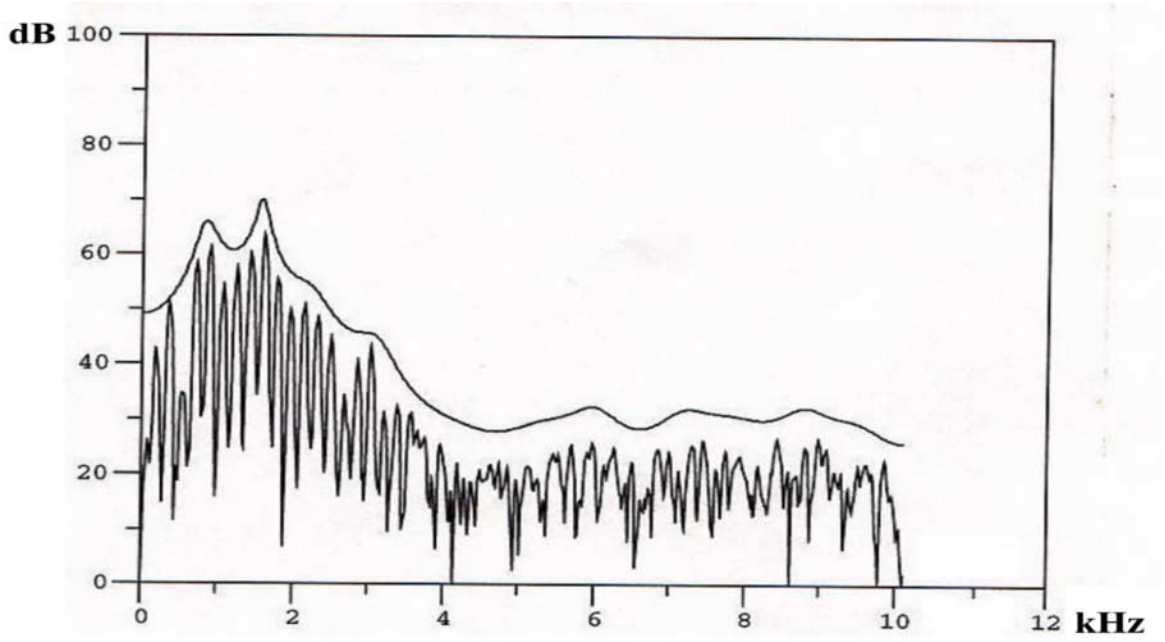


Figure 46. Spectre de la voyelle [a], extraite de l'énoncé émotionnellement neutre.

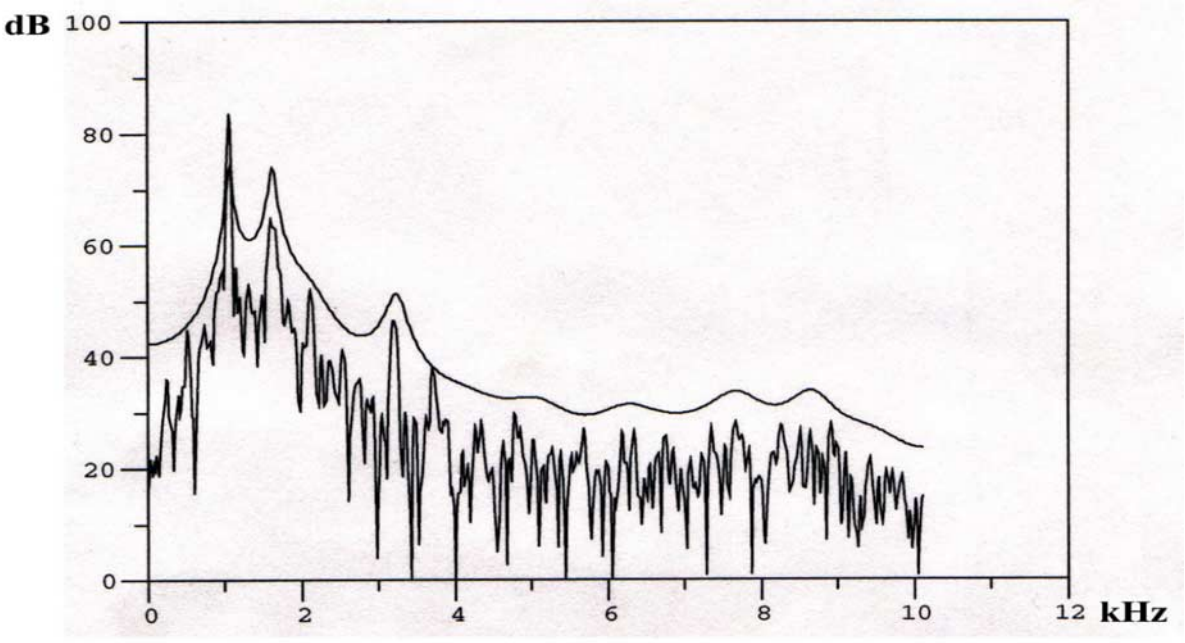


Figure 47. Spectre de la voyelle [a], extraite de l'énoncé émotionnellement négatif (détresse).

V.4. Analyse communicative

Pleurer de détresse est un trait fondamental et particulier à l'être humain (Darwin, 1872, p156)[99]. Quand on pleure, la respiration devient spasmodique et violente. L'acte de pleurer entraîne des modifications majeures au niveau de la glotte, ce qui s'entend au moment où l'inspiration l'emporte sur la résistance de la glotte, l'air affluant aux poumons[100]. Sur le plan de la communication, les pleurs en tant que l'expression émotionnelle de la détresse signalent un appel urgent à l'aide ou à la compréhension de l'interlocuteur. Vu l'expression explicite de la détresse dans la voix du locuteur (en pleurs) et l'urgence communicative de cette émotion (du point de vue psycho-évolutionnaire), la perception de l'émotion de détresse dans la parole est censée être simple et évidente. Or, dans quelle mesure la perception de cette émotion est-elle évidente pour l'auditeur, avec différentes tailles d'information ? Est-il vrai que la détresse du locuteur est toujours bien reconnue par l'auditeur, même dans la condition limitée où l'auditeur n'entend qu'une partie de l'énoncé au lieu de l'énoncé entier ? Est-ce que la reconnaissance de cette émotion serait toujours stable quelle que soit la position de la fenêtre d'extraction de cette partie dans l'énoncé ?

V.4.1. Expérience 6 : Le rôle des différentes parties de l'énoncé dans la communication de la détresse

L'Expérience 6[101] examine les problématiques, posées ci-dessus, par deux étapes d'analyse. La première étape consiste à voir si l'identification de l'émotion de détresse est comparable entre deux conditions de stimuli ; l'une où l'auditeur entend les stimuli d'énoncés entiers, et l'autre où l'auditeur entend les stimuli de parties, qui sont le résultat de la segmentation de l'énoncé en trois parties, initiale, médiane et finale, et sont

[99] Darwin (1872, p155) définit les pleurs comme l'expression primaire de la souffrance, causée par la douleur physique ou la détresse mentale. A propos de la particularité des pleurs humains, il cite une remarque des gardiens du jardin zoologique : qu'ils n'entendent jamais les pleurs ('*sobbing*') du singe, quoi que les singes crient bruyamment et halètent longtemps, quand ils sont capturés.

[100] Gratiolet (1865, p126) 'De la physionomie et des ,ouvements d'expression,' cité par Darwin (18720, p156).

[101] Cette expérience fait partie de la présentation de Chung (1998) dans l*es XXIIème Journées d'Etudes sur la Parole*, qui a eu lieu les 15-19 juin à Matigny en Suisse.

présentées de façon isolée. La deuxième étape d'investigation concerne la comparaison du taux d'identification entre les trois parties (initiale, médiane et finale) de l'énoncé.

Du point de vue de la théorie informationnelle (voir II.2), il s'agit de la répartition de l'information émotionnelle dans l'énoncé, ce qui peut être formulé de la manière suivante : est-ce que l'information de l'émotion de détresse est répartie au cours de l'énoncé de façon uniforme, ou bien est-elle concentrée dans une certaine partie de l'énoncé de façon particulière ? Si l'information de l'émotion était constamment présente au cours de l'énoncé, l'identification de l'émotion ne serait pas influencée quelle que soit la partie qu'on écoute. En même temps, si l'émotion était communiquée de façon purement qualitative[102], l'identification de l'émotion ne serait pas différente soit avec les stimuli d'énoncé soit avec les stimuli de partie. Or, si l'information émotionnelle était exprimée plus dans une certaine partie de l'énoncé que dans d'autres, l'identification de l'émotion serait sensible à la grande quantité d'information de cette partie, donc elle serait meilleure avec cette partie qu'avec les autres.

Notre hypothèse de l'Expérience 6 est que les pleurs de détresse seraient mieux identifiés avec la partie finale de l'énoncé qu'avec les autres parties, initiale et médiane. Cette hypothèse se base sur notre observation acoustique des énoncés émotionnels du corpus anglais et sur les autres remarques sur la particularité de la fin de l'énoncé dans la communication symbolique et symptomatique[103]. D'après notre analyse acoustique, le ralentissement et l'irrégularisation de la vibration des cordes vocales sont renforcés vers la fin de l'énoncé dans la voix pleurante (voir V.3). Ce phénomène a été noté par Hecker *et al.* (1968) et Williams & Stevens (1972) en tant que la manifestation du stress émotionnel dans la voix. Selon ces derniers, l'irrégularité de la vibration glottale est principalement due à la perte de contrôle du locuteur sur sa production articulatoire (l'aspect symptomatique), ce qui peut être éventuellement imité par l'acteur dans sa stylisation de la tristesse (l'aspect symbolique). Vu que la configuration des traits acoustiques de la fin de l'énoncé est importante dans la communication linguistique et paralinguistique[104], le

[102] Ici, il y a deux suppositions : que la qualité de voix en tant qu'information émotionelle est présente de façon constante dans l'énoncé et que l'identification de l'émotion se base principalement sur cette information.

[103] Dans le modèle de la comunication de Bühler (1934), le symble représente le signe motivé, conventionnel, tandis que le symptôme exprime l'état d'âme et les traits personnels du locuteur (voir II.2.1).

[104] La configuration du Fo, de la durée et de l'intensité de la partie finale de l'énoncé informe sur diverses modalités linguistiques et parlinguistiques, comme le statut syntaxique de l'énoncé (affirmatif ou interrogatif), l'attitude du sujet parlant (l'assurance, l'incertitude ou l'ironie), le cliché peronnel, etc. (voir les pages 147 et 148).

contrôle de la production vocale sur la partie finale de l'énoncé semble être essentiel dans la plupart des situations communicatives. S'il en est ainsi, la perte de contrôle sur la production vocale, dans le cas du bouleversement émotionnel, serait plus marquée vers la fin de l'énoncé, par rapport au début et au milieu de l'énoncé. C'est-à-dire, le contraste entre la régularité et l'irrégularité du fonctionnement de l'appareil vocal serait plus net dans la partie finale de l'énoncé que dans les autres parties, initiale et médiane. L'Expérience 6 teste cette hypothèse avec les données anglaises et les auditeurs américains.

V.4.1.1. La préparation des stimuli

A partir du corpus anglais, nous avons sélectionné 60 énoncés (30 neutres et 30 émotionnels) pour les stimuli d'énoncés ('Enoncés') dans l'Expérience 6. Ces derniers se composent de 12 énoncés de chacune des cinq locutrices, dont la moitié est neutre et la moitié est émotionnelle. La sémantique et la durée des énoncés étaient prises en compte dans notre sélection des stimuli[105]. Les Enoncés émotionnels ne sont pas sémantiquement plus émotionnels que les Enoncés neutres. La structure syntaxique n'est pas identique entre les énoncés. Etant donné que les Enoncés ont été prononcés dans une prise de souffle, il n'y a pas de silence perceptible à l'intérieur de l'énoncé. La durée moyenne (écart-type) des Enoncés est de 1180 (184,8)ms.

A la suite de la sélection des Enoncés, nous avons construit les stimuli de parties ('Parties'), en divisant chaque énoncé en trois parties, initiale, médiane et finale, de manière proportionnelle (voir la Figure 48). Dans la segmentation de l'énoncé, nous avons évité de découper au milieu du mot, afin de garder l'impression auditive naturelle des stimuli. Donc la frontière de la segmentation a été mise en accord avec la frontière du mot dans la mesure du possible. Etant donné qu'il était essentiel, dans l'Expérience 6, d'avoir les durées comparables entre les trois Parties (initiale, médiane et finale) d'un énoncé donné, il y avait des cas où il nous fallait diviser un mot (qui se trouve à la frontière des parties) en deux en vue du découpage proportionnel. Dans ce cas-là, la segmentation a été faite en fonction de la frontière de la syllabe, et les morceaux du mot ont été inclus ou exclus des Parties selon la naturalité de l'impression auditive des Parties concernées. Ainsi, 180 Parties (60 énoncés x 3 positions de l'énoncé) ont été obtenues par suite de la

[105] Ils sont les mêmes critères, appliqués dans notre Expérience 4.

segmentation des énoncés en trois parties, initiale, médiane et finale. La durée moyenne (l'écart-type) des Parties est de 346,8 (50,4) ms. La transcription phonétique et la description de la durée des énoncés sont présentées dans l'annexe. Le logiciel 'Mev'[106] a été utilisé pour la préparation des stimuli de cette expérience.

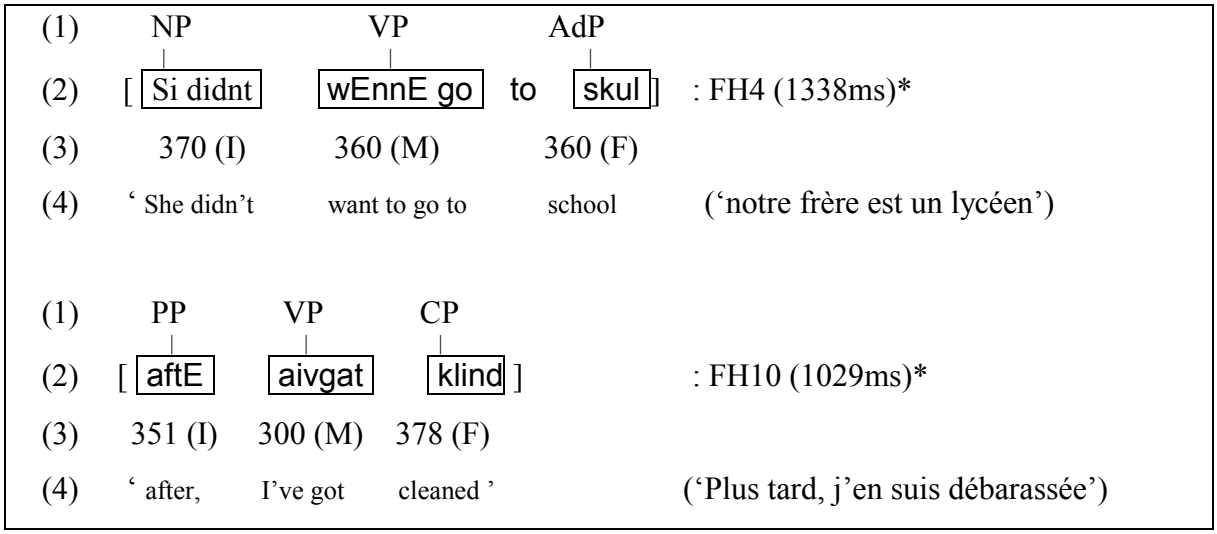


Figure 48. Exemples du découpage de l'énoncé en trois parties, Initiale, Médiane et Finale.
* Il n'y a pas de silence perceptible à l'intérieur de l'énoncé (voir le texte).
(1) Etiquetage syntaxique, (2) transcription phonétique (3) Durées des parties, (4) Transcription en alphabets anglais (Traduction française).

V.4.1.2. Test de perception

Vingt-deux Américains, étudiants de différents départements à l'université de Brown, ont participé au test de perception. La moitié des auditeurs était des hommes et l'autre moitié des femmes. Leur âge variait de vingt et un ans à trente ans. Ils se sont portés volontaires pour le test. Les auditeurs ont été divisés en deux groupes et chaque groupe a écouté un des deux types de stimuli (les Enoncés et les Parties) d'une locutrice donnée, ce qui a été conçu pour éviter tout effet d'apprentissage. Les auditeurs devaient juger si la locutrice pleurait ou pas, en choisissant l'une des deux cases étiquetées comme 'émotionnel (pleurs de détresse)' et 'neutre (non-émotionnel)'. Les instructions étaient indiquées au début du questionnaire de la façon suivante :

« Vous allez écouter les stimuli vocaux, qui sont des segments de la parole de cinq locutrices Américaines. La parole de chaque locutrice sera présentée dans une session et il

[106] Voir la note n°31 pour la description du programme 'Mev'.

y a cinq sessions dans le test qui suit. Les stimuli peuvent être des phrases ou des mots, donc leur durée varie largement. Les stimuli courts ne sont pas nécessairement compréhensibles en tant que tels, donc ne cherchez pas à trouver le sens du stimulus. La décision doit être basée sur votre impression subjective du stimulus sonore, mais pas sur votre compréhension lexicale du stimulus. Votre tâche est, après avoir écouté un stimulus, de décider si la locutrice est émotionnelle ou pas, c'est-à-dire si elle pleure de détresse ou pas. Dès que la décision sera prise, mettez une croix ('x') sur l'une des deux cases, marquées comme 'émotionnel (pleurs de détresse)' et 'neutre (non-émotionnel)'»

Les auditeurs ont passé le test individuellement dans une chambre sourde. Les stimuli étaient présentés aux auditeurs par l'intermédiaire d'un haut-parleur, et cela une fois, dans un ordre aléatoire. L'ordre de la session a été aussi contrebalancé entre les auditeurs. Les auditeurs ont répondu sur le questionnaire pendant deux secondes d'intervalle entre les stimuli d'énoncé et pendant une seconde d'intervalle entre les stimuli de partie. Le test total a duré environ 45 minutes.

V.4.1.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 1920 réponses des auditeurs, y compris 480 réponses pour les stimuli d'énoncés et 1440 réponses pour les stimuli de parties[107]. Une série d'analyses statistiques sont effectuées au moyen du logiciel 'SPSS,' en ce qui concerne le nombre de réponses correctes dans l'identification émotionnelle (la variable dépendante). La réponse correcte est déterminée par la correspondance entre la catégorie des stimuli dans le corpus, émotionnel ou neutre, et la réponse des auditeurs. Deux variables indépendantes sont examinées dans cette analyse, le facteur TYPE à deux niveaux (Enoncé et Partie) et le facteur POSITION à trois niveaux (Initiale, Médiane et Finale). L'influence du facteur sur les réponses des auditeurs est estimée par la différence du nombre de réponses correctes entre les différents niveaux du facteur. Le nombre de

[107] Parmi les 60 stimuli d'énoncé, quatre énoncés de chacune des trois locutrices, C, D et E, (12 énoncés au total) ont été omis au moment de la présentation des stimuli dans le test de percpetion pour une raison mécanique. Ensuite, les parties des 12 énoncés manquants (36 parties = 3 parties x 12 énoncés) ne furent pas présentés aux auditeurs, parce que notre expérience concerne la comparaison des réponses entre les Enoncés et les Parties (qui sont issues des mêmes énoncés). En ce qui concerne l'auditeur, deux auditeurs ont noté leurs réponses sur le questionnaire d'une façon décalée, donc leurs réponses ne sont pas prises dans les données à analyser. Finalement, l'analyse du résultat de cette expérience ne prend en compte que les réponses de 20 auditeurs (10 auditeurs de chaque groupe) pour les stimuli de 48 Enoncés et de 144 Parties.

réponses correctes pour un niveau donné de chaque facteur est appelé le taux d'identification.

L'effet du facteur TYPE sur les réponses des auditeurs est estimé par une analyse inter-sujets (le test-t indépendant), qui compare les taux d'identification des Enoncés et des Parties, évalués par deux différents groupes d'auditeurs. L'effet du facteur POSITION sur les réponses des auditeurs est estimé par une analyse intra-sujets (l'ANOVA à un facteur avec mesures répétées par sujets), qui compare les taux d'identification des positions initiale, médiane et finale, évalués par un même groupe d'auditeurs de façon répétitive[108]. La significativité de la différence du taux d'identification des Enoncés par rapport au taux d'identification pour chacune des trois positions de Parties, est estimée par une comparaison deux à deux, telle que 'Enoncé vs. Parties initiales,' 'Enoncé vs. Parties médianes' et 'Enoncé vs. Parties finales,' au moyen de trois tests-t indépendants.

V.4.1.4. Résultat

Le test-t indépendant montre un effet significatif du facteur TYPE, en indiquant que le taux d'identification émotionnelle avec les stimuli d'énoncés est plus élevé que celui avec les stimuli de parties ($t(125)=3,74$, $p<0,01$)[109]. Le passage du taux d'identification de 85,8% à 75,4% des Enoncé aux Parties montre que l'identification de l'émotion (pleurs de détresse) est diminuée d'environ 10%, quand l'auditeur n'entend qu'une partie de l'énoncé, au lieu de l'énoncé entier (voir la Figure 49). Ce résultat n'est pas surprenant mais il n'avait jamais été démontré auparavant avec des évidences expérimentales.

En ce qui concerne les stimuli de parties ('Parties'), l'ANOVA avec des mesures répétées montre un effet significatif du facteur POSITION sur les réponses d'auditeurs ($F(2,18)=26,81$, $p<0,01$). Cela indique que l'identification de l'émotion varie en fonction de la position dans l'énoncé duquel la Partie est extraite. Les taux d'identification nous informent que l'émotion de détresse (indiquée par les pleurs) est mieux identifiée avec les

[108] Deux ensembles de données sont utilisés pour les deux analyses, l'analyse inter-sujets et l'analyse intra-sujets. Ce sont les données dont le contenu est le même mais l'ordre des entrées est différent. L'analyse inter-sujets (comme le test-t) est effectuée sur les données dont les entrées (nombres de réponses corectes) sont ordonnées par stimulus, et l'analyse intra-sujets (comme l'ANOVA avec des mesures répétées) est effectuée sur les données dont les entrées sont ordonnées par auditeur en fonction des trois niveaux du facteur POSITION, inditiale, médiane et finale.

[109] Vu la grande différence du nombre de stimuli entre les Enoncés (48) et les Parties (144), l'égalité de la variance entre les deux groupes de stimuli n'est pas assumée, ce qui explique un degré de liberté de 125, au lieu d'être de 190.

Parties extraites à partir de la position finale dans les énoncés (‘Parties finales’) que celles extraites à partir des positions initiale et médiane (‘Parties initiales’ et ‘Parties médianes’). La meilleure identification de l’émotion avec les Parties finales est confirmée par les comparaisons deux à deux entre les trois taux d’identification : le taux d’identification des Parties finales est significativement plus haut que ceux des Parties initiales et des Parties médianes, tandis que les deux derniers ne sont pas différents en termes statistiques (voir la Figure 50).

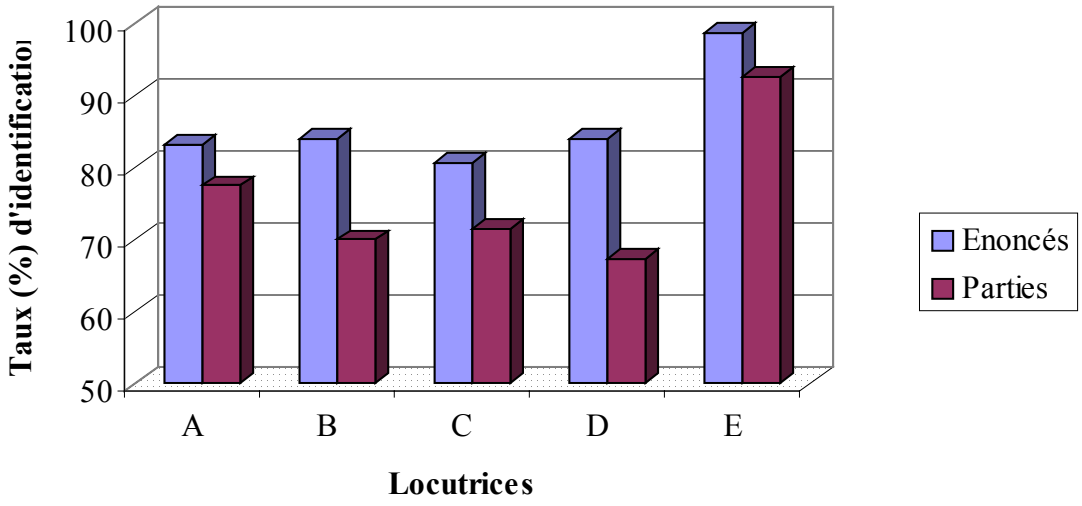


Figure 49. Taux (%) de l’identification émotionnelle des Américains pour les stimuli d’énoncés (‘Enoncés’) et les stimuli de parties (‘Parties’), qui sont des extraits de la parole de cinq locutrices Américaines, A, B, C, D et E.

Ce résultat est trouvé en commun en ce qui concerne les énoncés des cinq locutrices Américaines. L’effet du facteur POSITION sur l’identification de l’émotion concerne spécifiquement les stimuli émotionnels[110], pourtant la meilleure identification émotionnelle avec les Parties finales de l’énoncé est aussi trouvée en ce qui concerne les stimuli neutres. Enfin, les taux d’identification des Parties initiales, médianes et finales sont comparés deux à deux avec le taux d’identification des Enoncés, à l’aide de trois tests-t indépendants. Les résultats des tests montrent que le taux d’identification est significativement plus élevé pour les Enoncés que les Parties initiales ($t(18)=2,56$, $p<0,05$) et les Parties médianes ($t(18)=3,52$, $p<0,01$), tandis qu’il n’est pas très différent entre les Enoncés et les Parties finales ($t(18)=1,41$, $p>0,05$).

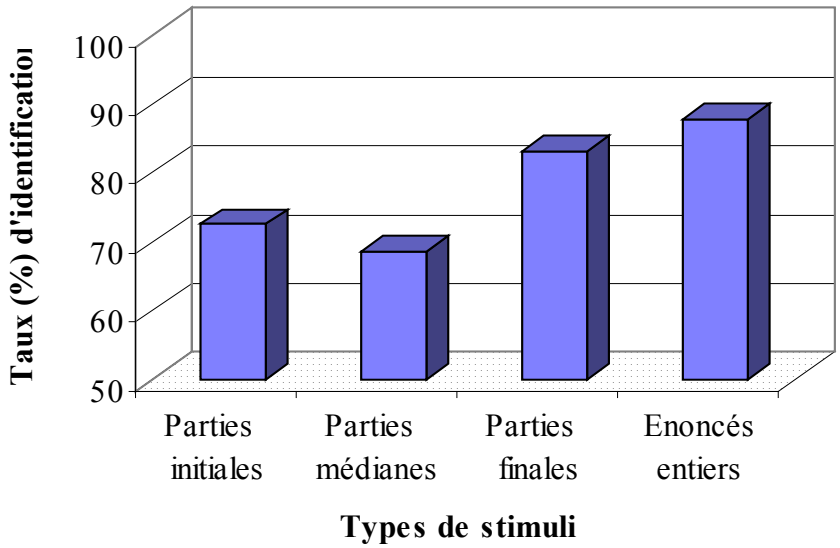


Figure 50. Taux (%) de l'identification émotionnelle des Américains avec les différents types de stimuli, les Enoncés entiers et les Parties extraites à partir des positions, initiale, médiane et finale, dans l'énoncé.

V.4.1.5. Discussion

D'après l'Expérience 6, l'émotion (les pleurs de détresse) est généralement mieux identifiée quand l'auditeur entend l'énoncé entier que quand il n'entend qu'une partie de l'énoncé. Or, en ce qui concerne le dernier cas où les parties de l'énoncé sont présentées en forme isolée à l'auditeur, la précision de l'identification émotionnelle de l'auditeur dépend de la position dans l'énoncé duquel la partie est extraite. L'émotion est mieux identifiée par l'auditeur avec les parties de l'énoncé qui sont extraites à partir de la position finale dans l'énoncé qu'avec celles qui sont extraites à partir des autres positions, initiale et médiane, dans l'énoncé. Le taux d'identification émotionnelle de l'auditeur avec la partie finale de l'énoncé est si élevé qu'il est presque comparable à celui qui en est fait avec l'énoncé entier. Ces résultats signifient, du point de vue de l'expression émotionnelle, que la manifestation vocale du bouleversement émotionnel du locuteur (qui finir par pleurer de détresse) varie au cours de l'énoncé, les indices émotionnels étant exprimés d'une façon plus explicite vers la fin de l'énoncé qu'au début et au milieu de l'énoncé.

[110] C'est parce qu'il s'agit de la question : comment l'émotion est-elle identifiée en fonction des différentes parties de l'énoncé ?

Du point de vue de l'expression émotionnelle, ces résultats signifient que la manifestation vocale du bouleversement émotionnel du locuteur, qui finit par pleurer de détresse, varie au cours de l'énoncé, les indices émotionnels étant exprimés d'une façon plus explicite vers la fin de l'énoncé qu'au début et au milieu de l'énoncé. Le renforcement de l'irrégularité de la vibration des cordes vocales vers la fin de l'énoncé paraît largement responsable de cette variation expressive des différentes parties dans l'énoncé émotionnel. Un certain degré de l'irrégularité de la vibration glottale est trouvé dans la voix émotionnellement neutre, pourtant il n'est guère perceptible au niveau conscient de l'auditeur (Lieberman, 1961 ; Horri, 1982)[111]. Or, dans le cas du stress émotionnel, la vibration glottale devient plus irrégulière et plus lente, surtout vers la fin de l'énoncé, ce que Hecker *et al.* (1968) et Williams & Stevens (1972) attribuent à la perte du contrôle précis sur la production vocale et à la respiration irrégulière[112]. Cette tendance est retrouvée dans les énoncés émotionnels de notre corpus anglais, qui ont été produits au moment où la locutrice pleurait de détresse. Notre analyse acoustique des énoncés émotionnels montre que la vibration glottale devient extrêmement irrégulière et interrompue de façon sporadique, ce qui est le plus visible dans la partie finale de l'énoncé dans le spectrogramme (voir la Figure 45)[113]. Vu la relation étroite entre l'état physiologique et l'état émotionnel[114] (Tomkins, 1984[115] ; Scherer, 1986[116]), le renforcement de l'irrégularité de la vibration des cordes vocales vers la fin de l'énoncé semble refléter le renforcement de l'excitation émotionnelle du locuteur au cours de l'énoncé, auquel le contrôle phonatoire du locuteur est lié en fonction inverse. Ce genre de contrôle n'est pas nécessairement conscient, il repose plutôt sur le niveau inconscient en majorité puisqu'il s'agit du fonctionnement automatique des appareil vocaux. Malgré l'automaticité physiologique, le fonctionnement des appareils phonatoires peut être perturbé par le bouleversement psychologique émotionnel, ce qui résulte en l'irrégularité de la vibration des cordes vocales en cas de détresse. Cet indice vocal de la détresse est bien reconnu par l'auditeur dans

[111] La syllabe finale de l'énoncé est souvent allongée et glottalisée en tant que démarquage de la finalité, mais ce genre d'allongement et de glottalisation d'un degré modéré n'est pas lié à la perception émotionnelle.
[112] Hecker *et al.* ont aussi noté la variation individuelle dans ces changements vocaux par le stress.
[113] La variation individuelle est négligeable dans notre corpus anglais. Cela peut être dû au fait que le degré d'intensité émotionnelle de la détresse (vécue par nos locutrices Américaines) est si grand qu'il recouvre la variabilité individuelle dans la manifestation vocale de cette émotion.
[114] La plupart des théories de l'émotion se basent sur l'interaction entre éléments physiologiques et éléments psychologiques dans l'expérience émotionnelle, et nombre d'études acoustiques ont démontré cette relation en termes de corrélats acoustiques de l'émotion (voir II.5.1).
[115] Voir II.3.2 pour le revue de la théorie de l'*affect* de Tomkins (1962, 1984).
[116] Dans le modèle du processus componentiel (Scherer, 1986), l'expression vocale de l'émotion est considérée comme le résulat des changements physiologiques, causés par l'évaluation psychologique du sujet parlant sur des événements significatifs dans son environnement. Voir II.4.1 pour le revue du modèle du processus componentiel.

l'identification de l'émotion selon notre Expérience 6. La découverte de la *primauté de la partie finale de l'énoncé dans l'expression et la perception de l'émotion de détresse* (les pleurs de détresse) apporte de nouvelles perspectives dans l'étude de l'émotion, en ce qui concerne la détection automatique de l'émotion et la stylisation émotionnelle de la voix synthétique. Ces dernières seront discutées dans le chapitre VII à l'égard de l'application potentielle des résultats de notre étude.

Dans cette expérience, la neutralité de la partie finale des énoncés neutres est bien identifiée en tant que telle, tandis que la partie finale des énoncés neutres est identifiée comme émotionnellement positive dans notre Expérience 5 faite avec le corpus coréen (voir IV.6.1). Cette divergence des résultats des deux expériences peut être expliquée par le fait suivant. D'une part, les auditeurs n'avaient que deux choix de réponse (émotionnel ou neutre) dans le test de perception de cette expérience, tandis qu'ils en avaient trois (positif, neutre et négatif) dans l'Expérience 5. D'une autre part, la partie finale de l'énoncé neutre dans notre corpus anglais n'a pas de traits prosodiques particuliers, tandis que la partie finale de l'énoncé neutre dans notre corpus coréen se compose d'une longue syllabe finale avec une intonation montante suivie par une intonation descendante (contour de Fo *montant-descendant*), ce qui semble exprimer l'émotion positive de la locutrice WJ vis-à-vis de l'interlocuteur. Cette supposition est testée dans le chapitre suivant, VI, par une expérience avec des stimuli synthétiques, en ce qui concerne le rôle de l'intonation et de la durée dans la communication de l'émotion.

V.5. Conclusion du chapitre V.

Dans le chapitre V, nous avons étudié comment l'émotion de détresse est exprimée dans la parole spontanée en anglais et comment cette émotion est perçue par l'auditeur quand il n'entend qu'une partie de l'énoncé émotionnel au lieu de l'énoncé entier. Nous avons d'abord décrit l'acquisition du corpus anglais et des traits prosodiques de l'expression émotionnelle de la détresse avec des mesures acoustiques, puis avons effectué une expérience perceptive afin de répondre à la deuxième question. Ici, nous allons réviser ce qui a été discuté dans ce chapitre, en précisant le statut de cette étude dans la structure générale de la présente thèse.

Le corpus anglais consiste en des discours de cinq locutrices américaines, enregistrés à partir d'une série d'entretiens télévisés, dans lesquels il y a des moments où la locutrice était tellement émotionnelle qu'elle a fini par pleurer de détresse, en racontant son problème personnel. Vu que l'entretien s'est effectué de manière improvisée et que les locutrices ont pleuré au moment où elles ne pouvaient plus se retenir, l'expression émotionnelle de la détresse de ces locutrices est considérée comme le reflet authentique de leur état émotionnel. Dans l'analyse acoustique, le Fo moyen et le débit de parole des énoncés émotionnels produits au moment des pleurs sont comparés à ceux des énoncés neutres produits au moment de l'absence d'émotion particulière. Le résultat montre que le Fo augmente et que le débit ralentit quand le sujet parle en sanglotant. Les mesures du Fo moyen et du débit sont trouvées efficaces à caractériser l'émotion de détresse, étant un mélange de stress et de tristesse. L'augmentation du Fo par l'excitation émotionnelle (ce qui correspond au bouleversement émotionnel dans notre corpus anglais) reflète l'aspect de stress de l'émotion de détresse, alors que le ralentissement du débit reflète l'aspect de tristesse de cette émotion.

En nous basant sur l'observation du spectrogramme des énoncés émotionnels, à savoir que la vibration glottale devient extrêmement irrégulière vers la fin de l'énoncé, nous avons effectué un test de perception dans lequel les pleurs de détresse étaient identifiés par deux groupes d'auditeurs, l'un écoutant des énoncés en forme entière, et l'autre écoutant des parties des énoncés, en forme isolée, qui ont été extraites

respectivement à partir de la position, initiale, médiane et finale, de l'énoncé. La comparaison des taux d'identification entre les deux groupes montre un abaissement significatif du taux d'identification des stimuli de parties par rapport aux stimuli d'énoncés, sauf dans le cas où les parties viennent de la position finale de l'énoncé. Autrement dit, la partie finale de l'énoncé est hautement informative dans l'identification de l'émotion de détresse, ce qui est comparable à l'énoncé entier, tandis qu'il n'en est pas ainsi pour la partie initiale et la partie médiane de l'énoncé. L'irrégularité de la vibration glottale, en tant qu'indice de l'émotion de détresse, est expliquée par la perte de contrôle du sujet pleurant sur sa production vocale. Le renforcement de l'irrégularité de la vibration glottale vers la fin de l'énoncé, lié à la meilleure identification de l'émotion de la détresse dans la partie finale de l'énoncé, semble être dû à l'augmentation de l'excitation émotionnelle au cours de l'énoncé, laquelle perturbe le fonctionnement des appareils phonatoires et respiratoires. La partie finale de l'énoncé, étant la fin de l'unité de souffe et la fin de l'utité de sens, paraît être un endroit plus sensible à la modification physiologique du locuteur et un endroit auquel l'auditeur fait plus d'attention dans la communication parlée.

La primauté de la partie finale de l'énoncé dans la communication émotionnelle de détresse est un résultat nouveau et intéressant, bien qu'il doit à être confirmé par d'autres analyses avec plus d'évidences expérimentales. L'étude avec le corpus coréen a servi de point de départ pour notre itinéraire à la recherche de la communication de l'émotion dans les différentes langues. L'étude avec le corpus anglais est largement inspirée de cette étude en ce qui concerne la méthode et la problématique. Le résultat du corpus anglais fournit une autre évidence de la particularité de la partie finale de l'énoncé dans la communication émotionnelle, et nous attendons d'autres analyses ultérieures qui prouvent la validité de notre résultat avec les données en d'autres langues et des sujets plus nombreux.

Chapitre VI

Etude vérificative avec des stimuli synthétiques

Résumé

Ce chapitre présente deux expériences avec des stimuli synthétiques, qui ont été effectuées en vue de la vérification par synthèse de nos résultats des chapitres précédents. Dans la première expérience, nous avons examiné l'influence de la modification du Fo contour et de la durée sur la perception de l'émotion. Quatre contours de Fo (*montant*, *descendant*, *montant-descendant* et *plat*) et deux durées (*longue* et *courte*) ont été synthétisés à partir de trois syllabes extraites de la parole naturelle. L'impression émotionnelle de ces 24 stimuli synthétiques (4 Fo contours x 2 durées x 3 syllabes originales) a été évaluée par dix Coréens et dix Américains en termes d'émotion POSITIVE et d'émotion NEGATIVE. Le résultat montre que les contours intonatifs contenant un élément 'montant' (comme le contour *montant* et le contour *montant-descendant*) produisent un biais perceptif vers une émotion POSITIVE, tandis que les contours 'non-montants' (comme le contour *descendant* et le contour *plat*) produisent un biais perceptif vers une émotion NEGATIVE. Les résultats des Coréens et des Américains sont similaires, sauf que la contribution du contour plat à la perception de l'émotion NEGATIVE est peu significative dans le cas des Américains. Dans la deuxième expérience, nous avons examiné si l'insertion du trait prosodique dans les différentes parties (initiale, médiane et finale) de l'énoncé influence la perception de l'émotion. Quatre contours de Fo *montant-descendant*, identifiés comme l'indice de l'émotion positive, ont été choisis comme les traits de cible. Chaque trait de cible a été inséré dans la partie initiale, médiane ou finale de l'énoncé au moyen de la resynthèse. L'impression émotionnelle (POSITIVE ou NEGATIVE) de ces 12 stimuli (4 traits de cible x 3 positions) a été évaluée par dix Coréens et dix Américains. D'après le résultat, l'émotion de l'énoncé était plus fréquemment perçue comme POSITIVE quand le contour montant-descendant était placé dans la partie finale de l'énoncé que quand il était placé dans la partie initiale ou médiane de l'énoncé. Il est à préciser que l'expérience du chapitre VI concerne la modification volontaire des traits prosodiques qui vise à communiquer le sens émotionnel, alors que l'expérience du chapitre V concerne la modification involontaire des traits prosodiques qui résulte du bouleversement émotionnel du locuteur malgré l'intention du locuteur. Ces deux types de modification des traits prosodiques en tant qu'expression émotionnelle ont été expliqués respectivement sur base du modèle de configuration et celle du modèle de covariation.

VI. Etude vérificative avec des stimuli synthétiques

VI.1. Préliminaires

Le chapitre VI présente deux expériences, effectuées avec des stimuli synthétiques, sur la contribution respective du contour intonatif et de la durée à la perception émotionnelle et l'influence de la position de la cible prosodique dans l'énoncé sur la perception émotionnelle. Les problématiques de ces expériences viennent de l'observation des résultats de nos expériences précédentes, effectuées avec des stimuli de la parole naturelle en coréen. Dans l'Expérience 4, nous avons constaté que l'auditeur a une tendance à identifier une émotion positive dans un énoncé, qui porte un contour intonatif de la forme de cloche ('*montant-descendant*') sur sa longue syllabe finale, malgré que cet énoncé ait été produit au moment où la locutrice parlait sans excitation émotionnelle. L'Expérience 5 nous a précisé la situation, avec pour résultat que la partie finale de cet énoncé neutre, laquelle contient le contour intonatif *montant-descendant* et l'allongement vocalique, est souvent perçue comme émotionnellement positive, tandis que la partie initiale et la partie médiane de cet énoncé neutre, lesquelles n'ont pas de tels traits prosodiques, sont perçues en tant que telles, c'est-à-dire comme neutres. Cette tendance a été observée chez les Coréens aussi bien que chez les auditeurs étrangers, Français et Américains. Vu cette variation perceptuelle émotionnelle en fonction de la variation des traits prosodiques, vous allons vérifier, dans ce chapitre, dans quelle mesure ces traits prosodiques influencent la perception émotionnelle dans un contexte indépendant.

Le chapitre VI examine principalement trois facteurs dans la perception de l'émotion ; la variation intonative, la variation de la durée et la position de la partie dans l'énoncé. Les deux premiers facteurs sont examinés dans l'Expérience 7 et le dernier dans l'Expérience 8. Les stimuli de ces expériences sont préparés au moyen de la resynthèse par manipulation systématique, et les résultats sont comparés entre les auditeurs coréens et les auditeurs américains, prenant en compte le facteur culturel.

VI.2. Analyse par synthèse

VI.2.1. Expérience 7 : Les contributions respectives du contour de Fo et de la durée à la perception de l'émotion

L'Expérience 7 examine comment le contour de Fo ('contour de Fo') et la durée influencent la perception de l'émotion, en utilisant des stimuli synthétiques. Cette expérience a pour fonction de vérifier des phénomènes observés dans nos expériences précédentes, en utilisant des stimuli synthétiques. Dans l'Expérience 5, nous avons constaté que l'auditeur a une tendance à percevoir une émotion positive dans une partie de l'énoncé, laquelle contient un contour intonatif en forme de cloche ('*montant-descendant*') sur sa longue syllabe finale, bien que l'énoncé d'origine ait été produit au moment où la locutrice parlait sans excitation émotionnelle. Etant donné que ces traits prosodiques sont souvent utilisés dans l'expression de la politesse par les jeunes Coréennes, nous avons supposé un lien entre l'intonation *montante-descendante* (accompagnée d'un allongement vocalique) et l'impression émotionnelle positive. Dans l'Expérience 7, nous allons vérifier l'existence de ce lien dans un contexte indépendant, en gardant le facteur du timbre vocal constant. La question est : est-ce que la perception de l'émotion positive et de l'émotion négative varie en fonction du contour de Fo ('*montant*,' '*descendant*,' '*montant-descendant*' et '*plat*') et de la durée ('*long*' et '*court*') du stimulus auditif ? Cette expérience est effectuée avec des auditeurs coréens et américains, afin de voir si la perception de l'émotion avec les différents traits prosodiques est comparable entre les deux groupes d'auditeurs qui sont culturellement différents.

La contribution des traits prosodiques (fréquentiels et rythmiques) à la communication émotionnelle n'est pas une nouvelle idée dans l'étude de l'émotion vocale (Skinner, 1935 ; Scherer & Oshinsky, 1977[117] ; Ladd *et al*., 1985 ; Leinonen *et al*., 1997). En même temps, il est aussi bien connu qu'il n'existe pas de relation univoque entre tel trait prosodique et telle signification émotionnelle (Fónagy, 1971b, p50 ; Pakosz, 1982, p158, Mozziconacci, 1998, p5). Ces deux phénomènes apparemment contradictoires (la

[117] Scherer & Oshinsky (1977) appellent ce phénomène 'l'utilisation des indices acoustiques dans l'attribution de l'émotion,' du point de vue de l'auditeur (voir la page 132).

communication de l'émotion par des traits prosodiques vs. l'absence de correspondance univoque entre le trait prosodique et la signification émotionnelle) conduisent les chercheurs à examiner la signification des traits prosodiques en termes dimensionnels, plutôt qu'en termes catégoriques spécifiques.

Udall (1960), dans l'article intitulé '*Signification d'attitude exprimée par le contour intonatif*,' propose que l'évaluation de l'attitude (émotionnelle) varie en fonction du contour de Fo dans les dimensions d'activation, de valence et de force[118]. Parmi ces dimensions, celle de valence ('positive-négative') est la plus pertinente et le jugement des auditeurs sur cette dimension est plus cohérent que celui sur les dimensions d'activité et de force. Dans un même type d'analyse plus élaborée (1964), elle précise que le contour *montant* à la fin de l'énoncé est associé à la perception de l'attitude agréable (positive), tandis que le contour *plat* est associé à la perception de l'attitude désagréable (négative)[119]. Cette remarque est comparable à ce que Kaiser (1962) a observé dans les données d'expressions émotionnelles avec des voyelles singulières en hollandais. Selon lui (*ibid*., p306), l'émotion positive (comme la gentillesse et la gaieté) est souvent exprimée par un contour *montant-descendant*[120], tandis que l'émotion négative (comme la tristesse et le mécontentement) est exprimée par un contour *descendant*[121]. Udall et Kaiser n'ont pas oublié de mentionner la possibilité de la variation due à la culture dans leurs résultats, malgré qu'aucun n'ait vérifié cette possibilité avec l'étude expérimentale.

En ce qui concerne la signification émotionnelle du contour intonatif dans la dimension d'activation (liée à l'intensité émotionnelle), Pakosz (1982) formule cinq principes pour décoder cette signification, en se basant sur des données anglaises : 1) le contour variable signifie une plus grande intensité émotionnelle que le contour statique ; 2) la grande plage de variation signifie la grande intensité émotionnelle[122] ; 3) la combinaison du contour *montant* et du contour *descendant* en direction endocentrique (en forme

[118] Dans l'expérience d'Udall, 16 contours de Fo synthétiques, construits à partir de phrases anglaises, ont été évalués sur dix axes sémantiques (comme 'poli-impoli,' 'intéressé-indifférent,' 'tendu-détendu,' etc.), et les réponses des auditeurs sont analysées en fonction de trois critères, 'agréable-désagréable,' 'intéressé-indifférent,' 'autoritaire-soumis,' qui correspondent respectivemnt aux dimensions de valence, d'activation et de force. Ces trois dimensions ont été introduites dans notre discussion dans la page 96.

[119] Elle a aussi noté l'association entre le contour de Fo descendant avec une grande plage et l'attitude autoritaire et l'association entre le contour de Fo montant final et l'attitude soumise, ce qui rappelle la notion du code fréquentiel ('*Frequecy code*') proposée par Ohala (1983).

[120] Il a aussi constaté que certains sujets féminins expriment la gentillesse avec une intonation *montante*.

[121] L'intonation descendante pour l'expression de la tristesse a été aussi notée par Fairbank & Pronovost (1939).

[122] Udall (1964, p257) a aussi noté la relation entre la plage de Fo et le degré d'intensité émotionnelle perçu.

concave) signifie une plus grande intensité émotionnelle que celle en direction excentrique (en forme convexe) ; 4) la variation complexe du contour signifie une plus grande intensité émotionnelle que la variation simple ; 5) le contour à un haut niveau de Fo signifie une plus grande intensité émotionnelle que celui à un bas niveau de Fo. La communication de l'émotion dans la dimension d'activation peut être aussi accomplie par la variation du débit de parole[123] : un débit rapide signale une activation de l'émotion du type joie, colère, peur, surprise tandis qu'un débit lent signale une activation de l'émotion du type tristesse et dégoût (Scherer, 1974).

Quand à l'exemple de l'analyse par synthèse, Carlson *et al*. (1992) ont échangé les traits prosodiques (comme le contour intonatif et la durée) entre phrases neutres et phrases émotionnelles, au moyen de la synthèse, et ont démontré la pertinence de ces traits dans la communication de l'émotion. Protopapas & Lieberman (1995) ont utilisé des stimuli synthétiques (construits par la modification de la valeur de Fo avec un même segment de la parole naturelle) dans un test de perception, et ont constaté la corrélation entre la valeur de Fo et le degré d'émotion perçue. Mozziconacci & Hermes (1998) ont effectué une série d'expériences perceptives avec des stimuli de différents contours intonatifs, synthétisés à la base des valeurs extraites de la parole naturelle en néerlandais, et ont démontré que différents contours intonatifs peuvent créer différentes impressions émotionnelles dans la voix synthétique.

VI.2.1.1. Préparation des stimuli

Les stimuli de l'Expérience 7 sont préparés par resynthèse des syllabes extraites à partir de la parole naturelle. La resynthèse consiste à modifier le contour de Fo et la durée des syllabes d'une façon paramétrique explicitée par la suite. Pour cette expérience, trois syllabes, [de], [ne] et [ge], de la locutrice WJ[124] ont été choisies en tant que segments à modifier. Ce sont les dernières syllabes de différents énoncés, qui font partie de la catégorie d'émotion neutre d'après l'Expérience 1 (voir IV.3.1.4). Chaque syllabe consiste en un contour de Fo *montant-descendant*, d'une intensité modérée et d'une durée relativement longue, de plus ou moins 500ms. Aucune syllabe ne contient de glottalisation ou

[123] D'après l'expérience de Scherer & Oshinsky (1977), le débit est l'indice le plus puissant, parmi les variations fréquentielles, temporelles et d'amplitude, dans la communication de l'émotion sur la dimension d'activation.
[124] Voir IV.2.3 pour la description de la locutrice WJ.

d'aspiration, ce qui nous permet de considérer la qualité de voix des syllabes comme neutre. Les valeurs de Fo et de la durée des trois syllabes originales sont présentées dans le Tableau 15, ces valeurs servent de valeurs de référence dans la fabrication des contours de Fo et des durées synthétiques.

Syllabe	Durée (ms)	Fo moyen	Fo min/max	Variation de Fo (Hz)	Jitter (%)
[de]	567	206	186 / 226	186 → 226 → 180	0
[ge]	505	197	186 / 212	186 → 212 → 185	0
[ne]	455	204	193 / 213	192 → 223 → 180	-0,1

Tableau 15. Les valeurs acoustiques des syllabes originales de l'Expérience 7 ; durée (ms), Fo moyen (Hz), Fo min/max (Hz), variation de Fo (Hz), et jitter (%).

Les stimuli synthétiques sont construits par la modification du contour de Fo et de la durée des syllabes originales, à l'aide du logiciel 'Winpitch'[125]. Cette méthode de resynthèse a l'avantage, par rapport à la synthèse complète, de garder le spectre de la syllabe originale à travers les modifications des paramètres acoustiques. Pour une syllabe originale donnée, huit syllabes synthétiques sont créées à travers des modifications de quatre contours de Fo (*montant*, *descendant*, *montant-descendant* et *plat*) et de deux durées (*long* et *court*). La resynthèse est effectuée sur la partie vocalique de la syllabe d'une façon relative. C'est-à-dire que chaque syllabe est modifiée en fonction de ses propres valeurs de Fo et de la durée selon la procédure suivante. Etant donné que le logiciel 'Winpitch' effectue la synthèse selon les valeurs du contour de Fo qu'on dessine manuellement, les contours *montant*, *descendant*, *montant-descendant* et *plat* sont synthétisés par l'interpolation manuelle des valeurs de Fo maximum, minimum et moyenne, dans différentes directions. Le contour *montant* est créé par une interpolation des valeurs à partir du Fo minimum jusqu'au Fo maximum. Le contour *descendant* est fait par une interpolation dans un sens inverse. Le contour *montant-descendant* est reconstruit par l'interpolation Fo début, Fo maximum et Fo fin. Le contour *plat* est créé avec une valeur constante de Fo moyen à travers la syllabe.

[125] Voir note n°52 pour la description du logiciel 'Winpitch'.

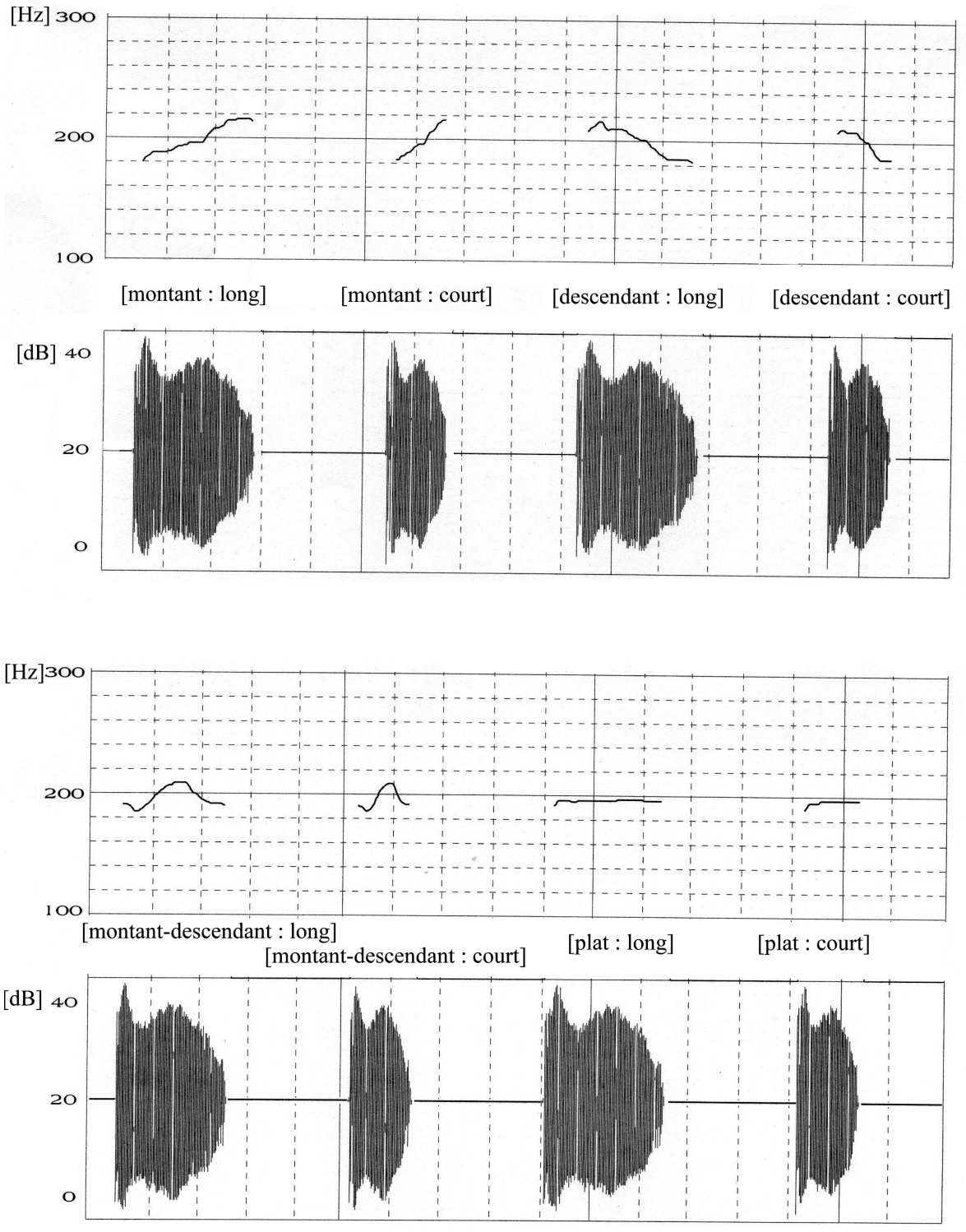


Figure 51. Exemple de huit variations des stimuli synthétiques, produites par la combinaison de quatre contours de Fo (*montant*, *descendant*, *montant-descendant* et *plat*) et de deux durées (*long* et *court*).

En même temps, nous avons modifié la durée des stimuli synthétiques de deux façons ; la durée *longue* et la durée *courte*. La première correspond à la durée originale de chaque syllabe et la dernière correspond à une moitié de la durée, qui résulte d'un rétrécissement de 50% de la durée *longue*. Ainsi, les durées *courtes* de 283,5ms, 222,5ms et 252,5ms sont créées à partir des durées *longues* de 567ms, 455ms et 505ms, pour les syllabes respectives [de], [ne] et [ge]. Les exemples des huit conditions des stimuli synthétiques sont illustrés dans la Figure 51.

Après manipulation, 24 syllabes sont resynthétisées par la modification des trois syllabes dans les huit conditions (3 syllabes x 4 contours de Fo x 2 durées). Ces syllabes sont insérées dans une phrase, appelée '**Phrase-type**,' afin de fournir un contexte de phrase à l'auditeur au moment de la présentation des stimuli. Cette phrase est l'un des énoncés de la catégorie d'émotion neutre (voir Tableau 7), elle a une intonation relativement monotone avec une qualité de voix neutre (ni cassage de voix ni bruit d'aspiration). Le Fo moyen (l'écart-type) de la phrase-type est de 192Hz (29Hz), et les Fo maximum et minimum sont de 162 et 227Hz[126]. La durée de la phrase-type est de 1780ms, excepté la syllabe finale. La syllabe finale de cette phrase[127] est allongée avec une intonation *montante-descendante*, et c'est à la place de cette syllabe que la syllabe synthétique ('stimulus de cible') est insérée dans la phrase-type[128]. L'insertion des syllabes synthétiques dans la position finale de la phrase s'explique par le fait que les syllabes originales ont été extraites à partir de la fin d'énoncés. Etant donné que les syllabes [de], [ne] et [ge] et la syllabe finale de la phrase-type sont échangeables en tant que particule finale de la phrase coréenne, l'insertion des syllabes synthétiques dans la phrase-type résulte en des phrases sémantiquement naturelles. Au total, les 24 stimuli de phrase sont préparés en tant que stimuli de l'Expérience 7 [129].

[126] Le contour mélodique de la phrase-type est illustré dans la Figure 54.

[127] La syllabe finale de la phrase-type est [ku], elle n'est pas utilisée en tant que stimuli de cette expérience (Expérience 7) mais elle est incluse dans les stimuli de l'expérience suivante (Expérience 8).

[128] Autrement dit, les syllabes de stimuli sont concaténées à la fin de la phrase-type.

[129] En fait, pour l'Expérience 7, nous avons utilisé trois types de stimuli dans trois tests de perception indépendants ; (1)'Syllabes isolées,' (2)'Syllabes insérées dans une phrase de la parole réitirée' et (3) 'Syllabes insérées dans une phrase de la parole naturelle'. C'est pour savoir quel mode de présentation des stimuli est le plus approprié à ce genre d'expérience. Les résultats des tests avec les trois types de stimuli sont comparables. Pourtant, le troisième type de stimuli ('Syllabes insérées en fin de phrase naturelle') paraît le plus approprié à ce genre d'expérience, parce qu'il donne l'impression la plus naturelle en tant que stimuli dans un contexte de la phrase.

VI.2.1.2. Test de perception

Dix Coréens et dix Américains ont participé au test de perception. Ce sont des étudiants de différentes universités américaines. Ces deux groupes d'auditeurs représentent respectivement les natifs de la langue coréenne et les natifs de la langue anglaise. Leur âge varie de vingt-deux ans à trente-trois ans. La moitié des auditeurs était des hommes et l'autre moitié des femmes. Les auditeurs ont passé le test de perception individuellement dans une pièce insonorisée. Le test s'est effectué avec le logiciel 'AVSUPER'[130]. La tâche de l'auditeur était de choisir quelle sorte d'émotion est exprimée dans le stimulus, en appuyant sur l'un des deux boutons étiquetés comme 'émotion positive' et 'émotion négative'[131]. Au début du test, l'auditeur recevait des instructions écrites sur papier comme suit :

« Vous allez écouter les stimuli de phrases coréennes, qui varient en intonation et en durée à la fin. Votre tâche est, après avoir écouté un stimulus, de décider quelle sorte d'émotion est exprimée dans le stimulus. Ne cherchez pas à trouver la signification de la phrase ; votre décision doit être purement basée sur l'impression subjective de l'acoustique de la phrase. Dès que vous aurez décidé de l'émotion du stimulus, appuyez sur l'un des deux boutons, marqués comme 'Emotion positive' et 'Emotion négative'. »

Les stimuli étaient présentés à travers un haut-parleur une seule fois dans un ordre aléatoire. Chaque auditeur a eu à juger l'émotion des stimuli deux fois dans deux sessions différentes. Etant donné que le logiciel AVSUPER présentait le stimulus suivant une seconde après que l'auditeur avait donné sa réponse, chaque auditeur a passé le test selon son propre rythme. Le logiciel a enregistré le choix de réponse ('émotion positive' ou 'émotion négative') et le temps de réaction (ms)[132] pour chaque stimulus.

[130] L'AVSUPER est un logiciel dévelopé pour le test de perception, par Mertus (1995) dans le département des sciences cognitives et linguistiques à l'université de Brown aux Etats-Unis. Ce logiciel présente les stimuli, audio ou visuels ou les deux, et enregistre les choix de réponse et le temps de réaction à la fois.

[131] Dans cette expérience, nous avons supprimé la modalité 'neutre' dans les choix de réponse, en considération du fait suivant. Etant donné que les stimuli synthétiques n'étaient pas parfaitement naturels, si l'auditeur recevait trois options, 'positive,' 'neutre' et 'négative,' il aurait facilement choisi l'émotion neutre par paresse, ce qui rendrait le résultat moins net. Afin d'empêcher ce genre de réponse par paresse, il nous fallait forcer l'auditeur à choisir l'une des deux pôles émotionnels, 'positive' et 'négative'.

[132] Le temps de réaction était mesuré par l'intervalle entre la fin du stimulus et le moment où le sujet appuyait sur le bouton, mais il n'est pas pris en compte dans la présente analyse des résultats.

VI.2.1.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 240 réponses des auditeurs coréens et 240 réponses des auditeurs américains, au total 480 réponses (24 stimuli x 10 auditeurs x 2 groupes d'auditeurs). Ce sont les jugements de la deuxième session : les jugements de la première session sont considérés comme apprentissage, donc ils ne sont pas repris dans les données à analyser. Les réponses sont enregistrées comme données en forme dichotomique ('émotion positive' ou 'émotion négative'). Les deux groupes de réponses, coréennes et américaines, sont examinés de manière indépendante, et leurs résultats sont comparés au fur et à mesure que l'analyse a lieu.

En ce qui concerne les réponses coréennes, le test Q de Cochran[133] est effectué pour voir si la perception de l'émotion varie en fonction du contour de Fo et de la durée du stimulus vocal. Dans cette analyse, la fréquence de réponses 'émotion positive' (= réponses POSITIVES) et celle de réponses 'émotion négative' (= réponses NEGATIVES) sont comparées entre quatre niveaux du facteur CONTOUR (*montant*, *descendant*, *montant-descendant* et *plat*) et entre deux niveaux du facteur DUREE (*long* et *court*). Etant donné que le test Q de Cochran ne concerne qu'un facteur à la fois, les deux facteurs, CONTOUR et DUREE, sont analysés de façon indépendante. Pour les analyses des deux facteurs, deux ensembles de données sont établis à la base des mêmes données par la différente structuration des données ; l'un construit par le rangement des données en fonction de quatre niveaux du facteur CONTOUR, l'autre construit par le rangement des données en fonction de deux niveaux du facteur DUREE[134].

Les réponses américaines sont analysées de la même manière que les réponses coréennes, en ce qui concerne la fréquence des réponses POSITIVES et des réponses NEGATIVES en fonction des facteurs CONTOUR et DUREE.

---

[133] Le test Q de Cochran est une analayse non-paramétrique avec mesures répétées pour des données nominales dichotomiques. Dans cette expérience, l'analyse paramétrique comme l'analyse de variance ('ANOVA') n'est pas applicable, parce qu'elle présuppose que la variable dépendante soit sur une échelle d'intervalles, ce qui exige au moins trois modalités dans le test de perception.

[134] Ici, il sembe être utile de préciser la raison de l'utilisation de l'analyse paramétrique dans l'expérience 6, où l'auditeur n'avait que deux choix, 'émotionnel' et 'neutre'. Etant donné qu'il s'agissait de la tâche de la reconnaissance de la présence de l'émotion dans l'expérience 6, plutôt que de la tâche de l'attribution de l'émotion au stimulus auditif (ce qui est fait dans cette expérience), les réponses de l'expérience 6 étaient prises dans les données en termes de l'identification correcte, c'est-à-dire, avec des nombres de réponses correctes pour les trois niveaux du facteur

VI.2.1.4. Résultats

En ce qui concerne les réponses coréennes, le test Q de Cochran montre un effet significatif du facteur CONTOUR sur la fréquence des réponses 'émotion positive' et celle des réponses 'émotion négative' ($q(3)=13,25$, $p<0,01$). Le résultat indique que les contours *montant* et *montant-descendant* produisent plus de réponses POSITIVES que de réponses NEGATIVES tandis que les contours *descendant* et *plat* produisent plus de réponses NEGATIVES que de réponses POSITIVES (voir la Figure 52)[135]. Du point de vue compositionnel, le résultat se traduit comme suit : les contours qui contiennent un élément *montant* (ce que nous appelons les contours ' Montants'), sont perçus émotionnellement plus positifs que ceux qui ne le contiennent pas (ce que nous appelons les contours 'Non-montants'). Quant au facteur DUREE, il y a plus de réponses POSITIVES que de réponses NEGATIVES pour la durée *longue* que pour la durée *courte*, ce qui montre qu'il y a plus de réponses NEGATIVES que de réponses POSITIVES pour la durée *courte* que pour la durée *longue* (voir les chiffres soulignés dans le Tableau 16A). Or, cette différence de réponses POSITIVES et NEGATIVES en fonction des durées, *longue* et *courte*, n'est pas statistiquement significative ($q(1)=0,30$, $p>0,05$). Le fait qu'il y a toujours plus de réponses POSITIVES pour les contours *montant* et *montant-descendant* que pour les contours *descendant* et *plat*, quelle que soit la durée, suggère qu'il n'y a pas d'interaction entre les deux facteurs, CONTOUR et DUREE.

L'analyse des réponses américaines montre des résultats similaires à ceux de l'analyse des réponses coréennes, sauf dans le cas du contour de Fo *plat*. D'après le test Q de Cochran, l'effet du facteur CONTOUR sur la fréquence des réponses POSITIVES et des réponses NEGATIVES est significatif ($q(3)=29,68$, $p<0,01$). Il y a plus de réponses POSITIVES que de réponses NEGATIVES pour les contours *montant* et *montant-descendant*, tandis qu'il y a plus de réponses NEGATIVES que de réponses POSITIVES pour les contours *descendant* et *plat*, (voir la Figure 53). Or, la différence de la fréquence des réponses POSITIVES et des réponses NEGATIVES est moins marquée par comparaison à celle des réponses coréennes. En général, les contours 'Montants' sont perçus émotionnellement plus positifs que les contours 'Non-montants'. L'effet du facteur DUREE n'est pas significatif selon un autre

POSITION (*initiale*, *médiane*, et *finale*). Vu que ces données étaient sur l'échelle d'intervalles, nous avons pu effectuer l'ANOVA avec mesures répétées dans l'expérience 6 (voir V.4.1.3).

[135] Etant donné qu'il n'y a que deux choix de réponse, 'émotion positive' et 'émotion négative,' le résultat des réponses POSITIVE est une image de miroir du résultat des réponses NEGATIVES.

test Q de Cochran (q(1)=2,50, p>0,05), il en est de même pour les réponses coréennes. Malgré l'absence d'effet significatif du facteur DUREE, on constate une tendance selon laquelle la durée *longue* est perçue comme émotionnellement plus positive que la durée courte (voir les chiffres soulignés dans le Tableau 16B). Le fait qu'il y a toujours plus de réponses POSITIVES que de réponses NEGATIVES pour les contours *montant* et *montant-descendant*, quelle que soit la durée des stimuli, nous indique qu'il n'y a pas d'interaction des facteurs CONTOUR et DUREE.

A.

Réponse Positives	Montant		Non-montant		
	Mont.	M-D	Desc.	Plat	Total
Long	**19**	**20**	13	10	62
Court	**18**	**16**	10	14	58
Total	**37**	**36**	23	24	120

Réponse Négatives	Montant		Non-montant		
	Mont.	M-D	Desc.	Plat	Total
Long	11	10	**17**	**16**	54
Court	12	14	**20**	**20**	66
Total	23	24	**37**	**36**	120

B.

Réponse Positives	Montant		Non-montant		
	Mont.	M-D	Desc.	Plat	Total
Long	**22**	**20**	12	16	70
Court	**22**	**16**	8	14	60
Total	**44**	**36**	20	30	130

Réponse Négatives	Montant		Non-montant		
	Mont.	M-D	Desc.	Plat	Total
Long	8	9	**18**	**14**	49
Court	8	14	**22**	**17**	60
Total	16	23	**40**	**31**	110

Tableau 16. Nombre[136] de réponses POSITIVES et de réponses NEGATIVES pour les stimuli dans les huit conditions, y compris quatre contours de Fo (*Montant*, *Descendant*, *Montant-Descendant* et *Plat*) et deux durées (*Long* et *Court*), d'après les réponses coréennes (en haut) et les réponses américaines (en bas).

[136] Etant donné que la phrase-type a été originellement produite au moment où la locutrice n'était pas émotionnellement excitée (autrement dit, la qualité de voix représente l'émotion 'neutre'), l'émotion des stimuli peut être perçue comme soit positive, soit négative. Donc, parmi les 60 réponses totales (POSITIVES et NEGATIVES) pour un contour de Fo donné, il y a la moitié de possibilité de la réponse POSITIVE (N=30) et la moitié de possibilité de la réponse NEGATIVE (N=30). L'excès par rapport à ce nombre-référence (N=30) des réponses POSITIVES ou des réponses NEGATIVES dans un contour donné est considéré comme la préférence peceptuelle de cette émotion ('positive' ou 'négative') pour ce contour. Par exemple, pour le contour de Fo *montant*, il y a 37 réponses POSITIVES et 23 réponses NEGATIVES ; le grand nombre de la réponse POSITIVES par rapport au nombre de réponses NEGATIVES est interprété comme la préférence de l'émotion 'positive' pour le contour de Fo *montant*.

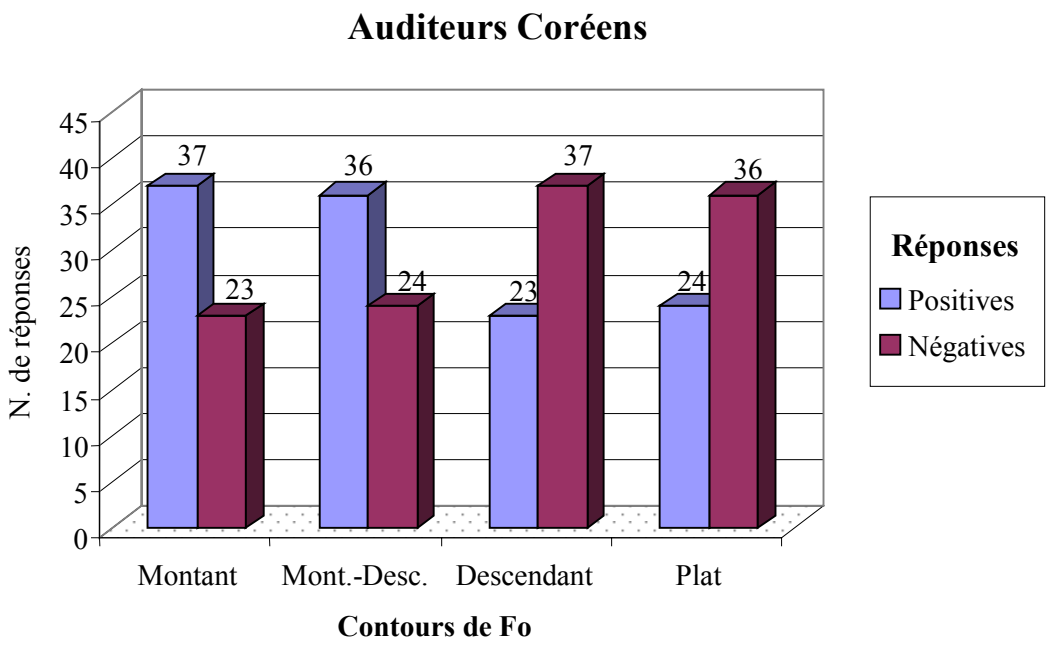


Figure 52. Nombre de réponses POSITIVES et de réponses NEGATIVES pour quatre contours de Fo, *Montant*, *Descendant*, *Montant-Descendant* et *Plat*, d'après les réponses coréennes.

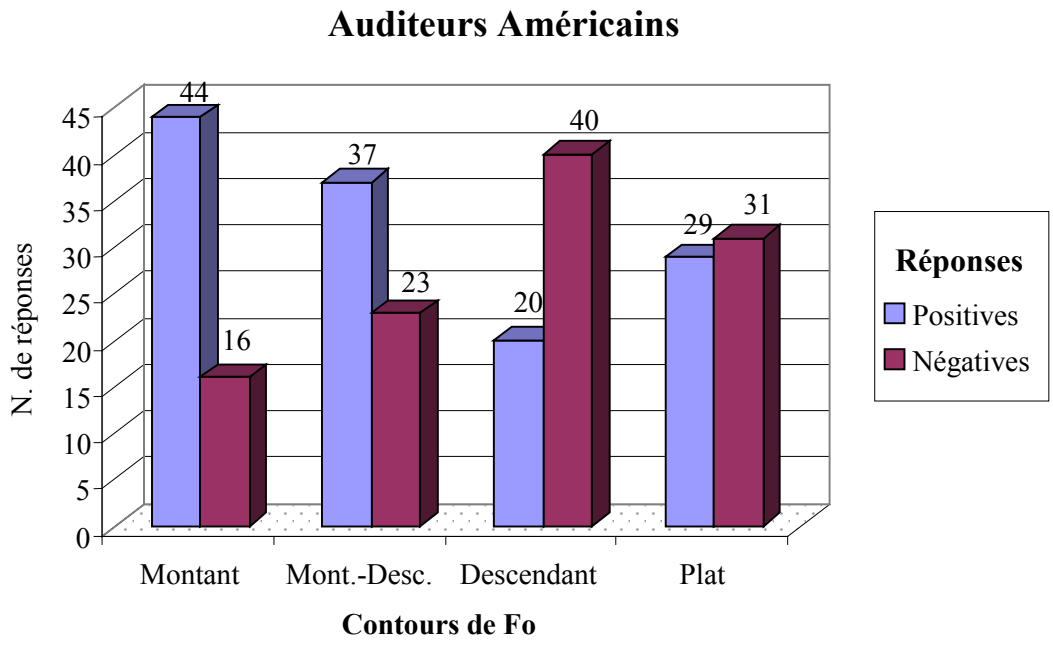


Figure 53. Nombre de réponses POSITIVES et de réponses NEGATIVES pour quatre contours de Fo, *Montant*, *Descendant*, *Montant-Descendant* et *Plat*, d'après les réponses américaines.

VI.2.1.5. Discussions

Dans l'Expérience 7, nous avons appris que la variation du contour de Fo influence la perception de l'émotion, positive ou négative, dans les stimuli auditifs. Le contour *montant* et le contour *montant-descendant*, qui contiennent en commun un élément intonatif 'montant,' favorisent la perception de l'émotion POSITIVE, tandis que le contour *descendant* et le contour *plat*, dits contours 'non-montants,' favorisent la perception de l'émotion NEGATIVE. Cette tendance est comparable pour les Coréens et les Américains, sauf que la contribution du contour *plat* à la perception de l'émotion négative n'est pas aussi évidente chez les Américains que chez les Coréens. Le résultat de l'Expérience 7 supporte notre explication du résultat de l'Expérience 5 à propos de la perception de l'émotion positive dans la partie finale de l'énoncé neutre. Dans l'explication, nous avons proposé la contribution du contour *montant-descendant* et de l'allongement vocalique à la perception de l'émotion positive : c'est-à-dire, la locutrice WJ a exprimé son émotion positive vis-à-vis de l'interlocuteur par une intonation *montante-descendante* et un allongement vocalique ; son intention était communiquée à l'auditeur, et l'auditeur a attribué l'émotion positive à la partie finale de l'énoncé neutre, dans laquelle les indices prosodiques sont présents, lors du test de l'identification de l'émotion[137]. Dans l'Expérience 7, nous nous étions attendue à une influence significative de la durée sur la perception de l'émotion, comme il en est pour le contour de Fo, et nous avons observé plus de réponses POSITIVES pour la durée *longue* et plus de réponses NEGATIVES pour la durée *courte* dans le résultat du test de perception. Pourtant, ce résultat n'a pas atteint le niveau de la significativité statistique (α=0,05). Il semble que la variation fréquentielle soit plus pertinente que la variation temporelle au niveau perceptuel.

La prédominance perceptuelle de l'information fréquentielle sur l'information temporelle a été aussi notée par d'autres chercheurs dans le domaine linguistique et dans le domaine paralinguistique. L'indice fréquentiel l'emporte sur l'indice temporel dans la perception de l'accent linguistique (Fry, 1958) et il en est ainsi pour la compréhension du

[137] Etant donné que nous n'avons pas précisé la distinction conceptuelle entre l'attitude et l'émotion à l'auditeur lors du test de perception, la réponse de l'émotion positive est considérée comme l'équivalence de la réponse de l'attitude polie dans notre interprétation du résultat. Il est possible que la réponse de l'émotion positive dans l'Expérience 5 signifie d'autres attitudes positives, outre la politesse. La vérification de cette possiblité nécessite une autre expérience perceptive.

discours (Swerts, 1993) ; les valeurs de Fo sont les meilleurs indices de l'excitation émotionnelle parmi les traits fréquentiels, temporels et d'amplitude (Scherer, 1982 ; Streeter *et al*., 1983). Sur le plan de la communication émotionnelle, l'intonation est considérée comme le moyen le plus important d'expression de l'émotion (Fónagy, 1990 ; Péter, 1997). La stylisation émotionnelle de la voix synthétique peut être faite essentiellement par l'ajustement du module d'intonation dans le système de la synthèse vocale (Murray & Arnott, 1995, p371). En ce qui concerne la fonction du contour de Fo dans la communication linguistique, Pierrehumbert & Hirschberg (1989) notent que le contour *montant* à la fin d'un énoncé ('*H boundary tone*') signale que l'auditeur doit faire attention à l'énoncé suivant ('forward-reference fonction') tandis que le contour de Fo *descendant* ('*L boundary tone*') n'a pas de telle fonction.

Vu que les traits prosodiques sont polyvalents[138] et multi-fonctionnels[139] dans la communication parlée, le résultat de la contribution du contour *montant* et du contour *montant-descendant* à la perception de l'émotion positive doit être interprété comme l'une des significations possibles de ces contours intonatifs, plutôt que comme la signification absolue de ces contours. En ce qui concerne l'influence des traits prosodiques sur la perception de l'émotion, les études précédentes ont montré leur aspect universel (voir Scherer & Oshinsky, 1977) et l'influence de la culture (voir McAndrew, 1986 ; Wahass & Kent, 1997). Notre résultat de l'Expérience 7 montre les deux aspects à la fois ; la similarité des Coréens et des Américains en ce qui concerne les contours de Fo *montant*, *montant-descendant* et *descendant*, et la divergence des deux groupes d'auditeurs en ce qui concerne le contour de Fo *plat*. Il nous semble précoce de conclure sur l'universalité et la spécificité à partir de notre résultat actuel. Ce genre de conclusion nécessite des expériences avec plus de paramètres manipulés et plus d'auditeurs dont les cultures sont suffisamment diversifiées.

[138] La nature polyvalente des traits prosodiques réfère à l'absence de relation univoque entre le trait prosodique et la catégorie émotionnelle (voir la page 186).

[139] La variation de Fo signale diverses informations, telles la modalité linguistique, l'émotion, l'attitude, le sexe, l'âge et la personnalité du locuteur (voir Cooper & Sorensen, 1981, p171-175).

VI.2.2. Expérience 8 : Vérification par la synthèse du rôle des différentes parties de l'énoncé

L'Expérience 8 examine si le placement des traits prosodiques dans les parties initiale, médiane et finale de l'énoncé influence la perception de l'émotion. Comme le cas de l'Expérience 7, cette expérience a aussi pour fonction de vérifier le phénomène observé dans l'Expérience 5 (voir la page 185) par synthèse. Parmi les trois facteurs suggérés pour l'explication de la perception de l'émotion positive dans les énoncés neutres (tels le contour de Fo *montant-descendant*, la durée *longue* et la position finale de ces traits prosodiques), l'Expérience 7 a démontré la contribution significative à la perception de l'émotion positive du contour de Fo *montant-descendant* plutôt que la contribution de la durée *longue*. L'Expérience 8 examine le dernier facteur, la position finale des traits prosodiques, en adressant la question si le placement de cet indice prosodique dans la position finale de l'énoncé joue un rôle pertinent dans la perception de l'émotion positive. Spécifiquement, est-ce que la présence du contour de Fo *montant-descendant* dans la partie finale de l'énoncé favorise la perception de l'émotion positive ?

VI.2.2.1. Préparation des données

Les stimuli de l'Expérience 8 sont construits à partir des trois syllabes ([de], [ne] et [ge])[140] et de la phrase-type, qui ont été utilisées dans l'Expérience 7 (Voir VI.2.1.1). Ces syllabes partagent en commun le contour de Fo *montant-descendant* et une durée d'environ 500ms. Une autre syllabe [gu], qui est la syllabe finale de la phrase-type originale, est ajoutée aux stimuli de l'Expérience 8, afin d'augmenter le nombre de stimuli du test de perception. Elle a aussi le contour *montant-descendant* avec une durée de 520ms. La variation de Fo au cours de la syllabe [gu] est de 189 → 235 → 181Hz. Les Fo moyens, maximum et minimum de cette syllabe sont respectivement de 209, 186 et 235Hz. La valeur de jitter de cette syllabe est de 0,1%[141]. Les quatre syllabes ('stimuli de cible') sont insérées dans les parties initiale, médiane et finale de la phrase-type, de façon respective.

[140] La version originale de ces syllabes est employée dans cette expérience, tandis que les versions synthétiques de ces syllabes étaient utilisées dans l'Expérience 7 (voir VI.2.1.1).
[141] Comparer avec les valeurs acoustiques des trois syllabes, [de], [ne] et [ge], dans le Tableau 15.

L'exemple du placement des stimuli de cible dans les trois positions initiale, médiane et finale de la phrase-type est illustré dans la Figure 54.

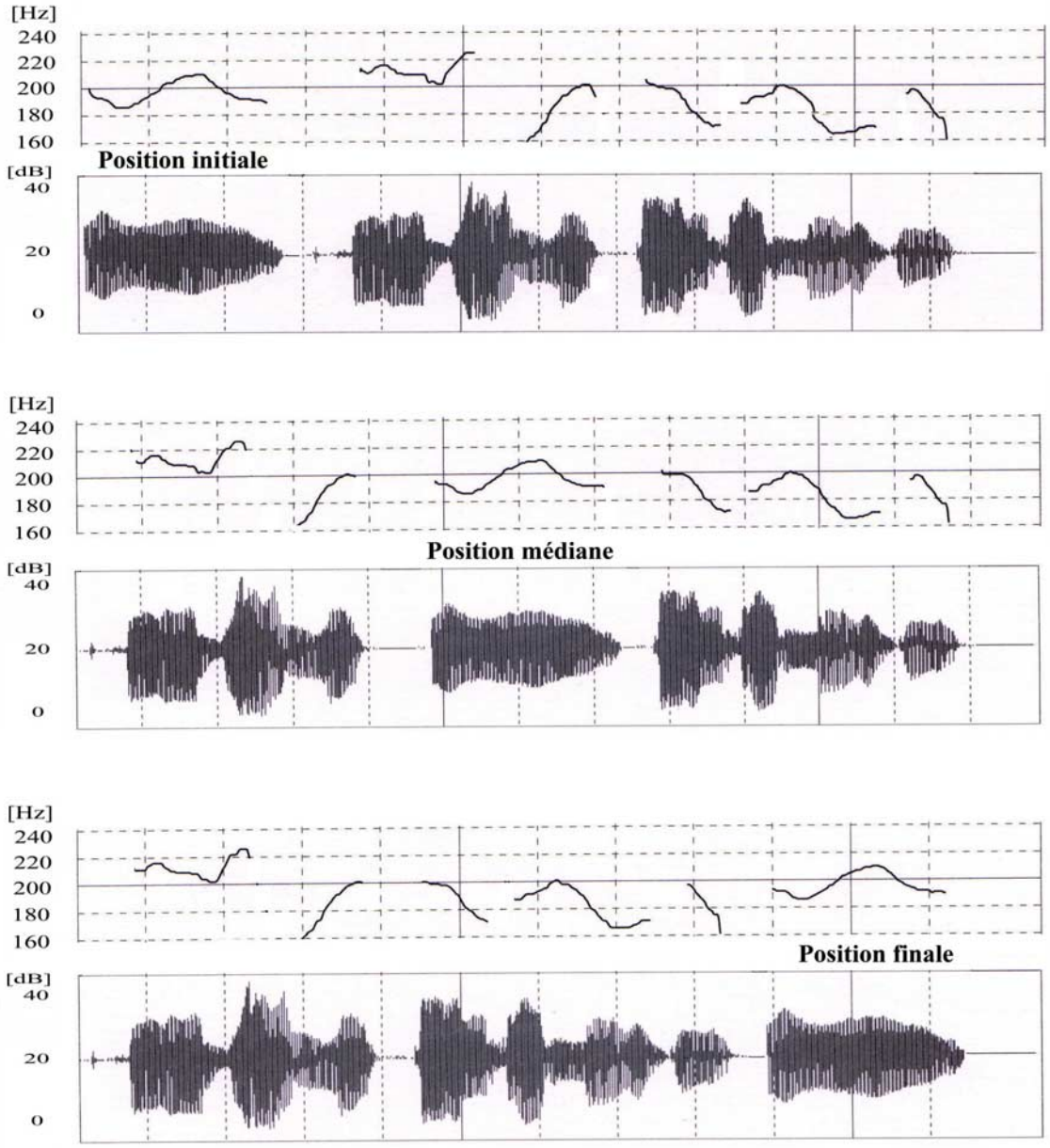


Figure 54. Placement du contour de Fo *montant-descendant* dans les trois positions, initiale, médiane et finale, de la phrase-type.

Dans ce genre de manipulation, il est essentiel d'assurer la naturalité des stimuli résultant de la resynthèse. Cette précaution prise, l'insertion des stimuli de cible dans la phrase-type est faite en respectant les frontières de mot, pour que les phrases resynthétisées ne soient pas choquantes à l'oreille de l'auditeur. Ainsi, la continuité perceptuelle du contour intonatif de la phrase resynthétisée est gardée dans la mesure du possible. La resynthèse est faite à l'aide du logiciel 'Winpitch'[142]. A la suite des manipulations, 12 phrases sont préparées par l'insertion des quatre syllabes de cible ([de], [ne], [ge] et [gu]) dans les trois parties (initiale, médiane et finale) de la phrase, en tant que stimuli du test de perception.

VI.2.2.2. Test de perception

Dix Coréens et dix Américains, qui n'ont pas participé aux études précédantes, ont participé au test de perception. Ce sont des natifs de la langue coréenne et de la langue anglaise. La plupart des auditeurs sont des étudiants de l'université, il y a aussi une secrétaire et une infirmière. Leurs âges sont de vingt ans à trente-trois ans. La moitié des auditeurs était des hommes et l'autre moitié des femmes. Ils se sont portés volontaires pour le test. L'auditeur a passé le test de perception individuellement avec le logiciel 'AVSUPER'[143] dans une pièce insonorisée. L'auditeur devait décider quelle sorte d'émotion est exprimée dans le stimulus, en appuyant sur l'un des deux boutons étiquetés comme 'émotion positive' et 'émotion négative'[144]. Au début du test, l'auditeur recevait des instructions écrites sur papier comme suit :

« Vous allez écouter des phrases coréennes qui ont des intonations similaires. Votre tâche est, après avoir écouté un stimulus, de décider quelle sorte d'émotion est exprimée dans le stimulus. Ne cherchez pas à trouver la signification de la phrase ; votre décision doit être purement basée sur l'impression subjective de l'émotion à propos de l'intonation de la phrase. Dès que vous aurez décidé de l'émotion du stimulus, appuyez sur l'un des deux boutons marqués comme 'Emotion positive' et 'Emotion négative'. »

[142] Voir la note n°52 pour la description du logiciel 'Winpitch'.
[143] Voir la note n°130 pour la description du logiciel 'AVSUPER'.
[144] Comme le cas de l'Expérience 7, la modalité 'neutre' est exclue des choix de réponse dans l'Expérience 8. La raison de cette exclusion est expliquée dans la note n°131.

L'auditeur a passé le test lors de deux sessions, séparées par 10 minutes d'intervalle. Dans la session d'exercice, il a eu à juger l'émotion des stimuli présentés à travers un haut-parleur, une seule fois et dans un ordre aléatoire. Dans le test principal, il a jugé l'émotion des stimuli trois fois et dans un ordre différent. Etant donné que le logiciel AVSUPER présentait le stimulus suivant une seconde après que l'auditeur avait donné sa réponse, chaque auditeur a passé le test selon son propre rythme. Le logiciel a enregistré le choix de réponse ('émotion positive' ou 'émotion négative') et le temps de réaction (ms) pour chaque stimulus dans un fichier informatique. Les choix de réponse seuls sont pris en compte dans nos données à analyser.

VI.2.2.3. Analyse statistique

A la suite du test de perception, nous avons obtenu 360 réponses des auditeurs coréens et 360 réponses des auditeurs américains, soit un total de 720 réponses (12 stimuli x 10 auditeurs x 3 jugements x 2 groupes d'auditeurs). Ce sont des jugements du test principal. Les réponses sont enregistrées dans les données en forme dichotomique ('émotion POSITIVE' ou 'émotion NEGATIVE'). Les deux groupes de réponses, coréennes et américaines, sont examinés de manière indépendante, et leurs résultats sont comparés à la fin.

Le test Q de Cochran[145] est effectué pour tester si la perception de l'émotion varie en fonction du placement du trait prosodique (contour de Fo *montant-descendant*) dans la partie *initiale*, *médiane* et *finale* de l'énoncé. Ce test examine si une réponse, POSITIVE ou NEGATIVE, est particulièrement fréquente dans un groupe de stimuli, comme la position *initiale*, *médiane* ou *finale*. Etant donné que le test Q de Cochran, l'analyse non-paramétrique avec mesures répétées, ne permet pas de comparaison à posteriori (test post-hoc), la fréquence de la réponse POSITIVE ou de la réponse NEGATIVE est comparée deux à deux entre les trois niveaux du facteur POSITION (*initiale*, *médiane* et *finale*) de manière respective.

[145] Voir la note n°133 pour la descritpion du test Q de Cochran.

VI.2.2.4. Résultat

En ce qui concerne les réponses coréennes, le test Q de Cochran montre un effet significatif du facteur POSITION sur la fréquence des réponses POSITIVES et des réponses NEGATIVES ($q(2)=15,05$, $p<0,01$). La fréquence des réponses POSITIVES varie significativement selon la position de l'indice prosodique (contour de Fo *montant-descendant*) dans la phrase-type (voir la Figure 55). D'après la comparaison deux à deux, la réponse POSITIVE est significativement plus fréquente quand l'indice prosodique est présent dans la position *initiale* ou dans la position *finale* que quand il est présent dans la position *médiane*. La fréquence de la réponse POSITIVE est plus élevée quand l'indice prosodique est placé dans la position *finale* que quand il l'est dans la position *initiale* (différence non-significative). Etant donné qu'il n'y avait que deux modalités de réponse (POSITIVE et NEGATIVE) dans le test de perception, le nombre de réponses POSITIVES est le complémentaire du nombre de réponses NEGATIVES dans les 120 réponses totales pour un niveau du facteur (*initiale*, *médiane* ou *finale*), donc le résultat concernant les réponses POSITIVES est l'équivalent du résultat concernant les réponses NEGATIVES. Vu l'émotion neutre de la phrase-type originelle[146], nous nous sommes attendue à une moitié de réponses POSITIVES et une moitié de réponses NEGATIVES dans le test de perception avec le choix forcé ('positive' ou 'négative'). Or, le résultat du test montre plus de réponses NEGATIVES que de réponses positives, ce qui semble être dû à la nature des stimuli synthétiques. L'insertion d'une syllabe extérieure à l'intérieur de la phrase confère une impression artificielle à la phrase synthétique et introduit une bizarrerie sémantique de la phrase, ce qui aurait dû affecter particulièrement les auditeurs coréens. Malgré le faible nombre de la réponse POSITIVE, la variation du nombre de réponses POSITIVES en fonction des positions (*initiale*, *médiane* et *finale*) de l'indice est statistiquement significative, comme il en est pour les réponses NEGATIVES. L'essentiel de notre résultat est que la proportion de la réponse POSITIVE et de la réponse NEGATIVE. est différente en fonction de la position de l'indice dans les positions initiale, médiane et finale de la phrase. La distribution des réponses POSITIVES en fonction des trois positions de l'indice dans la phrase-type est présentée dans la Figure 55 et celle des réponses NEGATIVES est présentée dans la Figure 56.

[146] La phrase-type a été originnellement produit au moment où la locutrice n'était pas émotionnellement excitée, elle fait partie de la catégorie d'émotion neutre, d'après l'Expérience 1 (voir Tableau 7).

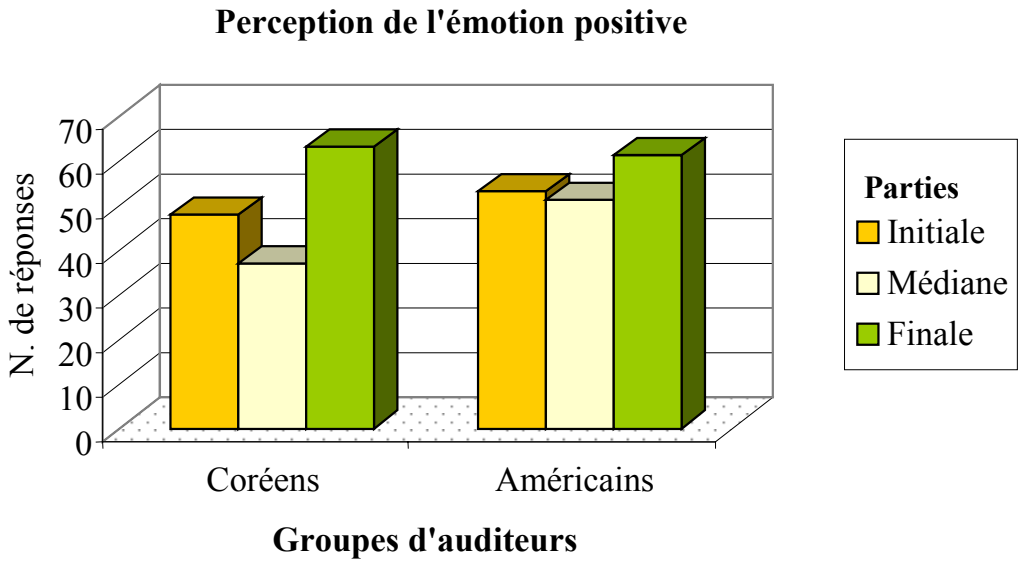


Figure 55. Nombre de réponses POSITIVES en fonction des positions (*initiale*, *médiane* et *finale*) de l'indice dans la phrase-type ; d'après les réponses coréennes et américaines.

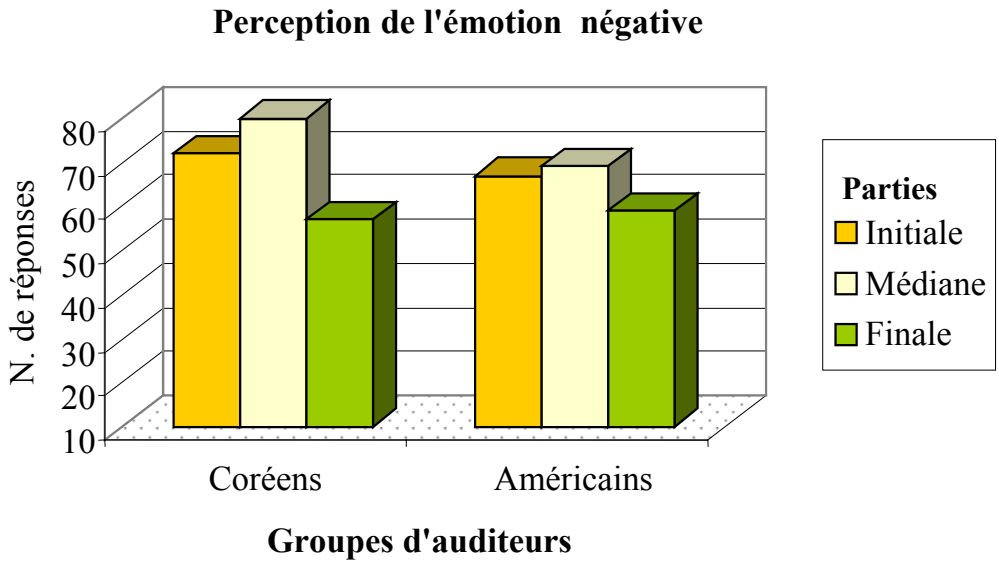


Figure 56. Nombre de réponses NEGATIVES en fonction des positions (*initiale*, *médiane* et *finale*) de l'indice dans la phrase-type ; d'après les réponses coréennes et américaines.

Quand aux réponses américaines, comme le cas des réponses coréennes, la réponse POSITIVE est la plus nombreuse quand l'indice prosodique est placé dans la position *finale* de la phrase, et elle est la moins nombreuse quand l'indice prosodique est placé dans la position *médiane* de la phrase (voir la Figure 55). Pourtant, la variation du nombre de réponses POSITIVES en fonction des différentes positions de l'indice prosodique n'est pas

suffisamment grande que pour atteindre le niveau de la significativité statistique ($q(2)=3,65$, $p>0,05$). Vu que la phrase-type est en coréen, le manque de connaissance de la langue coréenne de l'auditeur Américain semble avoir rendu peu pertinent le différent placement de l'indice dans les parties (*initiale*, *médiane* et *finale*) de la phrase, ce qui aurait dû résulter en l'absence d'effet significatif du facteur POSITION dans les réponses américaines. Parmi ces réponses, le nombre de réponses POSITIVES pour le placement *initial* de l'indice dans la phrase est presque aussi bas que celui pour le placement *médian* de l'indice, à la différence du résultat des réponses coréennes dans lequel le nombre de réponses POSITIVES pour le placement *initial* de l'indice est significativement plus haut que celui pour le placement *médian* de l'indice.

En résumé, l'émotion de la phrase est plus souvent perçue comme positive quand le contour de Fo *montant-descendant* est présent dans la position *finale* de la phrase que quand il est présent dans les autres positions, *initiale* et *médiane*, de la phrase. Le placement du contour *montant-descendant* dans la position *initiale* de la phrase contribue secondairement à la perception de l'émotion positive, après la contribution majeure du placement dans la position *finale* de la phrase. Si ces tendances sont trouvées chez tous les deux groupes d'auditeurs, elle est pourtant beaucoup plus évidente pour les Coréens que pour les Américains.

VI.2.2.5. Discussion

L'Expérience 8 a démontré que le placement d'un indice acoustique dans les différentes parties (*initiale*, *médian*e ou *finale*) de la phrase influence la perception de l'émotion de cette phrase. Le placement du contour de Fo *montant-descendant* dans la partie *finale* de la phrase favorise la perception de l'émotion positive plus que celui dans la partie *médiane*. Le placement de ce contour dans la partie *initiale* produit un effet intermédiaire. L'efficacité de la partie finale de la phrase dans la communication parlée est souvent notée par les chercheurs dans les différents aspects linguistiques et paralinguistiques. Le contraste de l'intonation *montante* et de l'intonation *descendante* pour la distinction de la phrase affirmative et de la phrase interrogative a lieu surtout à la fin de la phrase. La modification du contour intonatif pour l'ajout sémiotique de modalité au message référentiel est surtout faite à la fin de la phrase. Par exemple, le renforcement de l'intonation *descendante* en fin de phrase exprime l'attitude d'assurance ou d'impérativité

du sujet parlant, tandis que l'inverse de cette intonation en fin de phrase (résultant en une intonation finale *montante*) exprime l'attitude d'incertitude ou d'inquiétude du sujet parlant (Léon, 1993, p148). En ce qui concerne la synthèse de parole, l'intonation finale de la phrase joue un rôle essentiel dans l'ajout de l'impression émotionnelle à la parole synthétique (Mozziconacci, 1998, p103). Notre résultat de l'Expérience 8 va dans le même sens pour l'efficacité de la partie finale de la phrase, en tant que position préférable de l'indice émotionnel.

Avant de conclure la discussion de notre Expérience 8, il faut mentionner une autre interprétation possible du résultat de cette expérience. Etant donné que la syllabe du contour de Fo *montant-descendant* est originellement extraite à de la position finale de l'énoncé, on peut se demander si la meilleure perception de l'émotion positive avec les syllabes insérées à la fin de la phrase-type est simplement due à la cohérence entre la positon originelle des syllabes et leur position dans la phrase synthétique. Tout en admettant cette confusion potentielle au niveau méthodologique, nous insistons pourtant sur la validité de notre interprétation des résultats dans VI.2.2.4 avec recours des évidences indirectes, explicitées par la suite. Si le grand nombre des réponses POSITIVES pour le placement *final* de l'indice (contour de Fo *montant-descendant*) dans la phrase est purement dû à la cohérence générale entre la position de l'indice dans la phrase originale et celle dans la phrase synthétique, le nombre des réponses POSITIVES pour le placement *initial* de l'indice et celui pour le placement *médiane* de l'indice doivent être les mêmes, parce que les deux placements concernent les positions différentes de l'originale ; mais le placement *initial* de l'indice a produit plus de réponses POSITIVES que le placement *médian* de l'indice, ce qui est hautement visible dans les réponses coréennes. Cette différence supporte l'existence de l'effet de la position de l'indice dans la phrase sur la perception de l'émotion. Donc, nous concluons que la partie *finale* de la phrase en tant que lieu de l'indice prosodique a un statut spécial au niveau perceptuel, par rapport aux parties *initiale* et *médiane* de la phrase.

VI.3. Conclusion du chapitre VI.

Dans le chapitre VI, nous avons vu que la modification des traits prosodiques (comme le contour intonatif et la durée) influence la perception de l'émotion sur l'axe de valence ('positive' vs. 'négative') et que le placement du trait prosodique dans les différentes parties (*initiale*, *médiane* et *finale*) de la phrase influence aussi le même genre de perception émotionnelle. Les contours contenant un élément intonatif 'montant' (ex. le contour *montant* et le contour *montant-descendant*) orientent la perception de l'émotion vers le sens positif, tandis que les contours du type 'non-montant' (ex. le contour *descendant* et le contour *plat*) orientent la perception de l'émotion vers le sens négatif. Cette orientation s'effectue de façon relative mais avec une tendance cohérente. Les Coréens et les Américains montrent une tendance similaire en ce qui concerne l'attribution de l'émotion aux contours de Fo *montant*, *montant-descendant* et *descendant*, mais ils diffèrent en ce qui concerne l'attribution de l'émotion au contour de Fo *plat* ; les Coréens attribuent l'émotion négative à ce dernier contour de façon explicite tandis que les Américains attribuent un peu plus d'émotion négative que d'émotion positive au contour de Fo *plat*. L'allongement de la durée semble aussi favoriser la perception de l'émotion positive dans notre Expérience 7, pourtant cet effet n'est pas significatif. La pertinence perceptuelle de la partie finale de la phrase, par rapport aux parties initiale et médiane, est confirmée dans l'Expérience 8. Le placement du contour de Fo *montant-descendant* dans la partie *finale* de la phrase favorise la perception de l'émotion positive plus que celui dans les parties *initiale* et *médiane* de la phrase. Cette tendance est trouvée en commun chez les Coréens et les Américains, même si elle n'est pas statistiquement significative chez ces derniers. La correspondance majeure des résultats des Coréens et de ceux des Américains dans les deux expériences de ce chapitre nous mène à la conclusion de l'universalité de la perception de l'émotion. Pourtant, vu les différences subtiles, nous réservons la conclusion définitive à un travail ultérieur.

Du point de vue de la modélisation de l'émotion vocale (voir II.4.2), les discussions de ce chapitre concernent les déterminants externes, plutôt que les déterminants internes, de la production de la voix émotionnelle. L'effet des déterminants externes (effet 'pull') consiste en l'expression de l'intention ou de l'émotion du locuteur vis-à-vis de

l'interlocuteur (ce qui se passe en majorité au niveau conscient), tandis que l'effet des déterminants externes (effet 'push') consiste en la manifestation de l'excitation émotionnelle et de l'état physiologique du locuteur, ce qui n'est pas toujours contrôlable par la conscience. Les syllabes utilisées en tant que stimuli des expériences de ce chapitre ont été prononcées dans un état où la locutrice WJ n'était pas émotionnellement excitée (disons neutre)[147]. Pourtant, une certaine modification du contour de Fo et de la durée de ces syllabes, par la locutrice ou par la synthèse vocale, a créé l'impression émotionnelle positive pour l'auditeur, ce qui est interprété comme la communication de l'intention émotionnelle par la modification des traits prosodiques. Ce genre de modification est considéré comme l'effet 'pull,' motivé par un but communicatif du sujet parlant dans l'élaboration de la modalité de ses énoncés. Dans ce chapitre, nous avons vu que l'effet de la modification des traits prosodiques dans la partie finale de l'énoncé est particulièrement efficace, par rapport à celui dans la partie initiale ou médiane de l'énoncé. Or, cela n'explique pas directement les résultats de nos autres expériences (Expérience 5 et Expérience 6), à savoir pourquoi l'excitation émotionnelle du locuteur est particulièrement mieux exprimée dans la partie finale de l'énoncé que dans les autres parties, initiale et médiane. Dans ce dernier cas, il s'agit des déterminants internes de la production de l'émotion vocale. Afin de confirmer la particularité de la partie finale de l'expression émotionnelle, concernant non seulement les déterminants externes mais aussi les déterminants internes, il est nécessaire d'effectuer une autre expérience, similaire à l'Expérience 8, mais avec des stimuli produits dans un état émotionnel (par exemple, bouleversé par la détresse). Cette expérience attend notre prochain travail.

[147] Voir VI.2.1.1.

Chapitre VII

Discussions et conclusion générale

Dans ce dernier chapitre, nous résumons les apports de ce travail et présentons la perspective de notre recherche.

VII.1. Récapitulation des discussions

Dans le présent travail, nous avons poursuivi une série d'analyses acoustiques et perceptives sur l'expression et la perception de l'émotion extraite de parole spontanée. Le résumé des analyses se trouve en entête de chaque chapitre. Nous nous contenterons de récapituler ci-dessous les points qui nous semblent originaux dans cette thèse, du moins dans la limite de notre connaissance de la littérature sur ce sujet. Nos analyses acoustiques ont visé à mieux connaître les meilleurs indices acoustiques, corrélats de l'expression de la joie et de la tristesse, en mesurant neuf paramètres : le Fo moyen, le Fo maximum, le Fo minimum, la moyenne des 20% des valeurs les plus basses de Fo ('*Fo Moy Bas*'), la plage de Fo, la perturbation de Fo ('jitter'), la perturbation d'intensité ('shimmer'), le débit et la distribution spectrale. L'augmentation observée du Fo maximum et de la plage de Fo par l'excitation émotionnelle confirme les résultats des études précédantes (Hutter, 1968 ; Streeter *et al.*, 1983 ; Protopapas & Lieberman, 1995). Par contre, nos résultats suggèrent que la mesure du Fo moyen n'est pas toujours une bonne solution pour détecter l'excitation émotionnelle ; ce n'était valable que pour la joie dans le cas de notre corpus coréen. Le Fo minimum et le 'Fo Moy Bas,' ce dernier étant proposé dans cette étude à la suite d'une suggestion de Klaus Scherer (communication personnelle), se sont révélés des bons indices de l'excitation émotionnelle de la tristesse (voix larmoyante).

Les analyses perceptives ont examiné comment l'auditeur perçoit la joie et la tristesse dans différentes conditions d'audition. Les paramètres étudiés sont la connaissance culturelle de l'auditeur, la taille des stimuli présentés à l'auditeur, le contour intonatif et la durée des stimuli synthétiques et la position de la partie dans l'énoncé. Le fait que l'identification de la joie et de la tristesse de la locutrice coréenne par des auditeurs coréens, français et américains soit d'une précision supérieure à celle qui aurait été due au hasard confirme l'universalité de la perception de l'émotion, fréquemment notée dans les théories

de l'émotion (Tomkins, 1962 ; Plutchik, 1980). La meilleure identification des Coréens parmi les trois groupes d'auditeurs montre également l'influence de la culture sur la communication de l'émotion, en accord avec les résultats de Kramer (1963), de McCluskey et al. (1975) et de Van Bezooyen (1984). La perception de l'émotion est aussi influencée par la position de la partie (initiale, médiane et finale) dans l'énoncé présentée à l'auditeur : l'émotion est mieux exprimée et mieux reconnue dans la partie finale de l'énoncé que dans les parties initiale et médiane. Nous avons attribué cette meilleure identification de l'émotion dans la partie finale de l'énoncé à la présence des vibrations glottales extrêmement irrégulières et ralenties vers la fin de l'énoncé en cas de détresse (voix larmoyante) et à la présence d'un contour montant-descendant sur la dernière syllabe de l'énoncé lors de l'expression de l'émotion positive vis-à-vis de l'interlocuteur. Ces derniers phénomènes ont été observés par quelques chercheurs (Hecker *et al.* 1968 ; Fonagy, 1983 ; Léon, 1993 ; Mozziconacci, 1998), mais cette observation n'avait jamais été prouvée avec des évidences expérimentales: c'est donc chose faite. Notre démonstration de *l'efficacité de la partie finale de l'énoncé dans la l'expression et la perception de l'émotion* à travers des expériences systématiquement manipulées apporte une nouvelle confirmation des caractéristiques de la parole émotionnelle dans la recherche de la communication de l'émotion vocale.

VII.2. Perspective de la recherche

En traitant les données de nature spontanée, nous avons dû limiter le nombre de locuteurs pour maximiser le contrôle expérimental. L'utilisation de parole spontanée est un problème majeur, car il faut recueillir beaucoup de données pour en extraire des parties qui correspondent à certains critères prédéfinis. Il faudrait vérifier les résultats de nos expériences avec un plus grand nombre de locuteurs provenant d'autres cultures. L'étude de la variabilité des expressions émotionnelles entre les individus à l'intérieur de la culture est également digne d'intérêt . Seule une recherche collective, internationale peut résoudre ce problème. A ces différences inter et intra cultures intéressantes, il serait aussi intéressant de comparer l'expression des émotions chez les hommes et chez les femmes. Un autre sujet intéressant serait de comparer l'expression de l'émotion *vécue* et l'émotion *stylisée* (le même locuteur devrait reproduire plus tard son propre état émotionnel). Vu la rareté de ce genre de données émotionnelles, les données ainsi acquises seraient précieuses. Dans ce travail, nous n'avons traité que l'expression vocale de l'émotion. L'interaction ou la compensation entre les différentes modes d'expression émotionnelle, vocale, faciale et gestuelle, n'étant pas négligeable (Hess *et al.*, 1988), une recherche interdisciplinaire sur ce genre d'interaction contribuerait à mieux comprendre la communication de l'émotion. Les nouvelles technologies de stockage (DVD via Internet, etc.) permettront l'échange de données multimodales. Les paramètres comme l'intensité moyenne, les fréquences de formant, la forme de la vibration glottale, la fréquence des pauses (sonores et silencieuses) ne sont pas inclus dans notre analyse acoustique du présent travail. Ces paramètres étant facilement mesurables avec la technologie actuelle, il faut les inclure en espérant qu'ils capteront les variations prosodiques plus subtiles de l'expression vocale de l'émotion.

Dans cette étude, nous avons pu démontrer la répartition non uniforme des indices acoustiques de l'émotion dans les parties initiale, médiane et finale de l'énoncé, utilisant les données de la parole naturelle en coréen et en anglais. Nous avons suggéré que l'irrégularité et le ralentissement des vibrations glottales par l'excitation émotionnelle de la tristesse peuvent être considérées comme la manifestation d'un geste involontaire (exprimé sans intervention consciente du locuteur) et la configuration du contour de Fo *montant-descendant* en fin d'énoncé en cas d'expression de l'émotion positive vis-à-vis de

l'interlocuteur comme la manifestation d'un trait volontaire (motivé par l'intention du locuteur). Il est connu que la conscience des gens dans l'interaction sociale est apte à refouler l'émotion négative (comme la tristesse) et à exprimer l'émotion positive[148]. Cette distinction supposée entre les gestes émotionnels plus ou moins contrôlés et involontaires, qui se réfèrent respectivement au modèle de configuration et au modèle de convariation, nous paraît plus utile que l'argumentation de la plausibilité absolue d'un modèle ou de l'autre. De plus, il serait intéressant de voir dans quelle mesure chaque trait supposé volontaire ou involontaire influence le résultat d'expression vocale de l'émotion. Étant donné que notre analyse par synthèse de l'efficacité de la partie finale de l'énoncé dans la communication émotionnelle n'a traité que les stimuli de trait volontaire (le contour de Fo *montant-descendant* en tant qu'indice de l'émotion positive vis-à-vis de l'interlocuteur), nous comptons effectuer une expérience similaire mais avec les stimuli de trait supposé involontaire (une partie de l'énoncé dont la variation glottale est extrêmement irrégulière)[2]. Ainsi, nous pourrons compléter la vérification par synthèse de notre proposition du statut spécial de la partie finale de l'énoncé dans la communication de l'émotion, non seulement avec les données de la parole naturelle mais aussi avec les données de la parole synthétique.

La collaboration avec les ingénieurs est indispensable pour l'application de l'ensemble de ces connaissances dans le domaine pratique des technologies vocales. Les études futures ne peuvent être que multidisciplinaires pour être crédibles. En ce qui concerne les applications potentielles de ce type d'étude, il est suggéré dans notre travail que l'adéquation de la voix synthétique au contexte peut être améliorée par l'ajout de l'effet émotionnel. L'ajout d'un contour intonatif montant-descendant en fin d'énoncé peut créer une impression émotionnelle positive dans la voix synthétique et une glottalisation extrême dans la voix synthétique peut signaler un état de détresse[149]. La modification des indices acoustiques par l'excitation émotionnelle explique en partie la détérioration des résultats de la reconnaissance de la parole et du locuteur lorsque le sujet parle sous l'effet de l'émotion. Cet état de fait est difficilement améliorable puisque l'état émotionnel n'est pas

[148] Ici, nous parlons de la perte de contrôle cognitif. A propos de l'irrégularisation et le ralentissement des vibrations glottales sous le coup de détresse, Williams & Stevens (1972) ont même parlé de la perte de contrôle au niveau articulatoire inconscient.

[149] Nous sommes consciente du fait que ces traits sont polyvalents comme d'autres traits prosodiques. Le contour de Fo *montant-descendant* peut être aussi utilisé pour l'accentuation ou la focalisation d'une partie de l'énoncé (Pierrehumbert & Hirschberg, 1989 ; Chung & Kenstowicz, 1997), et un certain degré de glottalisation est souvent présent dans la voix sans émotion particulière, éventuellement servant d'une fonction démarcative (Dilley *et al.,* 1996),

connu à priori. Nos résultats d'expériences informent que l'état émotionnel du locuteur peut être repéré par des indices acoustiques, essentiellement localisés dans la fin de l'énoncé. Il pourrait être suggéré aux ingénieurs de baser leur recherche de critère sur les parties finales des énoncés. Le repérage de l'état émotionnel du locuteur peut être utile dans la communication homme-machine (tout au moins dans un temps futur) et déjà dans le milieu médical. Le développement de la reconnaissance de l'émotion du locuteur serait aussi utile pour évaluer automatiquement l'état psychologique des malades sous antidépresseur en tant qu'évaluation de l'efficacité des médicaments antidépresseurs. Des réalisations limitées existent déjà. Malgré des difficultés conceptuelles et techniques, la recherche de l'implémentation de l'émotion dans l'ordinateur (via la synthèse et la reconnaissance automatique de la parole) est très active actuellement. Par exemple, l'entreprise informatique IBM, pour développer son image de marque d'une équipe toujours en avance sur son temps, a lancé récemment plusieurs projets pour élaborer l'ordinateur qui peut reconnaître l'état émotionnel du sujet devant l'ordinateur sur la base des expressions vocale, faciale et gestuelle. La recherche de Cassell sur l'agent automatique de la conversation animée (Cassell *et al*., 1994) et celle de Picard sur l'ordinateur avec émotion ('*affective computing*,' 1995) dans le laboratoire de multimédia à MIT visent aussi à l'amélioration de l'interaction entre l'homme et la machine dans le domaine de l'intelligence artificielle.

Nous pensons avoir montré dans cette thèse l'intérêt d'étudier l'émotion vocale avec les données de la parole spontanée. Ce genre de données n'est pas facile à analyser par sa complexité, mais c'est grâce à cette complexité que nous pouvons communiquer avec d'autres nos divers sentiments, idées, propositions, etc. Selon l'expression de Fónagy (1983, p19), *tous les sons concrets sont expressifs*. Cette expressivité révélant l'état émotionnel et le style personnel du locuteur rend la communication parlée riche et intéressante. L'émotion est exprimée dans la parole souvent sous formes discrètes. Pourtant, elle ne doit plus être considérée comme un accessoire secondaire de la parole mais elle doit être prise en compte dans l'étude phonétique en tant que l'un des aspects essentiels de la communication langagière. Vu cette importance de l'étude de l'aspect émotionnel de la parole, nous comptons à continuer dans cette voie de recherche pour mieux comprendre le mécanisme de *l'expression et la perception de l'émotion dans la communication parlée.*

Résumé

La parole est un moyen de communication extrêmement riche et subtile. Elle véhicule non seulement de l'information linguistique référentielle mais aussi entre autres de l'information sur la personnalité et sur l'état émotionnel du locuteur. L'émotion est une réponse motivationnelle et adaptative d'un organisme à l'environnement social. Elle est souvent présente dans la parole naturelle, alors qu'elle n'est guère prise en compte à l'heure actuelle dans les systèmes de synthèse et de reconnaissance automatique de la parole et du locuteur. D'où l'impression mécanique de la parole synthétique et un accroissement de l'instabilité des résultats de la reconnaissance automatique lorsque le locuteur parle sous le coup d'émotion. Afin de contribuer à la résolution de ces problèmes, la présente thèse étudie comment l'émotion modifie la production vocale du locuteur et comment l'auditeur perçoit l'émotion en fonction des facteurs acoustiques, linguistiques et culturels.

La dissertation consiste en une série d'analyses acoustiques et perceptives de la parole émotionnelle. Les données de base ont été acquises à partir d'extraits d'émissions télévisées, Coréennes et Américaines, enregistrées sur des cassettes vidéo. Cinq questions majeures ont été adressées : (1) Comment l'émotion du locuteur est-elle exprimée dans la parole spontanée au niveau prosodique et au niveau lexical ? (2) Quels sont les meilleurs indices acoustiques de la joie et de la tristesse (voix larmoyante sous le coup de stress) ? (3) Dans quelle mesure la perception des émotions primaires (joie et tristesse) est-elle universelle parmi les auditeurs provenant de différentes cultures (comparaison entre les jugements émis par des Coréens, des Américains et des Français) ? (4) Comment le contour intonatif et la durée des stimuli vocaux influencent-t-ils la perception de l'émotion positive (comme la joie) ou négative (comme la tristesse) ? (5) L'émotion est-elle communiquée dans un énoncé de façon uniforme ou bien de façon différente (par exemple, en fonction des parties initiale, médiane et finale de l'énoncé) ?

La dissertation se déroule en sept chapitres. Le chapitre I présente l'objectif et le plan de ce travail en tant qu'introduction générale. Les problématiques générales de l'étude de l'émotion sont aussi discutées de manière brève.

Le chapitre II présente une revue des études précédentes sur la communication de l'émotion à travers la voix dans les domaines sémiotique, philosophique, psychologique, linguistique, et de l'ingénieur. Cette revue comprend l'histoire de l'étude de l'émotion (depuis la philosophie ancienne grecque jusqu'à la psychologie moderne), les théories de la communication (Bühler et Shannon), les théories de l'émotion (Darwin, Tomkins et Plutchik), les modèles de la communication de l'émotion vocale (Fónagy, Léon et Scherer) et les résultats de diverses expériences concernant les indices acoustiques et perceptifs de l'émotion vocale (Lieberman & Michael, Williams & Stevens, Murray & Arnott, etc.). Parmi les modèles de la communication de l'émotion vocale, nous présentons en particulier le modèle de la covariation et le modèle de la configuration que nous allons utiliser pour expliquer nos résultats d'expériences dans les chapitres suivants. Sur le plan conceptuel, nous nous sommes attardée sur les définitions des divers termes émotionnels comme 'affect,' 'humeur ('*mood*'),' 'émotion,' 'sentiment,' 'passion' et 'attitude,' et puis avons explicité la définition du terme "EMOTION" dans notre travail.

Le chapitre III présente la méthodologie de ce travail, particulièrement le choix de la nature des données d'expressions émotionnelles. Nous expliquons ici les raisons pour lesquelles nous avons choisi les données d'expressions émotionnelles acquises à partir de la parole spontanée plutôt qu'à partir du jeu d'un acteur. Dans l'étude de l'émotion, deux types de données peuvent être distingués selon la manière de production : émotions exprimées par un sujet qui se trouve réellement dans un tel état émotionnel (l'*émotion vécue*) vs. émotions exprimées par un acteur qui imite un tel état émotionnel (l'*émotion stylisée*). Nous proposons que cette distinction soit similaire à celle proposée par Fagyal en 1995 entre *parole spontanée* et *parole lue*. Les avantages et les inconvénients de chaque type de données sont estimés selon une échelle de l'authenticité des données et du contrôle expérimental. L'émotion vécue et la parole spontanée sont des données favorables pour capter la richesse de la réalité de l'expérience improvisée du locuteur, mais ces données de nature spontanée ne donnent que des possibilités très limitées au niveau du contrôle expérimental des données. Or, l'émotion stylisée et la parole lue sont des données plus fréquemment utilisées dans les études précédentes à cause de la facilité d'acquisition mais

leur représentativité de la réalité naturelle est plus ou moins problématique. Dans le présent travail, nous avons choisi d'étudier le premier type de données (spontanées) (spontanées) pour examiner les phénomènes de la parole émotionnelle tels qu'ils sont. Une approche typologique fonctionnelle est adoptée dans cette étude pour que les données soient sélectionnées sur des critères fonctionnellement définis et qu'elles soient examinées en fonction des facteurs systématiquement manipulés dans une expérience donnée.

Le chapitre IV présente notre étude principale du corpus coréen. Le corpus coréen est extrait de 40 minutes du discours spontané d'une locutrice Coréenne. Au début de l'entretien télévisé, la locutrice était gaie en racontant de beaux moments avec sa famille et son fiancé, mais elle est devenue triste en parlant de ses conflits familiaux et a fini par pleurer de détresse. 110 énoncés ont été choisis d'une façon chronologique pour nos analyses acoustiques et perceptives. Dans un premier temps, nous avons effectué deux expériences pour décrire l'expression émotionnelle de chaque énoncé en ayant recours au jugement objectif de dix auditeurs coréens. Ils ont évalué l'expression prosodique de l'émotion (*indice vocal*) et l'expression lexicale de l'émotion (*indice sémantique*) des énoncés de notre corpus en termes d'intensité émotionnelle (*valeur d'activation*) et de positivité émotionnelle (*valeur de valence*). Deuxièmement, nous présentons les mesures acoustiques des énoncés : la fréquence du fondamental (Fo) moyen, les Fo maximum et Fo minimum, la plage de Fo, la moyenne des 20 % des valeurs les plus basses de Fo ('*Fo Moy Bas*'), les perturbations de Fo ('*jitter*') et d'intensité ('*shimmer*'), le débit et la distribution spectrale. Notre analyse acoustique montre que l'excitation émotionnelle de la joie crée surtout l'augmentation du Fo moyen tandis que l'excitation émotionnelle de la tristesse (voix larmoyante) peut être mieux repérée par la diminution du Fo minimum et du 'Fo Moy Bas'. L'augmentation du Fo maximum et de la plage de Fo est un bon indice de l'excitation émotionnelle générale (soit la joie, soit la tristesse). Les valeurs de jitter et de shimmer semblent augmenter sous la tension émotionnelle, et le débit de parole semble être accéléré par la joie et ralenti par la tristesse. Cependant, les variations du jitter, du shimmer et du débit dues à l'émotion ne sont pas suffisamment importantes pour atteindre la significativité statistique dans le cas de notre corpus coréen. Troisièmement, l'influence de la culture de l'auditeur sur la perception de l'émotion est examinée par un test de perception avec trois groupes d'auditeurs, Coréens, Américains et Français. Dans le test, ils ont tous identifié la joie et la tristesse de la locutrice Coréenne avec une précision supérieure à celle qui aurait été due au hasard. Cependant, les Coréens étaient

significativement plus précis que les Français et les Américains dans l'identification des émotions coréennes. Ces résultats confirment à la fois l'universalité des indices acoustiques pour la perception de l'émotion et l'influence partielle de la connaissance culturelle sur la perception de l'émotion. Quatrièmement, le rôle des différentes parties (initiale, médiane et finale) de l'énoncé dans la communication de l'émotion est examiné par un autre test de perception. Une étude spectrographique révèle que le bouleversement émotionnel (détresse) de la locutrice souvent s'accompagne de vibrations glottales irrégulières et ralenties dans la partie finale de l'énoncé, tandis que l'émotion positive de la locutrice vis-à-vis de l'interlocuteur s'accompagne souvent d'un cliché mélodique (un contour de Fo *montant-descendant*) sur la dernière syllabe de l'énoncé qui est particulièrement allongée. Dans le test, nous avons divisé des énoncés émotionnels positifs (joie), neutres et négatifs (tristesse) en trois parties, initiale, médiane et finale, et avons demandé aux auditeurs coréens, américains et français d'évaluer l'émotion des stimuli de parties présentés sous forme isolée. Le test a permis de mettre au jour le fait que l'émotion est mieux exprimée et mieux perçue dans la partie finale de l'énoncé que dans les autres parties, initiale et médiane. La découverte d'une répartition non uniforme des indices acoustiques de l'émotion tout au long de *l'énoncé* constitue un aspect original de cette étude. Ce résultat fut mis à l'épreuve et confirmé par d'autres expériences dans les chapitres suivants.

Le chapitre V présente une étude comparative avec un corpus anglais. Cette étude s'adresse spécifiquement à l'expression et la perception de la détresse en fonction des parties initiale, médiane et finale de l'énoncé. Le corpus anglais consiste en des extraits des discours spontanés de cinq locutrices Américaines dans une série d'entretiens télévisés. Chaque enregistrement de l'entretien contient des moments où la locutrice parle calmement (émotion neutre) et des moments où elle est tellement bouleversée qu'elle a fini par pleurer (voix larmoyante). L'analyse acoustique du corpus anglais montre que le Fo moyen augmente et le débit de parole se ralentit quand le sujet est en détresse, ce qui est similaire aux résultats du corpus coréen. La détresse des locutrices Américaines a aussi causé les vibrations glottales irrégulières ralenties vers la fin de l'énoncé. D'après une expérience perceptive, cette modification acoustique dans la partie finale de l'énoncé est responsable d'une meilleure identification de la détresse dans cette partie, par rapport aux parties initiale et médiane de l'énoncé. Le taux d'identification émotionnelle avec la partie finale de l'énoncé seule (présentée en forme isolée) est presque comparable à celui obtenu avec l'énoncé entier.

Le chapitre VI présente deux expériences avec des stimuli synthétiques, qui ont été effectuées en vue de la vérification par synthèse des résultats des chapitres précédents. Dans la première expérience, nous avons examiné l'influence de la modification du contour de Fo et de la durée sur la perception de l'émotion. Quatre contours de Fo (*montant*, *descendant*, *montant-descendant* et *plat*) et deux durées (*longue* et *courte*) ont été synthétisés à partir de la parole naturelle, et leur impression émotionnelle a été évaluée par dix Coréens et dix Américains. Les résultats montrent que les contours intonatifs contenant un élément 'montant' (comme le contour *montant* et le contour *montant-descendant*) produisent un biais perceptif vers une émotion POSITIVE, tandis que les contours 'non-montants' (comme le contour *descendant* et le contour *plat*) produisent un biais perceptif vers une émotion NEGATIVE. Les résultats des Coréens et ceux des Américains sont similaires, sauf que la contribution du contour *plat* à la perception de l'émotion négative est peu significative dans le cas des Américains. Dans la deuxième expérience, nous avons examiné l'influence de la position du trait prosodique dans les différentes parties (initiale, médiane et finale) de l'énoncé sur la perception de l'émotion. Le contour de Fo *montant-descendant*, identifié comme l'indice de l'émotion positive, a été choisi comme le trait de cible et inséré dans la partie initiale, médiane ou finale de l'énoncé au moyen de la resynthèse. L'évaluation des Coréens et des Américains sur l'impression émotionnelle des stimuli resynthétisés montre que l'émotion de l'énoncé est plus fréquemment perçue comme positive quand le contour *montant-descendant* est placé dans la partie finale de l'énoncé que quand il est placé dans la partie initiale ou médiane de l'énoncé. Il est à préciser que l'expérience du chapitre VI concerne la modification volontaire des traits prosodiques qui vise à communiquer le sens émotionnel, alors que l'expérience du chapitre V concerne la modification involontaire des traits prosodiques qui résulte du bouleversement émotionnel du locuteur sans regard pour son intention consciente. Ces deux types de manifestation de l'émotion conforment respectivement au modèle de *configuration* et au modèle de *covariation*.

Dans le chapitre VII, nous résumons les points originaux de ce présent travail, présentons plusieurs applications potentielles de nos résultats dans la technologie vocale et concluons la thèse avec notre perspective du travail futur. L'adéquation de la voix synthétique au contexte peut être améliorée par l'ajout de l'effet émotionnel. Par exemple, l'ajout d'un contour de Fo 'mondant' ou 'montant-descendant' peut créer une impression

positive dans la voix synthétique. En particulier, la synthèse de la voix munie d'émotion et de style personnel fournira aux aveugles un moyen de communication plus esthétique (e.x. la diction automatique des romans avec une variété vocale stylistique et émotionnelle). La modification significative des traits prosodiques par l'effet émotionnel, constatée dans nos résultats acoustiques, confirme le problème de la reconnaissance automatique de la parole et du locuteur lorsque le sujet parle sous le coup d'émotion. La détérioration des résultats de la reconnaissance automatique à cause de l'effet de l'émotion est difficilement améliorable puisque l'état émotionnel du locuteur n'étant pas connu à priori. La reconnaissance de l'émotion du locuteur peut être aussi bénéficiée par la connaisance des indices acoustiques de l'émotion dans la parole naturelle. Nos résultats d'analyses informent que l'état émotif du locuteur peut être repéré par des indices acoustiques, essentiellement localisés dans la fin de l'énoncé. Le repérage de l'état émotionnel du locuteur peut servir dans la communication homme-machine. Actuellement, quelques projets de l'entreprise informatique IBM ('*BlueEyes'* et '*multimodal-Speech Recognition and Animation*,'), la recherche de Cassell *et al.* (1994) sur l'agent automatique de la conversation animée et celle de Picard (1995) sur l'ordinateur émotionnel ('*affective computing*') visent à élaborer l'ordinateur qui peut reconnaître l'état émotionnel du sujet devant l'ordinateur sur la base des expressions vocale, faciale et gestuelle et qui peut lui répondre de façon appropriée au contexte. La reconnaissance de l'état émotionnel du locuteur peut être aussi utile dans le milieu médical. Par exemple, elle pourrait être utilisée pour évaluer automatiquement l'état psychologique des malades sous antidépresseur en tant qu'évaluation de l'efficacité des médicaments antidépresseurs. En conclusion, le point fort et le point faible du présent travail sont notés, et nous discutons de comment améliorer et comment approfondir la recherche de la *communication de l'émotion à travers la voix*.

Summary

Spoken language is a communication device which uses extremly rich and subtle cues. It conveys not only linguistic referential messages but also personal and psychological information of the speaker (ex. social identity, emotional state, etc.). Emotion is a motivational and adaptive response of man to his social environment. It is always present in human speech whereas it currently rarely incorporated in speech synthesis and recognition systems. This lack of emotional aspect of speech in automatic vocal system leads to the mechanic impression of synthetic speech and to the unreliable performance of speech recognition system when speaker is found to be under emotional stress. In order to contribute to the solution of these problems, the present study investigates how emotion modifies the vocal production of the speaker and how the listener perceives the emotion depending on linguistic, cultural, psychological, and acoustical factors.

The dissertation consists of a series of acoustic and perceptual analyses on the emotional speech. The data of emotional expressions were obtained from video-recordings of a series of televison interviews including one Korean and five American female speakers. Five major questions are addressed along the analysis : (1) How is emotion expressed in spontaneous speech on the prosodic level and on the lexical level ? (2) What are the relevant acoustic cues of joy and of sadness (distress expressed through tears) ? (3) How does the listener's cultural background influence the perception of emotion (comparison of jugements by Korean, American and French listeners) ? (4) How do the pitch contour and the duration of the vocal stimuli influence the perception of positive or negative emotion (such as joy or sadness) ? (5) How is emotion communicated in the utterance, whether in an uniform way or in a different way depending on the initial, middle and final parts of the utterance ? These questions are developed through seven chapters.

Chapter I presents the aim and the outline of this study. Some general issues of the study of emotion are briefly discussed.

Chapter II presents a review of previous studies on the vocal communication of emotion in the following fields : philosophy, semiotics, linguistics, psychology and engineering sciences. This review includes a history of studies on emotion (from the Ancient Greek philosophy to the modern psychology), communication theories (Bühler and Shannon), emotion theories (Darwin, Tomkins and Plutchik), vocal communication of emotion models (Fónagy, Léon, Scherer), and the results of a variety of experiments on the acoustic and perceptual cues of emotion (Lieberman & Michael, Williams & Stevens, Murray & Arnott, etc.). The covariation model as welle as the configuration model are also presented among the vocal communication of emotion models, which we will use to explain the results of our experiments in the following chapters. For the sake of conceptual clarification, the definition of the divers emotional terms - 'affect,' 'mood,' 'emotion,' 'feeling,' 'passion' and 'attitude' - are presented, and then the term EMOTION is defined in the frame of the present study.

Chapter III describes the methodology of this study. This methodology is based on the distinction between *live emotion* and *stylized emotion*, depending on the origin of emotional expression data. It is proposed that this kind of distinction is similar to that of *spontaneous speech* and *read speech*. Advantages and disadvantages of each type of data are explained based upon the inverse relationship of data authenticity vs. experimental control. This relationship model was originally suggested by Fagyal (1995) for the case of speech in general and was adopted with modification in this study for the case of emotional speech. The live emotion and the spontaneous speech data are preferable in order to capture the reality of the speaker's behavior and of his/her emotional experience, but this type of data gives only a very limited possibility of experimental control during data acquisition. The stylized emotion data and the read speech data are more controllable in terms of data construction, whereas their representation of natural emotion is more or less problematic. The first type of data (*spontaneous emotional speech*) was chosen in the current study in order to investigate the emotional speech in the way it occurs in our daily life. The functional typological approach was adopted as a theoretical frame : the data are selected by functionally defined criteria and are analyzed according to the factors systematically manipulated in a given experiment.

Chapter IV presents a study of Korean data. The latter consists of 40 minutes of TV interview of a Korean female speaker, in which she talked about her personal relational

problems. At the beginning of the interview, she was joyful telling the best moments with her family and her fiance, then grew sad and distressed by recalling painful moments of these relationships and finally ended up speaking in tears. 110 utterances were selected in a chronological order for our acoustical and perceptual analyses. Firstly, two perception tests were performed in order to assess how emotions are expressed and recognized on the prosodic level (*vocal cue*) and on the semantic level (*lexical cues*) of the utterances. In each test, ten Korean listeners judged the emotional expression of each utterance of the Korean data -randomly presented - in terms of emotional intensity (*activation value*) and of emotional positivity (*valence value*). Secondly, the acoustical properties of Korean data were measured : Fo mean, Fo maximum, Fo minimum, mean of the 20% lowest Fo values (‘*Low Fo Mean*’), Fo range, jitter, shimmer, speaking rate and spectral distribution. The acoustic analysis shows that the emotional excitement of joy particularly increases Fo mean, whereas that of sadness (distress expressed through tears) enhances the decrease of Fo minimum and of Low Fo Mean. The increase of Fo maximum and of Fo range is found to be a good indicator of the general emotional arousal (toward either positive or negative). The jitter and the shimmer values seem to increase under the emotional tension, and the speaking rate seems to increase with joy and decrease with sadness. However, these variations of the jitter, of the shimmer and of the speaking rate were not statistically significant in the case of our Korean data. Thirdly, the influence of cultural background on the perception of emotion was tested with three groups of listeners - Koreans, French, and Americans. The results point out to on the one hand universality of emotion and to some cultural influence on the perceptual precision of emotion on the other. Thus all three groups of listeners recognized the joy and sadness of our Korean speaker with accuracy beyond a chance level. However, the Koreans were significantly more precise than the French and the Americans. Fourthly, another perception test was performed regarding the role of different parts (initial, middle, and final) of the utterance in the communication of emotion. The issue was raised from our spectrographic observation of the Korean data. It shows that the emotional distress of the speaker is often expressed with slower and more irregular glottal vibrations especially in the final part of the utterance rather than at the beginning or at the middle of it. Furthermore, the speaker’s positive emotion toward the listener is often accompanied by a *rising-falling* Fo contour on the final syllable of the utterance which is particularly lengthened. In the perception test, each of the 15 selected utterances was divided into three parts - initial, middle, and final. The emotionality of each part (randomly presented in an isolated form) was evaluated by thirty Korean, American, and French

listeners. The results show the perceptual relevance of the acoustical modifications in the final part of the utterance, clearly indicating that the emotions were better identified in the final part of the utterance than in the beginning or in the middle of it. The *efficiency of the final part of the utterance in the communication of emotion* constitutes an original aspect of this thesis. The following chapters offer broader verification of the result with two new data - English and synthetic speech.

Chapter V presents a comparative study using English data. This study is focused on problems of the expression and perception of distress depending on the initial, middle, and final parts of the utterance. The English data consist of spontaneous discourse of five American female speakers, recorded from five TV interviews. The topics and formats of the interviews were similar to those of the Korean data. Each interview included the moments when the speaker was talking without any particular emotion (neutral emotion) and when she was so upset that she was ending up speaking in tears (sadness/distress). According to the acoustic analysis of 92 English utterances, there is an increase in Fo mean and a decrease in speaking rate when the speaker becomes upset in destresse. The glottal vibrations become slower and more irregular especially at the end of the utterance, which was also observed in our Korean data. A perception test was conducted with 22 American listeners. The result shows a better identification of the distress in the final part of the utterance, compared to the initial and middle parts of it. Furthermore, the correct identification rate of the final part alone is almost comparable to that of the whole utterance. These evidences of English data support again the primacy of the final part of the utterance in the expression and the perception of emotion.

Chapter VI describes two experiments with synthetic stimuli. These experiments were performed in view of the verification by synthesis of our results in the previous chapters. The first experiment examined the influence of Fo contour and of duration on the perception of emotion. Four types of Fo contour (*rising*, *falling*, *rising-falling*, and *flat*) and two types of duration (*long* and *short*) were synthesized from three syllables of natural speech (24 synthetic stimuli in total). The emotional impression (POSITIVE or NEGATIVE) of these stimuli was evaluated by twenty Korean and American listeners. The results show that the Fo contours containing a 'rising' element (such as *rising* and *rising-falling* contours) yield a perceptual bias toward a POSITIVE emotion, whereas the 'non-rising' Fo contours (such as *falling* and *flat* contours) yield a perceptual bias toward a NEGATIVE

emotion. The results of the Koreans and of the Americans are similar, except that the contribution of the *flat* contour to the perception of NEGATIVE emotion is barely significant in the case of American listeners. The second experiment tested whether or not the insertion of a prosodic feature in the different parts (initial, middle, or final) of the utterance influences the perception of emotion. Four *rising-falling* Fo contours, identified as an indicator of positive emotion, were chosen as target features, and each of them were inserted in the initial, middle, or final position of the utterance by means of vocal resynthesis. The emotional impression (POSITIVE or NEGATIVE) of these 12 stimuli was evaluated by ten Koreans and ten Americans. The results show that the emotion of the utterance was more frequently perceived as POSITIVE when the target feature was inserted in the final part of the utterance than in the initial or middle part. The results of this analysis by synthesis support again our argument that the final part of the utterance is more efficient than the initial and the middle parts when it comes to the communication of emotion. It was also noted that the discussion of Chapter VI is about the voluntary modification of the prosodic features which aims to communicate the emotional meaning, whereas the discussion of Chapter V deals with the involuntary modification of the prosodic features which is produced by the emotional upset and is expressed regardless of the speaker's intention.

Chapter VII summarizes the discussions of this dissertation and suggests several potential applications of our results in the field of voice technology. Then it concludes with the prospective of our future study. As for the potential applications, synthetic speech can be more appropriate to the context by varying its emotion and personal styles. For exemple, an addition of 'rising' or of 'rising-falling' contour at the end of utterance can create a positive impression in the synthétic voice. Especially, the synthetic speech wich emotion and personality would provide an esthetic communication tool to the blind (ex. automatic reading of novels with a variety of vocal styles and emotions). The significant prosodic modification by emotional effect, observed in our acoustical results, confirms the problem of the speech recognition system and of the speaker identification system when the speaker is found to be under emotional stress. The deterioration of speech recognition results because of the emotional effect is hard to be improved because the emotional state of speaker is not known à priori. By the way, the knowledge of the acoustical cues of emotions expressed in natural speech can be beneficial to the improvement of the emotion identification system. Our results inform that emotion can be detected by acoustical

measurement of speech signal, essentially localised in the final part of utterance. The identification of emotional state of the speaker is useful in the development of human-computer communication. Currently, several projects of IBM (such as '*BlueEyes Project*' and '*Multimodal-Speech Recognition and Animation Project*'), the research of Casell *et al.* (1994) on the 'Autonomous Animated Conversational Agent' and the project of Picard (1995) on the 'Affective Computing' aim to build a computer which can recognize user's émotional state and respond to him in an appropriate way. The emotion identification system would be also useful in the medical field. For exemple, it can be used to evaluate automatically the psychological state of the depressive patients as an assessment the efficiency of antidepression medication. As conclusion, merits and limits of the present study are noted, and several complementary experiments are suggested for further analysis on *the expression and perception of emotion in the spontaneous speech*.

Bibliographie

Abadjieva, E., Murray, I. R. & Arnott, J. L. (1993), Applying analysis of human emotional speech to enhance synthetic speech, *Proc. Eurospeech '93*, 909-912, Berlin, Allemagne.

Allen, J., Hunnicutt, M. S. & Klatt, D. (1987), *From text to speech: The MITalk system*, Cambridge University Press.

Ahn, S.-C. (1985), *The Interplay of Phonology and Morphonolgy in Korean*, Thèse de doctorat, Université de Illinois (Urbana-Champaign), Etats-Unis, publié par Hanshin Publishing Co. Séoul, Corée du sud.

Arnfield, S., Roach, P, Setter, J., Greasley, P, & Horton, D (1995), Emotional Stress and Speech Tempo Variation, *ESCA - NATO worshop on Speech uncer Stress*, 13-15, Lisbon, Portugal.

Arnold, M. B. (1960-1961), *Emotion and personality : vol.1, Psychological aspects ; vol.2, Neurological aspects*, Colombia University Press, New York.

Bard, P. & Rioch, D. A. (1937), A study of four cats deprived of neocortex and additional portions of the forebrain, *Johns Hopkins Hospital Bulletin* 60, 73-147.

Backrowski, J-A. & Owren, M. J. (1995), Vocal Expression of Emotion : Acoustique Properties of Speech Are Associated With Emotional Intensity and Context, *Psychological Science* 6 (4), 219-224.

Barik, H. C. (1975), Simultaneous interpretation : Qualitative and Lingusitic Data, *Language and Speech* 16, 237-270.

Barik, H. C. (1979), Cross-linguistic study of temporal characteristics of different types of speech materials, *Language and Speech* 20, 116-126.

Bateson, G. (1972), *Vers une écologie de l'esprit*, Seuil (éd.1980), Paris.

Bateson, G. & Ruesch, J. (1951), *Communication et société*, Seuil (éd. 1988), Paris.

Batliner, A., Kompe, R., Kießling, A., Mast, M., Nieman, H., & Nöth, E. (1998), M = Syntax + Prosody: A Syntactic-prosodic labelling scheme for large spontaneous speech databases, *Speech Communication* 25, 193-222.

Batson, C. D., Shaw, L. L., & Oleson, K. C. (1992), Differentiating Affect, Mood, and Emotion: Toward Functionally Based Conceptual Distinctions, Dans M. S. Clark (éd.), *Emotion*, 294-326, Newbury Park, CA, Etats-Unis.

Bechtereva, N. P. (1978), *The neurophysiological aspects of human mental activity*, Oxford University Press.

Beckman, M. & Edwards, J. (1991), The articulatory kinematics of final lengthening, *Journal of the Acoustical Society of America* 89(1), 369-382.

Beverige, W. I. B. (1950), *The art of scientific investigation*, Vintage Books, New York.

Bezooijen, R. A. M. G. van (1984), *The characteristics qnd recognizability of vocal expression of emotion*, Foris, Dordrecht, Pays-Bas.

Blaaw, E. (1994), The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech, *Speech Communication* 14, 359-376.

Blaaw, E. (1995), *On the perceptual classification of spontaneous and read speech*, Thèse de doctorat, Université de Utrecht, Pays-Bas.

Black, J. W. (1961), Relationships among Fundamental frequency, Vocal sound pressure, and Rate of speaking, *Language and Speech* 4, 196-199.

Blanche-Benveniste, C. (1997), *Approches de la langue parlée en français*, Ophrys, Paris.

Bluhme, T. (1971), L'identification de différents attitudes emotionnelles par l'intonation, *Travaux de l'institut de phonétique de Strasbourg* 3, 248-260.

Bonner, M. R. (1943), Changes in the speech pattern under emotional tension, *American Journal of Psychology* 56, 262-273.

Breznitz, Z. (1992), Verbal Indicators of Depression, *The Journal of General Psychology* 119(4), p351-363.

Brown, B. L. (1980), The detection of emotion in voice qualities, Dans H. Giles, W. P. Robinson & P. M. Smith (éd.), *Language : Social psychological perspectives*, 237-245, Pergamom Press, Oxford.

Brown, B. L., Strong, W. J. & Rencher, A. C. (1974), Fifty-four voices from two: the effects of simultaneous manipulation of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech, *J. Acoust. Soc. Am.* 55(2), 313-319.

Buck, R. (1985), Prime Theory: An integrated view of motivation and emotion, *Psychological Review* 92, 389-413.

Buck, R., Baron, R. Goodman, N. & Shapiro, B. (1980), Unitization of Spontaneous Nonverbal Behavior in the Study of Emotion Communication, *Journal of Personality and Social Psycholgoy* 39, 522-529.

Bühler, K. (1934), *Sprachtheorie*, dans *Karl Bühler, semiotic foundations of language theory*, traduction anglaise par Innis, R. E. (1982), Plenum Press, New York.

Cahn, J. E. (1989), *Generating Expression in Synthesized Speech,* Thèse de maîtrise, Massachusetts Institute of Technology, Cambridge, Etats-Unis.

Cahn, J. E. (1990), Generation of Affect in Synthesized Speech, *Journal of the American Voice I/O Society* 8, p1-19.

Cannon, W. B. (1927), The James-Lange theory of emotion – A critical examination and an alternative theory, *American Journal of Psychology* 39, 106-124.

Carlson, R. B., Granström, B. & Nord., L. (1992), Expériments with emotive speech – acted utterances and synthesiwed replicas, dans *ICSLP 92 International Conference on Spoken Language Processing*.

Cassell, J., Stone, M., Douville, B., Prevost, S., Achorn, B., Steedman, M., Badler, N. & Pelachaud, C. (1994), Modeling the Interaction between Speech and Gesture. Dans Ashwin Ram & Kurt Eiselt (éd.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, 153-158, Lawrence Erlbaum Associates, Publishers.

Cho, Y.-M. Y. (1987), Phrasal Phonology in Korean, *Harvard Studies in Korean Linguistics*, II, Harvard University Presse, Cambridge, Massachusetts, Etats-Unis.

Chomsky, N. (1965), *Aspects of Theory fo Syntax*, MIT Press, Cambridge, Massachusetts.

Chung, S.-J. (1994), *Les analyses acoustiques et perceptives de la parole émotionnelle en coréen et en français*, Mémoire de DEA, Université Paris III (Sorvonne Nouvelle), Paris, France.

Chung, S.-J. (1995a), Etudes acoustique et perceptive de la parole émotive en coréen et en Français, *Actes. XIII*èmes *Congré Iinternational des Sciences Phonétiques* 1, 266-269, Stockholm, Swède.

Chung, S.-J. (1995b), Les indices acoustiques et perceptives de l'émotion *Actes. Congré International de la Linguistique: XXII*èmes *anniversaire de la Société Linguistique Coréenne*, 25-27, Séoul, Corée du Sud.

Chung, S.-J. & Kenstowicz, M. (1997), Expression de la focalisation en coréen Séoulien, *Actes. VII*èmes *Symposium International sur la Linguistique Coréenne*, 93-105, Université de Havard, Cambridge, Massachusetts, Etats-Unis.

Chung, S.-J. (1998), Perception de l'émotion: les indices prosodiques et la durée minimale, *Actes des XXIIème Journées d'Etudes sur la Parole*, 159-162, 15-19 juin 1998, Matigny, Suisse.

Chung, S.-J. (1999), Expression vocale et perception de l'émotion en coréen, *Actes des XIV*èmes *Congré Iinternational des Sciences Phonétiques,* 1-8 aôut 1999, San Francisco, CA, Etats-Unis.

Collier, G. (1985), *Emotional Expression*, Lawrence Erlbaum Associates, Publishers, Londre.

Cooper, W. E. & Sorensen, J. M. (1981), *Fundamental Frequency in Sentence Production*, Springer-Verlag, New York, Heidelberg, Berlin.

Couper-Kuhlen, E. (1986), *An Introduction To English Prosody*, Edward Arnold.

Cosmides, L. (1983), Invariances in the Acoustic Expression of Emotion During Speech, *Journal of Experimental Psychology : Human Perception and Performance* 9(6), 864-881.

Costanzo, F. S., Markel, N., N., & Costanzo, P. R. (1969), Voice quality profile and perceived emotion, *Journal of Counseling Psychology* 16, 267-270.

Cowan, M. (1936), Pitch and Intensity Characteristics of Stage Speech, *Archives of Speech - Supplementary to December issue*, 7-92.

Cummings, K. E. & Clements, M. A. (1995), Analysis of the glottal excitation of emotionally styled and stressed speech, *J. Acoust. Soc. Am.* 98(1), 88-98.

Davidson, R. J. (1984), Hemispheric Asymmetry and Emotion, dans R. Scherer & P. Ekman (éd.), *Approches to Emotion*, Lawrence Erlbaum Associates, Publishers, Londre.

Dantzer, R. (1994), *Les émotions*, Que sais-je ?, Presses Universitaires de France.

Darwin, C. (1872), *The expression of emotion in man and animals*, J. Murray, Londre.

Davitz, J. R. (1964), Personality, Perceptual and Cognitive Correlates of Emotional Sensitivity, dans Davitz, J. R. (éd.), *The communication of Emotional Meaning*, 57-68, McGraw-Hill, New York.

Davitz, J. R. & Davitz, L. J. (1974), The communication of Feelings by Context-Free Speech, *Journal of Communication* 9, 6-13.

Dawkins, R. & Krebs, J. R. (1978), Animal signals: Information or manipulation?, dans J. R. Krebs & N. B. Davies (éd.), *Behavioral Ecology*, Blackwell, Oxford.

Dawes, R. M. & Kramer, E. (1966), A Proximity Analysis of Vocally Expressed Emotion, *Perceptual and Motor Skills* 22, 571-574.

DeJong, K. (1989), Initial Tones and Prominence in Seoul Korean, un papier présenté au *177th meeting of the Acoustical Society of America*, Syracuse, N.Y., Etats-Unis, apparu dans *Ohio State University Working Papers in Linguistics* (1994) 43, 1-14.

Delagado, C. (1970), Modulation of emotions by cerebral radio stimulation, dans P. E. Black (éd.), *Physological correlates of emotion*, 189-202.

Delattre, P. (1965), *Comparing the prosodic features of English, German, Spanish, and French*, Julius Groos Verlag, Heidelberg.

Démotrite, *Diogène Laërte*, traduit par Solovin, M. (1927), *Doctrines philosophiques et réflexions morales*, Librairie Félix Alcan, Paris.

Denenberg, V. H. (1981), Hemispheric laterality in animals and the effects of early experience, *The Behavioral and Brain Sciences* 4, 1-19.

Derryberry, D. & Rothbart, M. K. (1988), Emotion, attention, and temperament, dans C. E. Izard, J. Kagan, & R. B. Zajonc (éd.), *Emotion, Cognition, and Behavior*, 132-166, Cambridge University Press, New York.

Di Cristo, A. (1985), *De la micro-prosodie à l'intonosyntaxe*, Publication de l'Université d'Aix-en-Provence, Vol.2, Aix-en-Provence.

Di Cristo, A. & Rossi, M. (1981), Aspects phonétiques et phonologiques des éléments prosodiques, *Linguistiques*, Tome III, Fascicule 2, 24-83, Presses Universitaires de Lille.

Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996), Glottalization of word-initial vowels as a function of prosodic structure, *Journal of Phonetics* 24, 423-444.

Dimond, S., Farrington, L. & Johnson, P. (1976), Differing emotional reponse from right and left hemispheres, *Nature* 261, 690-692.

Doherty, E. T. & Shipp, T. (1988), Tape Recorder Effects on Jitter and Shimmer Extraction, *Journal of Speech and Hearing Research* 31, 485-490.

Doty, R. W. (1989), Some Anatomical Substrates of Emotion, and their Bihemispheric Coordination, *Expérimental Brain Research Series* 18, 56-81.

Duchenne, G. B. abbé, (1862), *Mécanisme de la physionomie humaine ; analyse électrophysiologique de l'expression des passaions*, Baillère, Paris.

Duez, D. (1978), *Essai sur la prosodie du discours politique*, Thèse de doctorat, Université de Paris III, Paris, France.

Dumas, G. (1900), La tristesse et la joie, dans *Nouveau Traité de psychologie*, Dumas G., Vol. II (Book III, éd. 1932), 221-443, Vol. III (book II, éd. 1933), 41-292, Paris.

Dumas, G. (1923), Expression des émotions, dans Dumas G., *Traité de psychologie*, 295-360, Alcan, Paris.

Dumitrache, H. (1994), *Attitudes et Emotions à travers la voix: Analyse de la pièce de Jean Cocteau, La voix humaine*, Thèse de Doctorat, Université de Paris VII, Paris, France.

Dusenberry, D. & Knower, F. H. (1939), Experimental Studies on Symbolisme of Action and Voice II, *Q. J. Speech* 25, 67-75.

Edwards, K. (1997), The face of time: Temporal cues in facial expressions of emotion, *Psychogical Science*.

Ekman, P., Ellsworth, P. & Frieson, W. V. (1972), *Emotion in the human face : Guidelines for research and an intergration of findings*, Pergamon Press, New York.

Ekman, P., Levenson, R. W. & Frieson, W. V. (1983), Autonomic Nervous System Activity Distinguishes Among Emotions, *Science* 221, 1208-1210.

Ekman, P., & O'Sullivan, M. (1991), Facial expression : methods, means and moues, dans R. S. Feldman & B. Rimé (éd.), *Fundamentals of human behavior*, 200-281, Cambridge Unviersity Press, Cambridge, New York.

Eskenazi, M. (1995), Hot topics in Speaking Style Research, dans G. Bloothooft, V. Hazan, D. Huber, & J. Llisterri (éd.), *European Studies in Phonetics and Speech Communication*, OTS Publications, Utrecht, Pays-Bas.

Fagyal, Z. (1995), *Aspects phonostylistiques de la parole médiatisée lue et spontanée: Age, préstige, situation, style et rythme de parole de l'écrivain M. Duras*, Thèse de Doctorat, Université de la Sorbonne Nouvelle, Paris.

Fairbanks, G. & Pronovost, W. (1938), Vocal Pitch During Simulated Emotion, *Science* 88, 382-383.

Fairbanks, G. & Hoaglin, L. W. (1939), An experimental study of the pitch characteristics of the voice during the expression of emotions, *Speech Monograph* 6, 87-104.

Fairbanks, G. & Hoaglin, L. W. (1941), An experimental study of the durational characteristics of the voice during the expression of emotions, *Speech Monograph* 8, 85-90.

Fant, G. Kruckenberg, A. & Nord, L. (1991), Prosodic and segmental speaker variation, *Speech Communication* 10, 521-531, Pays-Bas.

Faure, G. (1962a), L'intonation et l'identification des mots dans la chaîne parlée (exemples empruntés à la langue française), *Proc. 4th Int. Congr. Phon. Sci*., Helsinki (1961), 598-609.

Faure, G. (1962b), *Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise*, Didier, Paris.

Ficher, R. A. (1947), *The Design of Experiments*, fourth edition, Oliver and Boyd, Edinburg, Londre.

Fónagy, I. (1971a), Double coding in speech, *Semiotica* 3, 189-222.

Fónagy, I. (1971b), Synthèse de l'ironie, *Phonetica* 23(1), 42-51.

Fónagy, I. & Bérard, E. (1972), Il est huit heure: Contribution à l'analyse sémantique de la vive voix, *Phonetica* 26, 157-192.

Fónagy, I. (1976a), La vive voix: dynamique et changement, *Journal de Psychologie* 3-4, 273-303.

Fónagy, I. & Fónagy, J. (1976b), Prosodie professionnelle et changements prosodiques, *Le Francais moderne* 44/3, 193-227.

Fónagy, I. & Sap J. (1977), Traits prosodiques distinctifs de certaines attitudes intellectuelles et émotives, *Actes de VIIIèmes Journées d'Etude sur la Parole*, Aix-en-Provence, 25-27 Mai 1977, 238-246.

Fónagy, I., Fónagy, J. & Sap J. (1979), A la recherche de traits pertinents prosodiques du français parisien, *Phonetica* 36, 1-20.

Fónagy, I. (1978), A new method of investigating the perception of prosodic features, *Language and speech* 21, 34-40.

Fónagy, I. (1980), Interprétation des attitudes à partir d'information prosodiques, *Comprendre le language (actes du Colloque)*, 38-41, Didier, Paris.

Fónagy, I. (1981), Emotions, Voice and Music, *Research Aspects on Singing* 33, 51-79, Royal Swedish Academy of Music.

Fónagy, I. (1983a), *La vive voix: Essais de psycho-phonétique*, Bibliothèque scientifique Payot, Paris.

Fónagy, I. (1983b), Clichés melodiques, *Folia linguistica* 17, 153-185.

Fónagy, I. (1986a), Phonetics and Emotion I., *Quaderni di Semantica*, VII, No.1, juin, 1986.

Fónagy, I. (1986b), Les languages de l'émotion, *Quaderni di Semantica*, VII, No.2, décembre, 1986.

Fónagy, I. (1987), Vocal expression of Emotions and Attitudes – Dynamic distinctive features, *Affettività e sistemi semiotici – Le passioni nel discorso*, VS 47/48, 65-85, Bompiani.

Fónagy, I. (1990), The Changes of Vocal Characterology, *Acta Linguistica Hungarica* 40(3-4), 285-313.

Fohr, D. & Laprie, Y. (1989), Snorri : an Interactive Tool for Speech Analysis, *Proceedings of European Conference on Speech*, septembre 1989, Paris.

Frijda, N. H. (1982), The meaning of Emotional Expression, dans M. R. Key (éd.), *Nonverbal Communication Today, Current Research*, 103-119, Mouton Publishers, Berlin, New York, Amsterdam.

Frijda, N. H. (1986), *The emotions*, Cambridge University Press, Cambridge.

Fry, D. B. (1958), Experiments in the perception of stress, *Language and Speech* 1, 126-152.

Fujisaki, H. (1981), Dynamic Characteristics of Voice Fundamental Frequency in Speech and Singing: Acoustical Analysis and Physiological Interpretation, *The Fourth F.A.S.E. Symposium*, Lecture invitée, avril 21-24, Venice, Italie.

Gainotti, G. (1989), Features of Emotional Behavior Relevant to Neurobiology and Theores of Emotions, *Expérimental Brain Research Series* 18, 9-25.

Gardiner, H. M., Metcalf, R. C. & Beebe-Center, J. G. (1937), *Feeling and Emotion : A History of theories*, American Book Company, New York.

Goldman-Eisler, F. (1968), Psycholinguistics : Experiments in Spontaneous Speech, Academic Press, Londre.

Grosjean, F. & Deschamps, A. (1975), Analyse contrastive des variables temporelles de l'anglais et du français : Vitesse de parole et variables composantes, phénomènes d'hésitation, *Phonetica* 31, 144-184.

Guiraud, P. (1953), *Langage et versification dans l'oeuvre de Paul Valéry, Etude sur la forme poétique dans ses rapports avec la langue*, Klincksieck, Paris.

Gumperz, J. J. (1982), *Discourse strategies*, Cambridge University Press.

Hagège, C. (1985), *L'homme de parole*, Fayard, Paris.

Halberstadt, J. B., Niedenthal, P. M., & Kushner, J. (1995), Resolution of Lexical Ambiguity by Emotional State, *Science* 6(5), 278-282.

Hansen, C. H. & Hansen, R. D. (1988), Finding the Face in the Crowd: An Anger Superiority Effect, *Journal of Personality and Social Psychology* 54(6), 917-924.

Hecker, M. H., Stevens, K. N., von Bismarck, G., & Williams, C. E. (1968), Manifestation of Task-Induced Stress in the Acoustic Speech Signal, *J. Acoust. Soc. Amer.* 44(4), 993-1001.

Heiberger, V. L. & Horii, Y. (1982) Jitter and shimmer in sustained phonation, *Speech and Language : Advances in Basic Research and Practice* (éd. Lass, N. J.), Vol. 7, 299-332, Academic, New York.

Hess, U., Kappas, A. & Scherer, K. R. (1988), Multichannel Communication of Emotion : Synthetic Signal Production, dans K. R. Scherer (éd.), *Facets of Emotion : Recent Research*, 161-182, Lawrence Erlbaum Associates, publishers, Hove, Londre.

Hieke, A., Kowal, S. & O'connell, D. (1983), The trouble with articulatory pauses, *Language and Speech* 26, 203-215.

Hiki, S., Sugawara, K., & Oizumi, J. (1968), On the rapid fluctuation of voice pitch, *report of the Reseqrch Institute of Electrical Communication* 19, 237-239. Université de Tohoku, Sendai, Japon.

Hillenbrand, J. (1988), Perception of aperiodicity in synthetically generated voices, *J. Acoust. Soc. Am.* 83(6), 2361-2371.

Homayounpour, M. M., Goldman, J. Ph. & Chollet, G. (1993), Machine vs. Human Speaker Verification, *Actes du conférence IAFP*, Trier, Allemagne.

Horri, Y. (1980), Vocal Shimmer in Sustained Phonations, *Journal of Speech and Hearing Research* 23, 202-209.

Horri, Y. (1982), Jitter and Shimmer Differences among Sustained Vowel Phonations, *Journal of Speech and Hearing Research* 25, 12-14.

Hutter, G. L. (1968), Relations Between Prosodic Variables and Emotions in Normal American English Utterances, *Journal of Speech and Hearing Research* 11, 481-487.

Izard, C. E. (1979), *Emotions in Personality and Psychopathology*, Plenum Press, New York.

Izard, C. E. (1984), Emotion-cognition relationship and human development, dans C. E. Izard, J. E. Kagan & R. B. Zajonc (éd.), *Emotions, Cognition, & Behavior*, 17-37, Cambridge University Press.

Jacobson, E. (1967), *Biology of Emotions: New Understanding Derived from Biological Multidisciplinary Inverstigation; First Electrophysiological Measurements*, Charles D Thomas Publisher, Springfield, Illinois, Etats-Unis.

Jakobson, R. (1963), *Essais de linguistique générale*, traduit par N. Ruwet, Editions de Minuit, Paris.

James, W. (1884), What is emotion, *Mind* 19, 188-202.

James, W. (1894), The physical basis of emotion, *Psychologic Review* 1, 516-529.

Jun, Sun-Ah (1996), *The Phonetics and Phonology of Korean Prosody : Intonational Phonology and Prosodic Structure*, Garland Publishing, Inc., New York.

Jun, Sun-Ah (1998), The Accentual Phrase in the Korean prosodic hierarchy, *Phonology* 15, 189-226, Cambridge University Press.

Jürgens, U. (1979), Vocalization as an emotional indicator - A neuroethological study in the squirrel monkey, *Behaviour* 69, 88-117.

Jürgens, U. (1982), A neuroéthological approach to the classification of vocaliszation in the squirrel monkey, in C. T. Snowdon, C. H. Brown, & Petersen (éd.), *Primate communication*, 50-62, Cambridge University Press, Cambridge, Angleterre.

Kaiser, L. (1962), Communication of affect by single vowels, *Synthese* 14, 300-319.

Kang, H.-S. (1995), Acoustic and Intonational Correlates of Informational Status of Referring Expressions in Standard Korean, *Ohio State University Working Papers in Linguistics* 45, 98-130.

Kang, O.-M. (1992), *Korean Prosodic Phonology*, Thèse de doctorat, Université de Washington, Etats-Unis.

Kang, Y.-S. (1992), Prosodic Structure of Kroean, dans les *actes du 1992 Seoul International Conference on Linguistics*, 199-208.

Kant, E. (1798), *Anthropologie in pragmatischer Hinsicht*, traduction anglaise par Gregor, M. J. (1974), « *Anthropology from a pragmatic point of view* », Martinus Nijhoff, The Hague.

Karcevskij, S. (1931), Sur la phonologie de la phrase, *Travaux du cercle linguistique de Prague* 4, 188-227.

Kellar, E. (1994), *Signalize™ – Analyse du Signal pour la Parole et le Son*, Manuel d'Utilisation, version 3.0, Traduit en français par Brigitte Zellner, InfoSignal™ Inc.

Kim, K.-H. (1978), *The language of Emotion of Americans and Koreans*, Thèse de doctorat, Université Keimyung, Daegu, Corée du sud, publié par Hanshin Publishing Co. Séoul, Corée du sud.

Kim, M.-J. (1992), *A Phonetic Study of the Structure of Fundamental Frequency Contours*, Thèse de doctorat, Université Nationale de Séoul, Séoul, Corée du sud, publié par Hanshin Publishing Co. Séoul, Corée du sud.

Klingholz, F. & Martin, F. (1985), Quantitative spectral evaluation of shimer and jitter, *J. Speech Hear. Res.* 28, 169-174.

Ko, D.-H. (1988), *Declarative Intonation in Korean – An Acoustic Study of Fo declaration,* Thèse de doctorat, Université de Kansas, Kansas, Etats-Unis, publié par Hanshin Publishing Co. Séoul, Corée du sud.

Koike, Y., Takahashi, H., & Calcaterra, T. C. (1977), Acoustic measures for deteting laryngeal pathology, *Acta Otolaryngologica* 84, 105-117.

Koo, H.-S. (1986), *An Experimental Acoustic Study of the Phonetics of Intonation in Standard Korean*, Thèse de doctorat, Université de Texas à Austin, Etats-Unis, publié par Hanshin Publishing Co. Séoul, Corée du sud.

Kowal, S., O'connell, D. C., O'brien, E. A. & Bryant, E. T. (1975), Temporal Aspects of Reading Aloud and Speaking, *American Journal of Psychology* 88, 4, 549-569.

Kowal, S., Bassett, M. R., & O'connell, D. C. (1985), The Spontaneity of Media Interviews, *Journal of Psycholinguistic Research* 14 (1), 1-18.

Kramer, E. (1963a), The judgment of personal characteristics and emotions from nonverbal properties of speech, *Psychol. Bull.* 60, 408-420.

Kramer, E. (1963b), Judgment of portrayed emotion from normal English, filtered English, and Japanese speech, *Diss. Abstr.* 24, 1699-1700.

Kramer, E. (1964), Elimination of verbal cues in judgements of emotion from voice, *J. Abnorm. Soc. Psychol.* 68, 390-396.

Laan, G. P.M. (1997), The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style, *Speech Communication* 22, 43-65.

Labov, W. (1972), *Langugage in the inner city : Studies in the Black English vernacular*, University of Pennsylvania Press, Philadelpia.

Ladd, D. R. (1978), *The structure of intonational meaning : Evidence from English*, Indiana University Press, Bloomington, Londre.

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985), Evidence for the independent function of intonation contour type, voice quality, and Fo range in signaling speaker affect, *J. Acoust. Soc. Am.* 78, 435-444.

Landercy, A. & Renard, R. (1977), *Eléments de phonétique*, Didier, Bruxelles.

Laprie,Y. & Fohr, D. (1989), Snorri: an Interactive tool for Speech Analysis, *Proceedings of European Conference on Speech Technology*.

Laprie,Y. & Mercier, L. (1994), Un environnement logiciel pour un atelier phonétique, *Actes des 20ièmes Journnées d'Etudes sur la Parole*, 209-214.

Laukkanen, A. M., Vilkman, E., Alku, P., & Oksanen, H. (1996), Physical variations related to stress and emotional state: A preliminary study, *Journal of Phonetics* 24, 313-335.

Laver, J. (1980), *The phonetic description of voice quality*, Cambridge University Press.

Laver, J. (1994), *Principles of phonetics*, Cambridge University Press.

Lazarus, R. S. (1966), *Psychological stress and the coping process*, McGraw-Hill, New York.

Lazarus, R. S. (1982), Thoughts on the relations between emotion and cognition, *American Psychologist* 37, 1019-1024.

Lazarus, R. S. (1984), On the Primacy of Cognition, *American Psychologist* 39, 124-129.

Laziczius, G. (1966), *Selected Writings,* The Hague, Mouton.

LeDoux, J. E. (1989), Cognitive-emotional interactions in the brain, *Cognition and Emotion* 3, 267-289.

Lee, H.-B. (1964), *A Study of Korean (Seoul) Inotnation*, Thèse de maîtrise, Université de Londre, Angleterre.

Lee, H.-B. (1976), Intonation in Korean, *Language Research* (Seoul National University) 1, 131-143.

Lee, H.-Y. (1990), *The Structure of Korean Prosody*, Thèse de doctorat, Université de Londre, Angleterre, publié par Hanshin Publishing Co. Séoul, Corée du sud.

Lee, S.-H. (1989), Intonational Domains of the Seoul Dialect of Korean, un papier présenté au *177th meeting of the Acoustical Society of America*, Syracuse, N.Y., Etats-Unis.

Legros, C. (1995), Discussion on Emotion, *ESCA-NATO Workshop on Speech Under Stress*, 14-15 Septembre, Lisbon, Portugal, p99.

Lehiste, I. (1970), *Suprasegmental,* M.I.T. Press, Cambridge, Massachusetts, Etats-Unis.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M.-L. (1997), Expression of emotional-motivational connotations with a one-word utterance, *J. Acoust. Soc. Am.*102(3), 1853-1863.

Léon, P. R. (1970), Systématique des fonctions expressives de l'intonation, *Studia Phonetica* 3, 57-74.

Léon, P. R. (1971), *Essais de Phonostylistique*, Didier, Paris, Montréal, Bruxelles.

Léon, P. R. (1979), Modèle et fonctions pour l'analyse de l'énonciation, « Le document sonore authentique », numéro spécial du *Français dans el monde* (P. Léon, dir.) 145, 54-69.

Léon, P. R. (1993), *Précis de Phonostylistique: Parole et expressivite*, Nathan, Paris.

Léon, P. R. & Tennant, J. (1990), Bad French and Nice Guys : a Morphophonemic Study, *French Review* 63, 763-778.

Leventhal, H. (1980) Toward a comprehensive theory of emotion, dans L. Berkowitz (éd.), *Advances in experimental social psychology*, Vol.13, 139-207. Academic Press, New York.

Liberman, M. (1975), *The Intonational System of English*, Thèse de doctorat, MIT, Cambridge, Massachusetts, Etats-Unis.

Lieberman, P. (1961), Perturbations in vocal pitch, *J. Acoust. Soc. Am*. 33, 597-603.

Lieberman, P (1967), *Intonation, perception and Language,* Research Monograph No.38, Massachusetts Institute of Technology, Cambridge, Massachusetts, Etats-Unis.

Lieberman, P. (1975), *The intonational system of English*, Thèse de doctorat, Massachusetts Institute of Technology, publié par Garland Press (1979), New York.

Lieberman, P. (1997), Peak Capacity, *The Sciences*, 22-27.

Lieberman, P. & Michael, S. B. (1962), Some Aspects of Fundamental Frequency and Envelope Amplitude as Related to the Emotional Content of Speech, *J. Acoust. Soc. Am.* 34 (7), 922-927.

Lieberman, P. & Blumstein, S. (1988), *Speech physiology, speech perception, and acoustic phonetics*, Cambridge University Press, Cambridge.

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R. & Miller, M., (1985), Measures of the sentence intonation of read and spontaneous speech in American English, *J. Acoust. Soc. Amer.* 77 (2), 649-657.

Liénard, J.-S. (1977), *Les processus de la communication parlée : Introduction à l'analyse et à la synthèse de la parole*, Masson, Paris.

Louis, C. W. (1995), Hot Topics in the Field of Speech synthesis Assessment, in *European Studies in Phonetics and Speech Communication*, edited by Bloothooft *et al*. (1995), 123-126, Utrecht, Pays-Bas.

MacLean, P. D. (1949), Psychosomatic disease and the 'visceral brain', *Psychosomatic Medicine* 11, 338-353.

MacLean, P. D. (1970), The limbic brain in relation to the psychoses, dans l'édition de Black, P. E. (1970), *Physological correlates of emotion,* 129-146.

Martin, P. (1982), Phonetic Realizations of Prosodic Contours in French, *Speech Communciation* 1, 283-294, North-Holland Publishing Company.

Martin, P. (1995), *Winpitch*, Tutorial of the Program, Montréal.

Martinet, E. (1960), *Eléments de linguistique générale*, 1er éd., Coline, Paris.

McAndrew, F. T. (1986), A Cross-Cultural Study of Recognition Thresholds for Facial Expressions of Emotion, *Journal of Cross-Cultural Psychology* 17 (2), 211-224.

McCluskey, K. W., Albas, D. C., Niemi, R. R., Cuevas, C., & Ferrer, C. A. (1975), Corss-cultural differences in the perception of the emotional content of speech: a study of the development of sensitivity in Canadian and Mexican children, *Developmental Psychology* 11, 551-555.

McDougall, W. (1908), *Introduction to social psychology*, Methuen, Londre.

Menn, L. & Boyce, S. (1982), Fundamental Frequency and Discourse Structure, *Language and Speech* 25(4), 341-383.

Miermont, J. (1991), Théories de la communication, Editions Techniques - *Ensyclopédie Médico-Chirurgicale*, Psychiatrie, 37010 A[10], Paris.

Miller, G. A. (1973), *Communication, language and learning – Psychological perspectives*, Basic Books Publishers, éd., New York.

Moon, S.-J. & Lindblom, B. (1994), Interaction between duration, context, and speaking style in English stressed vowels, *Journal of the Acoustical Society of America* 96, 40-55.

Morel, M.-A. & Danon-Boileau, L. (1995), Valeur énonciative des variations de hauteur mélodique en français, *French Language Studies* 5, 189-202.

Morris, J. S., Firth, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J. & Dolan, R. J. (1996), A differential neural response in the human amygdala to fearful and happy facial expressions, *Nature* 383, 812-815.

Mozziconacci, S. J. L. (1998), *Speech Variability and Emotion: Production and Perception*, Thèse de doctorat, Université de Eindhoven, Pays-Bas.

Mozziconacci, S. J. L. & Hermes, D. J. (1997), A study of intonation patterns in speech expressing emotion or attitude: Production and perception, *IPO Annual Progress Report* 32, 154-160.

Mozziconacci, S. J. L. & Hermes, D. J. (1998), Pertinence perceptive des configurations intonatives en parole émotionnelle, *Actes des XXIIème Journées d'Etudes sur la Parole*, 163-166, Matigny, Suisse.

Murray, I. R. (1989), *Simulating emotion in synthetic speech*, Thèse de doctorat, Université de Dundee, Angleterre.

Murray, I. R. & Arnott, J. L. (1993), Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion, *J. Acoust. Soc. Am.* 93(2), 1097-1108.

Murray, I. R., Arnott, J. L., & Rohwer, E. A. (1995), The Strains of Emotional Stress in Synthetic Speech, *Proc. of ESCA-NATO Workshop on Speech under Stress*, Lisbon, Portugal, 71-74.

Murray, I. R. & Arnott, J. L. (1995), Implimentation and testing of a systems for productin emotion-by-rule in synthetic speech, *Speech Communication* 16, 369-390.

Murtus, J. (1985), *PITCH; Pitch Extraitor in Bliss*, version 4.07, Rapport dans le département des sciences cognitive et linguistique, Université de Brown, Providence, RI.

Murtus, J. (1992), *Mev; Speech Edition and Display*, version 1.08, Rapport dans le département des sciences cognitive et linguistique, Université de Brown, Providence, RI.

Mertus, J. (1995), *On-Line Subject Testing : Creating and Editing Experiments*, Rapport dans le département des sciences cognitive et linguistique, l'université de Brown, Providence, RI.

Nakajima, S. & Allen, J. F. (1993), A Study of Prosody and Discourse Structure in Cooperative Dialogue, *Phonetica* 50, 197-210.

Nash, H. (1974), Perception of Vocal Expression of Emotion by Hospital Staff and Patients, *Genetic Psychology Monographs* 89, 25-87.

Norris, M. P., Lynn Snow-Turek, L. B. A., & Finch, J. (1995), Influence of Depression On Verbal Fluency Performance, *Aging and Cognition* 2(3), 206-215.

O'Shaughnessy, D. (1976), *Modelling Fundamental Frequency, and its Relationship to Syntax, Semantics, and Phonetics*, Thèse de doctorat, Massachusetts Institute of Techonology, Cambridge, Massachusetts, Etats-Unis.

O'Shaughnessy, D. & Allen, J. (1983), Linguistic Modality Effects on Fundamental Frequency in Speech, *J. Acoust. Soc. Am.* 74 (4), 1155-1171.

Ohala, J. J. (1983), Cross-Language Use of Pitch: An Ethological View, *Phonetica* 40, 1-18.

Ohala, J. J. (1984), An Ethological Perspective on Common Cross-Language Utilization of Fo of Voice, *Phonetica* 41, 1-16.

Ortony, A. & Turner, T, J. (1990), What's basic about basic emotions?, *Psychological Review* 97 (3), 315-331.

Osgood, C. E., Suci, G. J. & Tannenbaum, P. H. (1957), *The measurement of meaning,* University of Illinois Press.

Pakosz, M. (1982), Intonation and Attitude, *Lingua* 56, 153-178.

Panksepp, J. (1986), The anatomy of Emotion, dans R. Plutchik & H. Kellerman (éd.), *Emotion: Theory, Research, and Experience, Vol. 3, Biological Foundation of Emotion*, Academic Press, Inc., New York, Londre.

Papez, J. W. (1937), A proposed mechanism of emotion, *Archives of Neurology and Psychiatry* 38, 725-743.

Péter, M. (1997), Affectivité, Expressivité, Valeurs stylistiques, dans J. Perrot (éd.), *Polyphonie – pour Iván Fónagy*, Mélanges offerts en hommange à Iván Fónagy par un groupe de disciples, collègues et admirateurs.

Picard, R. W. (1995), *Affective Computing*, M.I.T. Media Laboratory Perceptual Computing Section Technical Report, No. 321.

Pierrehumbert, J. (1980), *The Phonology and Phonetics of English Intonation*, Thèse de doctorat, MIT, Cambridge, Massachusetts, Etats-Unis.

Pierrehumbert, J. (1981), Synthesizing Intonation, *J. Acoust. Soc. Am.* 70(4), 985-995.

Pierrehumbert, J. & Hirschberg, J. (1989), The Meaning of Intonational Contours in the Interpretation of Discourse, dans Cohen, Pollack & Morgan (éd.), *Intentions in Communication*, 271-311, MIT Press, Cambridge, Massachusetts, Etats-Unis.

Pittam, J. & Scherer, K. R. (1993), Vocal Expression and Communication of Emotion, Dans M. Lewis & J. M. Haviland (éd.) *Handbook of Emotions*, Guilford Press, New York.

Ploog, D. (1986), Biological Foundations of the Vocal Expressions of Emotions, in *Emotion: Theory, Research, and Experience* 3, 173-197, Academic Press Inc.

Plutchik, R. (1962), *The Emotions: Facts, Theories, and a New Model*, Random House, New York. 111-113.

Plutchik, R. (1980a), *Emotion: A psychoevolutionary synthesis,* Harper and Row, New York.

Plutchik, R. (1980b), A General Psychoevolutionary Theory of Emotion, dans R. Plutchik & H. Kellerman (éd.), *Emotion: Theory, Research, and Experience,* Vol.1 *Theories of Emotion*, 3-33, Academic Press, New York.

Plutchik, R. (1984), Emotions : A General Psychoevolutionary Theory, dans K. R. Scherer & P. Ekman (éd.), *Approches to Emotion*, 197-219, Lawrence Erlbaum Associates Publishers, Londre.

Plutchik, R. & Kellerman, H. (1990), *Emotion, Psychopathology, and Psychotherapy*, dans *Emotion: Theory, Research, and Experience*, Vol. 5, Academic Press, Inc.

Plutchik, R. & Conte, H. R. (1997), *Circumplex Models of Personality and Emotion*, American Psychological Association.

Pratto, F. & John, O. P. (1991), Automatic Vigilence: The Attention-Grabbing Power of Negative Social Information, *Journal of Personality and Social Psychology* 61(3), 380-391.

Pollack, I., Rubenstein, H. Horowitz, A. (1960), Communication of Verbal Modes of Expression, *Language and Speech* 3, 121-130.

Prevost, S. & Steedman, M. (1994), Specifying intonation from context for speech synthesis, *Speech Communication* 15, 139-153.

Protopapas, A & Eimas, P. D. (1997), Perceptual Differences in infant cries revealed by modifications of acoustic features, *J. Acoust. Soc. Am.* 102(6), 3723-3734.

Protopapas, A & Lieberman, P. (1995), Effects of vocal Fo manipulations on perceived emotional stress, *Proceedings of the ESCA/NATO Tutorial and Research Workshop on Speech Under Stress*, Lisbon, Portugal.

Rolls, E. T. (1986), Neural systems involved in emotion in primates, dans R. Plutchik & H. Kellerman (éd.), *Emotion: Theory, Research, and Experience: Vol. 3 – Biological Foundations of Emotion*, 125-143, Academic Press, Inc., Londre.

Rigault, A. (1964), Rélexions sur le statut phonologique de l'intonation, *Proc. 9th Int. Congr. Ling.,* Cambridge, Mass (1962), 849-858.

Roessler, P., & Lester, J. W. (1976), Voice predicts affect during psychotherapy, *Journal of Nervous and Mental Disease* 163, 166-176.

Rogers, P. L., Scherer, K. R., & Rosenthal, R. (1971), Content-filtering human speech, *Behav. Res. Methods Instrumentation* 3, 16-18.

Rolls, E. T. (1986), Neural systems involved in emotion in primates, dans R. Plutchik & H. Kellerman (éd.), *Emotion: Theory, Research, and Experience: Vol. 3 – Biological Foundations of Emotion*, 125-143, Academic Press, Inc., Londre.

Rossi, M. (1981a), De la physiologie à la perception phonémique, *Linguistiques*, Tome III, Fascicule 2, 5-23, Presses Universitaires de Lille.

Rossi, M. (1981b), A model for predicting the prosody of spontaneous speech (PPSS model), *Speech Communication* 13, 87-107.

Rossi, M. (1988), Prosodie et technologies vocales, *1ères Journées Nationales au GRECO-PRC, Parole, Language, Matérial et Vision*, 63-80.

Sackeim, H. A., Weinman, A. L., Gur, R. C., Greenberg, M., Hungerbuhler, J. P., & Geschwind, N. (1982), Pathological laughing and crying : Functional brain asymmetry in the experience of positive and negative emotions, *Archives of Neurology* 39, 210-218.

Saussure, F. de. (1916), *Cours de linguistique générale*, Payot (5e éd. 1955), Paris.

Scherer, K. R. (1971), Randomized-splicing: A note on a sample technique for masking speech content, *J. Exptl. Res. Personality* 5, 155-159.

Scherer, K. R. (1974), Acoustic Concomitants of Emotional Dimensions: Judging Affect From Synthesized Tone Sequences, dans S. Weitz (éd.), *Nonverbal Communication*, 105-111, Oxford University Press, New York.

Scherer, K. R. (1979), Nonlinguistic Vocal Indicators of Emotion and Psychopathology, dans C. E. Izard (éd.), *Emotions in Personality and Psychopathology*, 495-529, Plenum Press, New York, Londre.

Scherer, K. R. (1982), Methods of research on vocal communication : paradigms and parameters, dans K. R. Scherer & P. Ekman (éd.), *Handbook of method in Nonverbal behavior Research*, 136-198, Cambridge University Press, Cambridge.

Scherer, K. R. (1984) On the nature and function of emotion: A component process approach, dans K. R. Scherer & P. Ekman (éd.), *Approaches to emotion*, 193-318, Lawrence Erlbaum Associates, Publishers, Londre.

Scherer, K. R. (1986), Vocal Affect Expression: A Review and a Model for Future Research, *Psychological Bulletin* 99(2), 143-165.

Scherer, K. R. & Oshinsky (1977), Cue Utilization in Emotion Attribution from Auditory Stimuli, *Motivation and Emotion* 1(4), 331-346.

Scherer, K. R., Wallbot, H. G., Matsumoto, D. (1988), Emotional experience in cultural context : A comparison between Europe, Japan, and The United States, dans Scherer, K. R. (éd.), *Facets of Emotion – Recent Research*, Lawrence Erlbaum Associates Publishers, Londre.

Scherer, K. R., Koivumaki, J., & Rosenthal, R. (1972), Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech, *Journal of Psycholinguistic Research* 1, 269-285.

Scherer, U., Helfrich, H., & Scherer, K. R. (1980), Internal push or external pull?, Determinants of paralinguistic behavior, dans H. Giles, P. Robinson, & P. Smith (éd.), *Language: Social psychological perspectives*, 279-282, Pergamon Press, Oxford, Angleterre.

Scherer, K. R., Summerfield, A. B., & Wallbott, H. (1983), Cross-national research on antecedents and components of emotion: A progress report, *Social Science Information* 22 (3), 355-385.

Scherer, K. R. & Ekman, P. (1984), *Approaches to emotion*, Lawrence Erlbaum Associates, Publishers, Londre.

Scherer, K. R., Ladd, D. R., & Silverman Kim E. A. (1984c), Vocal cues to speaker affect: Testing two models, *J. Acoust. Soc. Am.* 76(5), 1346-1356.

Scherer, K. R. & Kappas, A. (1988), Primate Vocal Epxression of Affective State, dans D. Todt, P. Goedeking, & D. Symmes (éd.), *Primate Vocal Communication*, 171-194, Springer-Verlag, Berlin.

Scherer, K. R., Johnstone, T., & Sangsue, J. (1998), L'état émotionnel du locuteur: facteur négligé mais non négligeable pour la technologie de la parole, *Actes des XXIIème Journées d'Etudes sur la Parole*, 249-257, Matigny, Suisse.

Scherer, R. C., Vail, V. J., & Guo, C. G. (1995), Required Number of Tokens to Determine Representative Voice Perturbation Values, *Journal of Speech and Hearing Research* 38, 1260-1269.

Schlosberg, H. (1954), Three dimensions of emotion, *Psychological Review* 61(2), 81-88

Schoentgen, J. & de Guchtenneere, R. (1995), Time series analysis of jitter, *Journal of Phonetics* 23, 189-201.

Shiffrin, R. M. & Schneider, W. (1977), Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory, *Psychological Review* 84(2), 127-189.

Selkirk, E. O. (1984), *Phonology and Syntax*, MIT Press, Cambridge, Massachusetts.

Skinner, E. R. (1935), A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness, *Speech Monograph* 2, p81-137.

Smith, G. A. (1977), Voice analysis for the measurement of anxiety, *British Journal of Medical Psychology* 50, 367-373.

Starkweather, J. (1961), Vocal Communication of Personality and Human Feeling, *The Journal of Communication*, 63-71.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983), Acoustic and Perceptual Indicators of Emotional Stress, *J. Acoust. Soc. Am.* 73(4), 1354-1360.

Swerts, M. & Geluykens, R. (1993), The Prosody of Information Units in Spontaneous Monologue, *Phonetica* 50, 189-196.

't Hart, J., Collier, R. & Cohen, A. (1990), *A perceptual study of intonation : An experimental-phonetic approach to speech melody*, Cambridge Unviersity Press.

Takahashi, H. & Koike, Y. (1975), Some perceptual dimensions and acoustical correlates of pathologic voices, *Acta Otolaryngologica, suppl.*, 338.

Tartter, V. C. (1980), Happy talk : Perceptual and acoustic effects of smiling on speech, *Perception and Psychophysics* 27, 24-27.

Thomas, R. & Alaphilippe, D. (1983), *Les attitudes*, Que sais-je ?, Presses Universitaires de France.

Titze, I. R., Horii, Y,. & Scherer, R. (1987), Some technical consideration in voice perturbation Measurements, *Journal of Speech and Hearing Research* 30, 252-260.

Tomkins, S. S. (1962), *Affect, Imagery, Consicousness : The Positive Affects ; Vol.1,* Springer Publishing Company Inc., New York.

Tomkins, S. S. (1963), *Affect, Imagery, Consicousness : The Negative Affects ; Vol.2,* Springer Publishing Company Inc., New York.

Tomkins, S. S. (1980), Affect as Amplification : Some Modifications in Theory, dans R. Plutchik & H. Kellerman (éd.), *Emotion : Theory, Research, and Experience*, Vol.1., 141-164, Academic Press, New York.

Tomkins, S. S. (1982), *Affect, Imagery, Consicousness : Vol.3, Cognition and Affect,* Springer Publishing Company Inc., New York.

Tomkins, S. S. (1984), Affect theory, dans K. R. Scherer (éd.), *Approches to Emotion*, 163-195, Lawrence Erlbaum Associates Publishers, Londre.

Trojan, F. (1948), *Der Ausdruck der Stimme und Sprache*, Wien.

Troubetzkoy, N. S. (1939), *Principes de phonologie*, traduction française par Cantineau (1957), Klincksieck, Paris.

Uldall, E. (1960), Attitudinal meanings conveyed by intonation contours, *Language and Speech* 3, 223-234.

Uldall, E. (1964), Dimensions of Meaning in Intonation, dans D. Boulinger (éd. 1972), *Intonation*, 250-259, Penguin Books, Baltimore, Etats-Unis.

Vaissière, J. (1983), Language-Independent Prosodic Features, dans Sprigner-Verlag (éd.), *Prosody: Models and Measurements*, 53-66, Berlin, Heidelberg, New York.

Vaissière, J. (1989), *Contribution à l'analyse des phénomènes de parole continue lue*, Texte préparé en vue de l'obtention du diplôme d'Habilitation à Diriger des Recherches, Strasbourg.

Vaissière, J. (1995), Phonetic Explanation for Cross-Linguistic Prosodic Similarities, *Phonetica* 52, 123-130.

Vaissière, J. (1997), Iván Fónagy et la notation prosodique, dans J. Perrot (éd.), *Polyphonie – pour Iván Fónagy*, 479-488, L'Harmattant, Paris.

Vaissière, J. (1998), Utilisation de la prosodie dans les systèmes automatiques: un problème d'intégration des différentes composantes, *Fait de la langue*, N. 13.

Van Bezooijen, R. (1984), *Characteristics and Recognizability of Vocal Expressions of Emotion*, Foris, Dordrecht, Pays-Bas.

Wallbot, H. G. & Scherer, K. R. (1988), How universal and specific is emotional epxression ? : Evidence from 27 countries on five continents, dans Scherer, K. R. (éd.), *Facets of Emotion – Recent Research*, 31-56, Lawrence Erlbaum Associates, Publishers, Londre.

Waters, J., Nunn, S., Gillcrist, B., & VonColln, E. (1995), The Effect of Stress on the Glottal Pulse, *Proc. of ESCA-Nato Workshop*, 9-11, Lisbon, Portugal.

Weaver, W. & Shannon, C. E. (1949), *Théorie mathématique de la communication*, CEPL Retz (éd.), Paris, 1975.

Wetzel, M. (1989), *Les passions*, Editions Quinquette, Paris.

Williams, C. E. & Stevens, K. N. (1969). On determining the Emotional State of Pilots During Flight: An Exploratory Study, *J. Aerospace Med*. 40, 1369-1372.

Williams, C. E. & Stevens, K. N. (1972). Emotions and Speech: Some Acoustical Correlates, *J. Acoust. Soc. Am.* 52, 1238-1250.

Williams, C. E. & Stevens, K. N. (1981). Vocal correlates of emotional states, dans J. K. Darby (éd.), *Speech Evaluation in Psychiatry*, 221-240, Grune-Stratton, New York.

Young, P. T. (1943), *Emotion in man and animal : Its nature and relation to attitude and motive*, John Wiley, New York.

Zajonc, R. B. (1980), Feeling and thinking: Preferences need no inferences, *American Psychologist* 35, 151-175.

Zajonc, R. B. & Markus, H. (1982) Affective qnd cognitive factors in preference, *Journal of Consumer Research* 9, 123-131

Zajonc, R. B. (1984), On the Primacy of Affect, *American Psychologist* 39, 117-123.

« Dictionnaire de linguistique », édité par J. Dubois, M. Giacomo, L. Guespin, C. Marcellesi, J.-B. Marcellesi, J.-P. Mevel (1979), Librairie Larousse, Paris.

« *Le Petit Robert : Dictionnaire de la langue française* » de Paul Robert, (1982), Le Robert, Paris.

« *Larousse : Précis de grammaire* », édition refondue (1979), Librairie Larousse, Paris.

« *Le Petit Robert : Dictionnaire de la langue française* » de Paul Robert, (1982), Le Robert, Paris.

« *Le Robert : Dictionnaire de la langue française* » de Paul Robert, revue et enrichie par Rey, A., deuxième édition (1985), Le Robert, Paris.

Annexes

Corpus Coréen

Transcription phonétique

Fichier	Emotion	Durée	Fo Moy.	Enoncés
WJ 11	positive	1246	241	ai Ze suNg'jcgijo
WJ 12	positive	2137	217	nuga gûrcke gabZ'agi ape dagaomjcn
WJ 13	positive	2558	244	jag'an dwiro ircke ZutSumhago mulcscnûn scNg'jcgigcdûnjo
WJ 14	positive	1591	257	gûruke mals'ûmûl hasigillA
WJ 15	positive	1667	278	scnbAnimhago naitSaiga manag'cdûnjo
WJ 16	positive	1715	243	tScûme uri deitûhalA ircgilA
WJ 17	positive	1400	264	c andwejohago marûlhAt'aga
WJ 18	positive	2152	219	a nAga ige gwaminbanûNi aninga hAscnûn
WJ 19	positive	1526	318	gûrcmjcnûn sigs'anûn gwentSanajo
WJ 21	positive	1250	253	je saNugi ob'anûnjo
WJ 22	positive	1206	222	maûmi dwege nclg'u
WJ 23	positive	1568	223	bepulZ'ul anûn saramijejo
WJ 24	positive	2921	250	d'ag wemorûl bwadu s'antakûrosû dalmZ'I anas'cjo
WJ 25	positive	2326	278	cd'cn Zagûmahan gcsûl soZuNhage anûn gûge
WJ 26	positive	2727	237	cd'cn kcdaran gcmman ZuNjohagu mwc ircke gûrcnûnge anigu
WJ 27	positive	2345	252	Zagûman saraNdo ircke soZuNhi halZ'ul anûn
WJ 28	positive	1853	246	gûrcn saramicsc dwege Zoas'cjo
WJ 31	neutre	1289	220	je ob'aga inZe g'umi
WJ 32	neutre	1768	218	munhwagoNganûl mandûnûn gcnig'an
WJ 33	neutre	2626	213	gûrcngcl wihAscn cd'cn Zagûmi piljohaZanajo
WJ 34a	neutre	1699	209	gûrAsc ircke ZcgûmtoNe obAwcnZ'ari
WJ 34b	neutre	1376	216	duls'ig moimjcnsc gûrcn
WJ 34c	neutre	1874	260	g'umûlirwcgals'uit'ago sANgakeg'cdûnjo
WJ 35	neutre	1243	248	g'umûl irwcganûnde
WJ 36	neutre	2791	223	doumi dwes'ûmjcnhanûn maûmûro ZcgûmtoNûl scnmulhAs'ûmnida
WJ 37	neutre	1842	218	je obAgwcnZ'arihana nccscjo
WJ 41	neutre	1018	271	Zcûmejo
WJ 42a	neutre	780	266	ob'aga tScûm
WJ 42b	neutre	1013	224	cmmarûl bwengejo
WJ 43	neutre	1550	233	Zcnsihwe gasc bweg'cdûnjo
WJ 44a	neutre	1742	214	gweNZaNi iri bab'ûnd'Ajcs'jo
WJ 44b	neutre	993	256	GûrAsc Zegajo

WJ 45	neutre	2093	236	bab'ûnd'Ajcsc inZe gûd'A d'akago gas'ûld'A
WJ 46	neutre	2241	233	mjcndodo jeb'ûge hago nagas'cja hAnnûnde
WJ 47	neutre	2415	235	gûrcke sinkjcNûl mani nos'ûngcjejo ob'anûn
WJ 48	neutre	1105	192	ZarinZi morûgu
WJ 49	neutre	2461	239	Zchigibi d'ari dwege manûn Zibigcdûnjo
WJ 410	neutre	1540	234	Zega ilgobZ'A d'arinde
WJ 51	neutre	1933	230	je gûrAsc Zchigibi dwege cmhagcdûnjo
WJ 52	neutre	1976	226	cnnidûri wcnag d'aldûri manki d'Amune
WJ 53	neutre	1500	205	cnnidûri dwege ZarAs'cjo
WJ 54	neutre	2090	229	Zalhagu d'o cmmado mullon ZalhasjcZ'iman
WJ 55	neutre	2260	228	cnnidûri ncmuncmu ZalhAwag'ld'Amune
WJ 56	neutre	1622	200	Zega ircnûnge
WJ 57a	neutre	1452	219	crûndûrûn ihAga angasiZjo
WJ 57b	neutre	1714	208	gûrûgu gûrAscnûn
WJ 58	neutre	2310	220	Zega sod'igu ob'aga mald'igcdûnjo
WJ 59	neutre	1872	212	gûrAsc a sod'iraN mald'inûn
WJ 510	neutre	2155	201	guNhabi anmad'a ircnmaldu ik'u
WJ 511	neutre	2477	230	d'o Zega saZurûl bomjcn siZibûl nûg'e dwendejo
WJ 512	neutre	2358	204	gûrcnge inZe cmmanûn ihArûl motasinûn gcjejo
WJ 513a	neutre	735	203	gûrcke saZudu
WJ 513b	neutre	1124	201	anmak'u
WJ 514a	neutre	1843	198	gûrûdu ncnûn siZibûl nûg'e gaja dwenûnde
WJ 514b	neutre	1670	201	we cmmamarûl andûnni hagusc
WJ 515a	neutre	2547	210	ihArûl motasejo
WJ 515b	neutre	1856	208	ZchiZibi wenag d'aldu manku
WJ 61	neutre	1302	210	gûrcnig'an ob'anûnjo
WJ 62	neutre	1780	206	I iri olta sANgakagu
WJ 63	neutre	1981	231	iredu dwege jclZ'cNZcgin saramijejo
WJ 64	neutre	1542	206	ob'ae ZaNZ'cmindejo
WJ 65	neutre	2093	224	modûn ire jclZ'cNZcgigu gûng'a Zigûmûn
WJ 66	neutre	1717	199	Zigûm saNhwaNi dwege crjcunig'anjo
WJ 67a	neutre	1802	233	dwege ob'adu himdûlg'cjejo
WJ 67b	neutre	1613	237	gûrcnge ob’a scNg’jcgijo
WJ 68	neutre	2302	246	tSwescnûl modûn irûl da tSwescnûl dahAsc jcls'imhi hAjo
WJ 69	neutre	1599	242	gûge ob'ae ZaNZ'cmijejo
WJ 71	négative	1018	244	je hûndûlljcs'cjo
WJ 72	négative	784	225	Zcnûnjo
WJ 73	négative	1045	218	wcnag ZchiZibi
WJ 74	négative	1231	201	d'o uAdo Zokcdûnnjo
WJ 75	négative	1522	273	ob'arûl saraNhaZiman
WJ 76	négative	1733	220	gaZogd'ûldo ZcNmal ZuNjohagcdûnjo
WJ 77	négative	2281	251	Zega cd'cngcl tAkalZ'I morûgennûngcjejo

WJ 78	négative	1606	200	t'AkaZani gaZogi ncmu Zoku
WJ 79	négative	2361	195	gaZogûl t'AkaZani ob'aga ncmu Zoku gûrAscnûn
WJ 710	négative	1514	219	gald'ûNûl ZcNmal manni haAg'cdûnjo
WJ 711	négative	1296	251	ibcn pûrorûl toNhAsc
WJ 712	négative	2778	182	uri sigg'urûri ob'arûl ZcNmal Zal bwaZwcs'ûmjcn
WJ 713	négative	1541	192	gûrcn maûmûro nawas'ûmnida
WJ 81	négative	2509	200	ircke Zega naon gc da morûsigcdûnjo
WJ 82a	négative	1283	228	ab'a cmma ZcNmal mianhAjo
WJ 82b	négative	1800	158	ircke mals'ûm andûrigu nawascjo
WJ 83	négative	2747	201	gûrigu cnnidûlhagu hjcNbudûlhantedu ZcNmal mianhAjo
WJ 84	négative	1655	212	ûm ZcNmal mianhagjeseNgakejo
WJ 85	négative	960	190	clmaZcne
WJ 86	négative	1064	219	terebiZjcnZuNe
WJ 87	négative	1705	237	Ainiranûn pûroga hAs'cZ'ijo
WJ 88	négative	2610	315	gû pûrorûl bomjcnsc musûn sANgagûl hAnnjahamjcn
WJ 89a	négative	1988	280	ob'aga gjcronûl anhan saNtAesc
WJ 89b	négative	975	226	narûl mannasc
WJ 89c	négative	2483	302	nan ZcNmal hANbokagunarago sANgagûl hAs'cjo
WJ 810	négative	2577	288	gû dûrama ZatSega gjclhonûl han saNtAesc
WJ 811	négative	1965	252	saraNhanûn jcinûl mannanûn gcjcZ'annajo
WJ 812	négative	588	198	cmma
WJ 813	négative	443	209	ab'a
WJ 814a	négative	1223	255	dasi hanbcnman bwaZusejo
WJ 814b	négative	1278	230	isaram cd'cn saraminga
WJ 814c	négative	1120	241	gwentSanûn saraminga
WJ 815	négative	1525	237	dasi hanbcnman bwaZusigujo
WJ 816	négative	916	192	gûrûgu maZimagûru
WJ 817	négative	1942	213	uri namdoNsANi gosamigcdûnjo
WJ 818	négative	1047	255	hjcNgjuna nunaga
WJ 819	négative	1000	215	nc goNbuhanûnde
WJ 820	négative	590	174	ircke
WJ 821	négative	2795	253	Ziban ZojoNhage mandûlgo sipcnnûnde ZcNmal mianhagu
WJ 822a	négative	1620	248	uri hjcNgjuni inZen crûniZi
WJ 822b	négative	1077	238	goNbu jcls'imi hagu
WJ 822c	négative	1105	267	nuna ihAhagenni
WJ 823	négative	743	260	ihAhAZura

Tableau 17. Transcription phonétique (Alphabets Phonétiques Internationaux) des énoncés du corpus coréen, et l'émotion, la durée et le Fo moyen des énoncés.

Transcription coréenne

WJ 11 아이, 제 성격이요,
WJ 12 누가 그렇게 갑자기 앞에 다가오면
WJ 13 약간 뒤로 아렇게 주춤하고 물러서는 성격이거든요.
WJ 14 그렇게 말씀을 하시길래,
WJ 15 선배님하고 나이차이가 많았거든요,
WJ 16 처음에 우리 데이트할래 이러길래
WJ 17 어 안돼요 하고 말을 했다가
WJ 18 아 내가 이게 과민반응이 아닌가 해서는
WJ 19 그러며는 식사는 괜찮아요

WJ 21 예 상욱이 오빠는요,
WJ 22 마음이 되게 넓구,
WJ 23 베풀줄 아는 사람이예요.
WJ 24 딱 외모를 봐두 산타크로스 닮지 않았어요?
WJ 25 어떤 자그마한 것을 소중하게 아는 그게
WJ 26 어떤 커다란 것만 중요하구 뭐 이렇게 그러는게 아니구
WJ 27 자그만 사랑도 이렇게 소중히 할줄 아는
WJ 28 그런 사람이어서 되게 좋았어요.

WJ 31 예, 오빠가 인제 꿈이
WJ 32 문화공간을 만드는 거니깐
WJ 33 그런걸 위해선 어떤 자금이 필요하잖아요.
WJ 34a 그래서 이렇게 저금통에 오백원짜리
WJ 34b 둘씩 모이면서 그런
WJ 34c 꿈을 이뤄갈수 있다고 생각을 했거든요
WJ 35 꿈을 이뤄가는데
WJ 36 도움이 됐으면하는 마음으로 저금통을 선물했습니다.
WJ 37 예 오백원짜리하나 넣어서요.

WJ 41 처음예요.
WJ 42a 오빠가 처음
WJ 42b 엄마를 뵌게요,
WJ 43 전시회 가서 뵀거든요.
WJ 44a 굉장히 일이 바쁜 때였어요.
WJ 44b 그래서 제가요,
WJ 45 바쁜 때여서 인제 그때 딱 하고 갔을 때
WJ 46 면도도 예쁘게 하고 나갔어야 했는데
WJ 47 그렇게 신경을 많이 못 쓴거예요 오빠는

WJ 48 자린지 모르구
WJ 49 저희 집이 딸이 되게 많은 집이거든요
WJ 410 제가 일곱째 딸인데,

WJ 51 예 그래서 저희집이 되게 엄하거든요.
WJ 52 언니들이 워낙 딸들이 많기 때문에
WJ 53 언니들이 되게 잘했어요.
WJ 54 잘하구 또 엄마도 물론 잘하셨지만
WJ 55 언니들이 너무너무 잘해왔기 때문에
WJ 56 제가 이러는게
WJ 57a 어른들은 이해가 안가시죠.
WJ 57b 그르구 그래서는,
WJ 58 제가 소띠구 오빠가 말띠거든요.
WJ 59 그래서 아 소띠랑 말띠는
WJ 510 궁합이 안 맞다 이런 말두 있구
WJ 511 또 제가 사주를 보면 시집을 늦게 가야 된대요.
WJ 512 그런게 인제 엄마는 이해를 못하시는 거예요.
WJ 513a 그렇게 사주두
WJ 513b 안 맞구,
WJ 514a 그르구 너는 시집을 늦게 가야 되는데
WJ 514b 왜 엄마 말을 안 듣니 하구서
WJ 515a 이해를 못하세요
WJ 515b 저희집이 워낙 딸두 많구

WJ 61 그러니깐 오빠는요
WJ 62 이 일이 옳다 생각하구
WJ 63 일에두 되게 열정적인 사람이예요.
WJ 64 오빠의 장점인데요.
WJ 65 모든 일에 열정적이구 근까 지금은
WJ 66 지금 상황이 되게 어려우니깐요
WJ 67a 되게 오빠두 힘들꺼예요.
WJ 67b 그런게 오빠 성격이예요.
WJ 68 최선을 모든 일을 다 최선을 다해서 다 열심히 해요.
WJ 69 그게 오빠 장점인예요.

WJ 71 예 흔들렸어요
WJ 72 저는요
WJ 73 워낙 저희집이
WJ 74 또 우애도 좋거든요.
WJ 75 오빠를 사랑하지만
WJ 76 가족들도 정말 중요하거든요.

WJ 77 제가 어떤걸 택할지 모르겠는 거예요.
WJ 78 택하자니 가족이 너무 좋구
WJ 79 가족을 택하자니 오빠가 너무 좋구 그래서는
WJ 710 갈등을 정말 많이 했거든요.
WJ 711 이번 프로를 통해서
WJ 712 우리 식구들이 오빠를 정말 잘 봐줬으면
WJ 713 그런 마음으로 나왔습니다.

WJ 81 이렇게 제가 나온 거 다 모르시거든요.
WJ 82a 아빠 엄마 정말 미안해요.
WJ 82b 이렇게 말씀 안 드리구 나와서요.
WJ 83 그리구 언니들하구 형부들한테두 정말 미안해요.
WJ 84 음, 정말 미안하게 생각해요.
WJ 85 얼마전에
WJ 86 테레비전중에
WJ 87 애인이라는 프로가 했었죠.
WJ 88 그 프로를 보면서 무슨 생각을 했냐하면,
WJ 89a 오빠가 결혼을 안한 상태에서
WJ 89b 나를 만나서
WJ 89c 난 정말 행복하구나라고 생각을 했어요.
WJ 810 그 드라마 자체가 결혼을 한 상태에서
WJ 811 사랑하는 여인을 만나는 거였잖아요.
WJ 812 엄마,
WJ 813 아빠,
WJ 814a 다시 한 번만 봐 주세요.
WJ 814b 이 사람 어떤 사람인가,
WJ 814c 괜찮은 사람인가.
WJ 815 다시 한 번만 봐 주시구요.
WJ 816 그르구 마지막으루
WJ 817 우리 남동생이 고3이거든요.
WJ 818 형균아, 누나가
WJ 819 너 공부하는데
WJ 820 이렇게
WJ 821 집안 조용하게 만들고 싶었는데 정말 미안하구,
WJ 822a 우리 형균아 인젠 어른이지?
WJ 822b 공부 열심히 하구,
WJ 822c 누나 이해하겠니?
WJ 823 이해해주라.

Transition de l'état émotionnel (Expression facale de l'émotion)

Corpus anglais

Transcription phonétique

Fichier	Emotion	Durée	Fo Moy.	Enoncés
FB 1	neutre	1391	264	hi teiks mai kar kiz
FB 2	neutre	2249	212	mai hEsbEnd bûnd El mai piktjuûz
FB 3a	neutre	1124	325	wai ar ju
FB 3b	neutre	1205	293	sou impElst bai mi
FB 4a	neutre	871	276	wat hAv ai dEn
FB 4b	neutre	1084	284	for Te ten iûz
FB 5a	neutre	1298	234	bEt ai ni:d help meikiN TAt
FB 5b	neutre	1100	228	meikiN TAt kúnektion
FB 6a	neutre	1014	247	ai ken júst teike mai kiz
FB 6b	neutre	1333	231	goiN tu tempErEri hauziN
FB 7	neutre	1269	246	fû difErEnt ri:sEns
FB 8	neutre	1382	204	fû spendiN tu mutS mEni
FB 9	neutre	1238	219	fû kukiN Te wroN fud
FB 10a	émotionnel	1000	324	weû ai ken gou
FB 10b	émotionnel	1042	358	duûriN Te dei
FB 11a	émotionnel	1759	318	if Teûs Aani|iN difErEnt
FB 11b	émotionnel	1434	331	ai em gona di:l wiT him fû Te list
FB 12a	émotionnel	956	305	aiv gotE kE:l him
FB 12b	émotionnel	1061	294	end Ask his pûmiSEn
FB 13	émotionnel	1113	337	end hi didnt laik TAt
FB 14	émotionnel	1434	272	hi bi:t mi úp fû TAt
FB 15	émotionnel	1780	240	ai spent |ûti dalarz En maiself
FB 16	émotionnel	1582	257	aiv bi:n goiN tu |erapi
FB 17	émotionnel	1063	316	ai ni:d tu get aut
FB 18a	émotionnel	1042	284	Teû ar pi:pl
FB 18b	émotionnel	947	298	pi:pl aut Teû
FE 1a	neutre	1002	184	ai |iNk its morû ov
FE 1b	neutre	1258	187	a lúve seksiN
FE 2	neutre	956	222	nou hi dEz nat
FE 3a	neutre	1159	235	aim nat TAt kaind ov pûsEn
FE 3b	neutre	720	199	hi knouz TAt
FE 4	neutre	800	181	fû E iû
FE 5	émotionnel	1022	219	Tis iz nat lúve
FE 6a	émotionnel	1145	248	of kûrs
FE 6b	émotionnel	1185	215	hi knouz ai lúv him
FE 7a	émotionnel	845	234	ai hAv Efeû
FE 7b	émotionnel	819	236	fû lAst iû
FE 7c	émotionnel	880	229	wiT his best frend
FE 8	émotionnel	1210	233	wiT his best frend

FG 1	neutre	934	184	siksti |auZEnd dallûz
FG 2	neutre	1025	201	hi wEs TE wan its dEn it
FG 3	neutre	1602	205	tel him hau lai daun Ep tu Tis morniN
FG 4	neutre	878	267	wel ai wEs saiko bEt
FG 5	neutre	1505	182	ju hAv tu meik sjuû ju ar oukei
FG 6	neutre	1757	211	wan Eklck end ju wû in nju jork siti
FG 7	neutre	2081	184	This is nat bûrmiNhAm alabama
FG 8	neutre	790	187	jes ju ar
FG 9	neutre	1946	206	ju pramist mi jul bi bAk At wan Eklck
FG 10	émotionnel	1813	216	ai dont si: hau Tis kid kEnvei Tis
FG 11	émotionnel	1293	224	juû maTû meid misteiks
FG 12	émotionnel	1641	238	pli:z let mi finiS it oukei
FG 13	émotionnel	1622	266	dont tel mi wat En alkoholik iz
FG 14	émotionnel	1126	217	hi ded bikoz ov it
FG 15	émotionnel	1086	253	dont tel mi wat it iz
FG 16	émotionnel	1212	221	ai nou wat hi put
FG 17	émotionnel	1166	215	faTû wEs En alkoholik
FH 1	neutre	2112	193	bikoz its morû Ediktd tu mai lEvû
FH 2	neutre	1560	182	end aid got hepetaizd nau
FH 3	neutre	1143	176	Entil ai did artikl
FH 4	neutre	1338	215	Si didnt wEnt tu gou tu sku:l
FH 5	neutre	1354	187	At TAt taim Si wEs sevEnti:n
FH 6	neutre	990	186	got sEme help
FH 7	neutre	1591	199	ai du hAv fû eiz
FH 8	émotionnel	1504	251	ai brEt mai dotû tu liv wiT mi
FH 9	émotionnel	1375	248	ai met E gai fûst ov El
FH 10	émotionnel	956	216	Aftû aiv got kli:nd
FH 11	émotionnel	1757	313	ai wEnt ju pli:z mai laif
FH 12	émotionnel	1571	187	aim lucki fû ai got E iû left
FH 13	émotionnel	1700	184	end mai lEvû bikoz ov TE eiz
FH 14	émotionnel	770	253	aiv kli:nd
FH 15	émotionnel	2058	217	aiv bin kli:nd fû kEpl ov iûz
FH 16	émotionnel	1642	232	teresi keim tu liv wiT Es
FI 1	neutre	971	186	wi wud jel bAk At hû
FI 2	neutre	1052	224	TAt didnt get Es Aniweû
FI 3	neutre	1642	195	Si hEs En opEn komjunikeiSEn
FI 4	neutre	1401	183	Si tEk wiT hû dEd end ai
FI 5	neutre	1174	189	aif Si hAz E prEblm
FI 6	neutre	610	195	Si hEd E fit
FI 7	neutre	1027	203	Ten wi El fû start it
FI 8	neutre	1304	207	ai min wiv jEst Ebaut
FI 9	neutre	1160	231	definitli ai rembû
FI 10a	émotionnel	1110	275	ai hAv tu teik ju
FI 10b	émotionnel	1170	364	ai dont wanna put
FI 10c	émotionnel	940	282	put ju sEmweû
FI 11a	émotionnel	1821	338	ou mai gad
FI 11b	émotionnel	1567	341	ou mai god

FI 12	émotionnel	922	285	ai lEv ju
FI 13	émotionnel	2000	239	Teû iz na|iN els ai ken du Sena
FI 14	émotionnel	2628	224	ju meid mi hAv tu hAv E tEf lEv

Tableau 18. Transcription phonétique (Alphabets Phonétiqes Internationaux) des énoncés du corpus anglais, et l'émotion, la durée et le Fo moyen des énoncés.

Transcription anglaise

FB 1	He takes my car keys.
FB 2	My husband burned all my pictures.
FB 3ab	Why are you so impulsed by me ?
FB 4ab	What have I done for the 14 years.
FB 5ab	but I need help making that, making that connection.
FB 6ab	I can just take my kids going to temporary housing.
FB 7	for different reasons
FB 8	for spending too much money
FB 9	for cooking the wrong food
FB 10ab	Where I can go during the day ?
FB 11ab	if there's anything different, I'm gonna deal with him for the list.
FB 12ab	I've gotta call him and ask his permission.
FB 13	and he didn't like that.
FB 14	He beat me up for that.
FB 15	I spent 30 dollars on myself.
FB 16	I've been going to therapy.
FB 17	I need to get out
FB 18ab	There are people, people out there.
FE 1ab	I think it's more of a love sexing.
FE 2	No, he does not.
FE 3ab	I'm not that kind of person, he knows that.
FE 4	for a year
FE 5	This is not love.
FE 6ab	Of course, he knows I love him.
FE 7abc	I have affaire for last year with his best friend.
FE 8	with his best frined
FG 1	sixty thousand dollars
FG 2	He was the one it's done it.
FG 3	Tell him how lie downs up to this morning .

FG 4 Well, I was psycho but,

FG 5 You have to make sure you are OK.

FG 6 One o'clock and you were in New York city

FG 7 This is not Burmingham Alabama.

FG 8 Yes, you are.

FG 9 You promised me you'll be back at one o'clock.

FG 10 I don't see how this kid convey this.

FG 11 Your mother made mistakes.

FG 12 Please let me finish it, OK.

FG 13 Don't tell me what an alcoholic is.

FG 14 He dead because of it.

FG 15 Don't tell me what it is.

FG 16 I know what he put.

FG 17 Father was an alcoholic.

FH 1 Because it's more addicted to my lover

FH 2 and I'd got hepetized now.

FH 3 until I did article

FH 4 She didn't want to go to school.

FH 5 At that time she was seventeen.

FH 6 got some help

FH 7 I do have four blond aids.

FH 8 I brought my daughter to live with me.

FH 9 I met a guy first of all.

FH 10 After I've got cleaned.

FH 11 I want you please my life.

FH 12 I'm lucky for I got a year left.

FH 13 and my lover because of the aids.

FH 14 I've cleaned.

FH 15 I've been cleaned for couple of years.

FH 16 Tracy came to live with us.

FI 1 We would yell back at her.

FI 2 That didn't get us anywhere.

FI 3 She has an open communication.

FI 4 She talk with her dad and I.

FI 5 If she has a problem

FI 6 She had a fit

FI 7 When we all for start it,

FI 8 I mean we've just about.

FI 9 Definitely I remember.

FI 10abc I have to take you, I don't wanna put put you somewhere

FI 11ab Oh, my god, Oh, my god,

FI 12 I love you.

FI 13 There is nothing else I can do Sena.

FI 14 You made me have to have a tough love.

Questionnaire du test de perception

Answer Sheet

Gender : M /F Age : _____ Native language : _____ Testing Date : ____/____/____

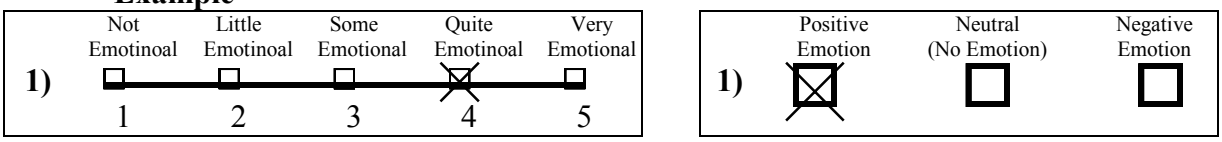
You are going to listen to a series of vocal stimuli, which are segments of speech extracted from an TV interview of a Korean lady. The stimuli can be phrases or words, so their length varies greatly. You have two tasks to accomplish in two different sessions :

(1) In the first session, you are to estimate, after listening to each stimulus, how much emotion is expressed in the stimulus ('**Estimation of Emotional Intensity**'). As soon as your decision is made, mark an 'x' on one of the five points, labeled as *'not emotional,' 'little emotional,' 'some emotional,' 'quite emotional'* and *'very emotional'*

(2) In the second session, you are to identify, after listening to each stimulus, what kind of emotion (*positive, neutral* or *negative*) is expressed in the stimulus ('**Identification of Emotional meaning**'). As soon as your decision is made according to your subjective impression, mark an 'x' on one of the three boxes, labeled as '*positive emotion*,' '*neutral emotion* (*no emotion*)' and '*negative emotion*'.

** Please do not be concerned about understanding meaning of the stimuli. Your decision should be made according to your subjective impression, not to the lexical meaning of the stimuli.

Example



****

Task 1: How much emotional does it sound?

	Not	Little	Some	Quite	Very
1)	1	2	3	4	5
2)	1	2	3	4	5
3)	1	2	3	4	5
4)	1	2	3	4	5
5)					

	Not	Little	Some	Quite	Very
6)	1	2	3	4	5
7)	1	2	3	4	5
8)	1	2	3	4	5
9)	1	2	3	4	5
10)					

Task 2: Is it a Positive or a Negative Emotion?

	Positive	Neutral	Negative
1)			
2)			
3)			
4)			
5)			

	Positive	Neutral	Negative
6)			
7)			
8)			
9)			
10)			

… … …

	Not	Little	Some	Quite	Very
76)	1	2	3	4	5
77)	1	2	3	4	5
78)	1	2	3	4	5
79)	1	2	3	4	5
80)					

	Positive	Neutral	Negative
76)			
77)			
78)			
79)			
80)			

	Not	Little	Some	Quite	Very
81)	1	2	3	4	5
82)	1	2	3	4	5
83)	1	2	3	4	5
84)	1	2	3	4	5
85)					

	Positive	Neutral	Negative
81)			
82)			
83)			
84)			
85)			

	Not	Little	Some	Quite	Very
86)	1	2	3	4	5
87)	1	2	3	4	5
88)	1	2	3	4	5
89)	1	2	3	4	5
90)					

	Positive	Neutral	Negative
86)			
87)			
88)			
89)			
90)			

	Not	Little	Some	Quite	Very
91)	1	2	3	4	5
92)	1	2	3	4	5
93)	1	2	3	4	5
94)	1	2	3	4	5
95)					

	Positive	Neutral	Negative
91)			
92)			
93)			
94)			
95)			

	Not	Little	Some	Quite	Very
96)	1	2	3	4	5
97)	1	2	3	4	5
98)	1	2	3	4	5
99)	1	2	3	4	5
100)					

	Positive	Neutral	Negative
96)			
97)			
98)			
99)			
100)			

Glossaire[150]

Accent : mise en relief de nature syntagmatique* ; l'accent d'`insistance` (fonction* expressive) n'affecte qu'une unité lexicale ('formidable) ; l'accent d'`intensité` ou tonique* (fonction* contrastive) contribue au rythme*. Acoustiquement, l'accent est lié essentiellement à la variation d'intensité de la voix, bien que la durée, la hauteur et/ou le timbre puissent également jouer un rôle compensatoire important.

Acoustique : étude physique des sons*, abstraction faite des sensations qu'ils provoquent. Une des difficultés de la phonétique réside dans l'établissement des corrélats physiologiques, acoustiques et perceptifs.

Activation : dans la psychologie de l'émotion, l'`activation` réfère à la stimulation des nerfs à cause de l'excitation émotionnelle. Le niveau de l'activation, estimé par des mesures physiologiques (ex. la fréquence, l'intensité et le rythme de la voix, la pression du sang, la vitesse du battement du coeur, le mesure galvanique de la réponse du peau, etc.), indique le niveau d'excitation émotionnelle d'un sujet. L'axe d'`activation` s'étend du `sommeil` à la `tension.` Cet axe a été proposé par Schlosberg (1954) en tant qu'une des trois dimensions majeures de l'émotion avec l'axe de `valence`* et l'axe de `puissance`*.

Aigu : élevé en hauteur*, de haute fréquence*

Amplitude : v. `Intensité`.

Articulation : ensemble des mouvements des organes supraglottiques. V. `Phonation`.

Bruit : signal acoustique ne possèdent aucune périodicité ;

Br. impulsionnel : signal acoustique apériodique très bref (ex. les occlusives) ;

Br. continu : signal acoustique apériodique possédant une certaine durée ;

Br. blanc : bruit continu dont le spectre s'étend sur toute la gamme des fréquences audibles avec une amplitude constante ; utile dans les expériences de perception de parole masquée ;

Br. coloré (rose) : bruit continu dont le spectre s'étend seulement sur une partie de la gamme des fréquences ; c'est en fait un bruit blanc filtré dans cette gamme de fréquences.

Cordes vocales : replis muscluo-membraneux du larynx*, entre lesquels se trouve la glotte* et qui constituent l'organe essentiel de la phonation* (sons* produits par vibration) V. `Voix`.

Cycle : v. `Fréquence`, `Période` et `Hertz`.

Débit : vitesse d'élocution, elle se mesure en syllabes par seconde. Syn. : tempo.

Décibel : unité logarithmique* de mesure de l'intensité* des sons. La gamme des intensités sonores s'étend, en dB `absolus`, de 0 (seuil d'audibilité) à 140 (seuil de douleur). (En d'autres termes, de 10^{-16} Watt/cm^2, ce qui correspond à une pression sonore de 20 μ Pa, à 10^{-2} Watt/cm^2, ce qui correspond à une pression sonore de 2.10^8 μ Pa. On peut aussi parler d'intensité sonore relative entre deux sons (tel son est 10 fois plus intense que tel autre). En ce cas, le dB n'est pas une unité fixe, mais `relative` (dB `relative`) : il indique un `rapport` (ex. : un son 100 fois plus intense qu'un autre est 20dB plus intense).

Dialecte : variété régionale d'une langue* qui, pour des raisons extralinguistique (historiques, régionales et socioculturelles), n'est pas été promue à un statut dominant.

Discours : ensemble d'énoncés produits par une personne ou un ensemble de personnes. V. `entretien`.

Distinctif : qui, pouvant être distingué, remplit un rôle, une fonction dans un système. Syn. : pertinent, fonctionnel.

Durée : qualité d'un son liée au facteur temps. Selon sa durée, le son est dit bref ou long.

[150] Les références principales de cette partie sont « *Le Robert : dictionnaire de la langue franaçaise* », « *Eléments de phonétique* » (Landercy & Renard, 1977) et « *Principles of Phonetics* » (Laver, 1994).

Enoncé : résultat de l'énonciation* (opposé à énonciation) ; segment de discours produit par l'énonciation d'un locuteur et dont les limites sont variables selon les critères (phonétiques ; syntaxiques, etc.).

Emotion : Etat de conscience complexe, généralement brusque et momentané, accompagné de troubles physiologiques (pâleur ou rougissement, accélération du pouls, palpitations, sensation de malaise, tremblements, incapacité de bouger ou agitation) ; sensation (agréable ou désagréable), considérée du point de vue affectif.

Enoncer : exprimer en termes nets, sous une forme arrêté (ce qu'on pense, ce qu'on a à dire) ; s'exprimer.

Enonciation : production individuelle d'un énoncé dans des circonstances données de communication.

Entretien : action d'échanger des paroles avec une ou plusieurs personnes ;

entretien entre deux personnes – dialogue ;

entretien entre plusieurs interlocuteur – colloque, conférence.

Fonction : rôle ; les linguistes distinguent, entre autres :

`F. contrastive` (ou culminative) : qui met en relief (p. ex. l'accent* en français) ;

`F. démarcative` (ou syntagmatique) : qui sert à découper (délimiter) des éléments de la chaîne parlée (p. ex. les pauses*) ;

`F. distinctive` : qui sert à distinguer (p. ex. l'accent anglais ; per'mit ~ 'permit) ;

`F. expressive` (ou émotive, ou phonostylistique) : qui informe sur l'affectivité du locuteur (p. ex. le trémolo*) ;

`F. métalinguistique` : qui sert à parler de la langue ;

`F. phatique` : qui sert à garder ou à couper le contact entre les interlocuteurs (p. ex. les clignements d'yeux, le mm d'approbation…) ;

`F. syntaxique` (ou logique) : qui établit une relation entre les différents éléments syntagmatiques.

On peut qualifier de `linguistique` tout ce qui correspond à la fonction de `communication`, et qui fait appel à des moyens expressifs `non marqués` par rapport à la norme attendue (le terme `linguistique` reprend donc ce que d'autre appellent : syntaxique, logique, phonologique, sémantique. V. Linguistique*). On peut qualifier de `supralingustique` tout ce qui correspond à la fonction d'`expression`, et qui fait appel à des moyens expressifs `marqués` (le terme `supralinguistique` reprend donc ce que d'autre appellent : émotif, expressif, phonostylistique).

Fonctionnel : qui a une fonction, un rôle dans le système étudié. Syn. : distinctif, pertinent.

Fondamental : inverse de la période* ; dans le spectre* d'un son périodique*, le fondamental, qui donne la sensation de hauteur*, est la composante la plus basse, le plus grand commun diviseur de l'ensemble des harmoniques* ; dans la voix*, le fondamental correspond à la fréquence* de vibration des cordes vocales. L'abréviation de la fréquence fondamentale est `Fo`. Alors que l'anglais emploie ***Fo*** pour la valeur physique de la fréquence fondamentale et ***Pitch*** pour la hauteur perçue de la voix, le français emploie indifféremment `Fo`, en recourant selon le cas aux adjectifs `physique` ou `subjective`. Syn. : premier harmonique.

Pitchmeter : appareil qui permet de mesurer l'évolution temporelle du fondamental* de la voix. Syn. : détecteur de mélodie.

Formant : dans le spectre* d'un son vocalique, zone de fréquences* de plus grande intensité* ; le timbre* vocalique est en relation avec la configuration formantique.

Fréquence : nombre de vibrations (cycles*, périodes*) par seconde ; elle est exprimée en hertz* (Hz) ; l'échelle des fréquences audibles s'étend de 16 à 16.000Hz.

Gosier : arrière-gorge et pharynx ; siège de la voix*, prolongement du pharynx* communiquant avec le larynx*. V. `Gorge`.

Grave :bas, dans l'échelle des hauteurs*, des fréquences*.

Glottalisation : occlusion des cordes vocales.

Glotte : espace situé entre les cordes vocales.

Harmonique : dans le spectre* d'un son périodique*, par exemple, un son musical, les *`harmoniques`* sont des fréquences* multiples de la fréquence fondamentale (le fondamental*) ; la répartition des harmoniques détermine le timbre*. N.B. Le fondamental est souvent appelé « premier harmonique ».

Hauteur : qualité d'un son qui semble liée principalement à sa fréquence*[151]. Selon la *hauteur*, le son est dit grave* ou aigu*. Cf. *Fondamental*.

Hertz (Hz) : unité de fréquence*, égale à une période* (ou un cycle*) par seconde.

Indices : signe* apparent qui indique quelque chose avec possibilité. V. Signe.

Intensité : qualité d'un son qui semble liée principalement à l'amplitude de ses vibrations ; elle est mesurée selon une échelle logarithmique (en décibels*, dB), selon l'*intensité*, le son est dit faible ou fort. En fait, physiquement, on parlera de pression, d'intensité ou de puissance sonores, proportionnelles à l'amplitude des vibrations, tandis qu'au niveau des sensations, on parlera d'intensité subjective. On ne saurait assez insister sur le caractère équivoque de la terminologie française. Alors que l'anglais emploie ***intensity*** pour l'intensité physique et ***loudness*** pour l'intensité subjective, le français emploie indifféremment intensité, en recourant selon le cas aux adjectifs physique ou subjective.

Intonation : au sens strict, courbe mélodique que l'on peut abstraire de l'analyse de la perception d'un énoncé parlé. Elle semble liée principalement aux variations de la hauteur* du ton* laryngien dans une phrase déterminée ; dans le contexte des procédés verbo-tonaux* de correction phonétique, on parlera d'intonations « montante » ou « descendante » à propos d'énoncés dont la courbe ainsi définie s'élève ou s'abaisse régulièrement du début à la fin. – L'intonation peut être distinguée de la mélodie* et considérée comme l'intégration perceptive globale des différents éléments prosodiques* (mélodie, tons, pauses, accents, rythme).

Langage : système de communication (terme général).

Langue : système théorique qui structure la parole* ; code, « système dont tous les termes sont solidaires » (Saussure), propre à un ensemble de sujets. V. Dialecte.

Larynx : comprenant la glotte*, organe essentiel de la phonation*, dont il constitue la source des sons voisée*. V. Voix.

Linguistique : nom de la science du langage ; adjectif relatif à la langue. Laver (1994) distingue trois types d'acte langagier (linguistique, paralinguistique, extralinguistique) selon le degré du codage communicatif. Tous les trois types d'acte langagier sont informatifs, mais seulement les actes linguistique et paralinguistique sont codés et communicatifs :

A. linguistique : acte associé à la langue ; sa fonction communicative est accomplie par le double codage de la langue parlée basé sur des règles phonologiques et grammaticales. Autres formes de l'acte linguistique (comme la communication par écrit et le langage par signes) existent mais ce sont des formes mineures, par rapport à la forme parlée, dans la communication linguistique.

A. paralinguistique : acte communicatif qui est non-linguistique et non-verbal, mais qui est pourtant codé. Cet acte sert à communiquer l'état émotionnel (tristesse, joie, colère, etc.) ou l'attitude du locuteur et à signaler le tour de parole dans la conversation. Les indices paralinguistiques (ex. le ton de la voix (timbre*), l'expression dans le visage et le geste) dépendent de la culture, et leur interprétation conventionnelle sont à être apprise dans une société donnée.

A. extralinguistique : acte informatif mais ni codé ni communicatif. Les aspects extralinguistiques de la parole sont riches en termes des évidences informatives de l'identité du locuteur (ex. la qualité de la voix et la plage* moyenne de hauteur et d'intensité).

Les caractéristiques socioculturels du locuteur, qui ne sont pas innés mais appris, sont à être révélés dans les actes linguistique et paralinguistique, tandis que les attributs physiologiques, qui sont plutôt innés, dépendent de l'acte extralinguistique. Cependant, ce genre d'attribution ne peut être directe puisque l'effet d'apprentissage et la nature innée sont mêlés dans les caractéristiques (socioculturels et physiologiques) du locuteur. Le rôle respectif des indices linguistiques, paralinguistique et extralinguistique dans la caractérisation du locuteur reste à étudier dans ce domaine de la recherche.

Logatome : groupe phonique sans valeur significative utilisé en phonétique, à des fins expérimentales et quelquefois pratiques (correction phonétique). Il apparaît en effet malaisé de déterminer ce qui, dans la perception de la parole, relève de l'audition au sens strict et d'autres facteurs, d'ordre sémantique, notamment.

Marque : signe* distinctif*.

Mélodie : sensation liée à l'évolution temporelle de la fréquence fondamentale* sur un énoncé. La courbe mélodique est obtenue à partir d'analyseur (*pitchmeter**). V. intonation.

[151] Les terms marqués d'un astérisque(*) font l'objet d'une définition propre. Les termes anglais sont indiqués en ***italique***.

Message (sonore) : énoncé destiné à être compris.

Paradigmatique : par opposition à syntagmatique*, ensemble de flexions d'un modèle donné. Les commutations phonologiques s'opèrent sur l'axe `paradigmatique`.

Parole : acte concret et individuel des sujets usant de la langue* ; cette définition saussurienne ne correspond pas à celle que nous lui donnons lorsque nous envisageons l'acte de parole tant du point de vue de la production que de celui de la perception ; dans cette perspective, la parole intègre non seulement des données linguistiques mais aussi psycho-socio-culturelles et situationnelles. Acoustiquement, la parole résulte de l'excitation des cavités supraglottiques par une source périodique (la voix*) et/ou une source de bruits (explosions ou frictions).

Pauses : arrêt ou suspension dans l'acte phonatoire.

Perception : action (ou son résultat) de saisir le stimulus sonore (ou le langage) par l'esprit et par le sens ; l'audition, au sens courant, elle-même très complexe puisqu'elle résulte de l'intégration de diverses voies (aérienne, osseuse, vibro-tactile), n'est qu'un élément de la perception, phénomène qui met en jeu de nombreux facteurs extra-auditifs d'ordre socioculturel, psychologique et spatio-temporel. Cet ensemble d'éléments où se mêlent le pertinent et le redondant, est perçu globalement et structuré en vue de la compréhension. Le processus complexe d'intégration d'un message oral ne peut être analysé en phases successives que pour les besoins de la discussion théorique et de la terminologie. Le champs sémitique peut être découpé selon l'opération suivante :

- `Détection` : dans le processus d'intégration d'un signal sonore, qu'il s'agisse d'un bruit sans signification ou d'un message oral, la détection apparaît comme le premier stade ; c'est une perception auditive dont on ne peut le type.
- `Distinction` ou `discrimination` : à ce stade, immédiatement après celui de détection, l'élément perçu est distingué d'un autre.
- `Décodage` ou `identification` : dans la perception de la parole, cette opération permet à l'auditeur d'assimiler un son de parole à un phonème inclus dans son système référentiel (code).
- `Reconnaissance` : structuration du décodage ; cette opération fait référence au travail d'élaboration d'un message codé par le recours pas toujours conscient à des moyens expressifs linguistiques et extralinguistique.

Période : intervalle de temps au cours duquel un point donné du mouvement vibratoire accomplit un cycle complet, exemple : mouvement d'un pendule.

Périodique : qui se reproduit régulièrement de la même manière dans le temps.

Pertinent : qui remplit une fonction, un rôle, dans le système étudié. Syn. : distinctif, fonctionnel.

Perturbation de fréquence (***Jitter***) : mesure de la micro-variation du Fo*, cycle* par cycle. Cette mesure a été proposée par Lieberman (1961) en tant que descripteur quantitatif de la qualité de la voix* et elle peut être calculée à travers diverses méthodes de calcul. Dans cette thèse, le ***jitter*** est mesuré par l'estimation du pourcentage de la différence des valeurs de Fo entre les cycles adjacents (V. p119 pour l'expression mathématique de ce calcul).

Perturbation d'intensité (***Shimmer***) : mesure de la micro-variation de l'intensité*, cycle* par cycle. Le ***shimmer***, comme le ***jitter****, varie selon les différentes phonations* et influence la perception* de la qualité de la voix* (comme le degré de rudesse vocale), mais la variation de ***shimmer*** est trouvée moins pertinente dans l'expression et la perception de l'émotion que celle de ***jitter*** (Klingholz & Martin, 1985). Dans cette thèse, le ***jitter*** est mesuré par l'estimation du pourcentage de la différence des valeurs d'amplitude* entre les cycles adjacents (V. p122 pour l'expression mathématique de ce calcul).

Pharyngal : relatif à une articulation réalisée essentiellement à partir d'une action de la base de la langue contre la paroi du pharynx* (ex. arabe /h/).

Pharynx : cavité situé entre l'épiglotte et la luette et dont le rôle de résonateur est très important.

Phonation : ensemble des actes produisant les sons de parole. V. `Articulation`.

Phonème : la plus petite unité fonctionnelle* (distinctive*, pertinente*) d'un système phonologique ; le `phonème` est noté entre traits obliques (/ /) ; il peut se réaliser en des sons* différents, notées entre crochets ([ ]) et appelés `variantes`, `allophones`, `réalisations phonétiques`.

Phonétique : étude des sons du langage ; cette étude, par opposition à la phonologie*, ne se préoccupe pas de la fonction des sons dans le système auquel ils appartiennent.

Phonologie : par opposition à la phonétique*, étude du système des sons* du langage*, de leur rôle dans la langue*.

Phonostylistique : étude des valeurs expressives de la parole*.

Phrase : forme sous laquelle une idée conçue par le sujet parlant s'exprime et se perçoit (sous forme parlée ou écrite) ; les éléments à l'intérieur de la `phrase` sont liées par des rapports phonétiques, grammaticaux, psychologiques ; c'est l'unité de l'analyse syntaxique.

Plage (de Fo) : dans l'analyse acoustique, la plage de Fo est calculée par l'écart entre deux mesures, les valeurs maximale minimale de Fo (FoMax et FoMin), dans une unité analysée (ex. énoncé).

Pragmatique : qui est apte à l'action sur le réel, qui est susceptible d'application pratiques, qui concerne la vie courante ; du point de vue sémiologique, la `pragmatique` est une étude des signes en situation.

Prosodie (éléments prosodiques) : concerne les éléments dynamiques de la chaîne parlée. Variations de hauteur*, intensité et de durée* qui déterminent la mélodie*, les tons*, les pauses*, les accents* et le rythme* et qui sont intégrés globalement au niveau perceptif par l'intonation*. Ces éléments sont aussi appelés suprasegmentaux*.

Puissance : dans la psychologie de l'émotion, la `puissance` distingue entre l'émotion initiée par le sujet et l'émotion initiée par l'environnement. L'axe de `puissance`, s'étend du `dégoût` à la `surprise` et la `peur`, est l'un des trois dimensions majeures de l'émotion avec l'axe de `activation`* et l'axe de `valence`*.

Rythme (de la parole) : organisation de la parole dans son déroulement temporel. Cette organisation peut se décrire en termes de découpage en unités rythmiques, de succession de syllabes accentuées et inaccentuées, de distribution syllabique dans les unités rythmiques, de durée relative des syllabes, de régularité des temps forts, etc.

Segmental : relatif aux phonèmes* (segments de la chaîne parlée).

Segmentation : délimitation d'un segment du signal acoustique continu en tant qu'unité linguistique discontinue (ex. phonème, syllabe, etc.) pour faciliter l'analyse des données de la parole.

Signal : message ou effet à transmettre, véhiculent de l'information au moyen d'un système de communication ; dans l'acoustique, un `signal` est une forme physique (tension électrique, etc.) sous laquelle se transmet une information ; dans la sémiotique, un `signal` est un signe naturel ou fabriqué qui fait agir le récepteur d'une certaine façon (ex. `fonction de signal` - fonction du langage lors qu'il agit sur le récepteur, opposé aux `fonction de symbole`* et `de symptôme`*.

Signal sonore : par l'opposition à message*, énoncé*, sans considération de son aspect sémantique.

Signe : objet matériel simple (figure, geste, couleur, etc.) qui, par rapport naturel ou par convention, est pris, dans une société donnée, pour tenir lieu d'une réalité complexe. F. de Saussure distingue entre le symbole* et le signe (pris maintenant au sens de *signe linguistique*). D'paèrs lui, le *signe linguistique* a un caractère arbitraire, c'est-à-dire il n'y a pas de lien naturel rudimentatire entre le signifiant et le signifié, tandis que le symbole n'est jamais tout à fait arbitraire (par exemple, le symbole de justice ne pourrait être remplacé par un char).

Sons (de la parole) : au sens strict, le plus petit segment d'un énoncé parlé. S'emploie couramment pour désigner un segment sonore correspondant à un phonème* : le son est noté entre crochets ([ ]).

Spectre : représentation amplitude*/fréquence* d'un phénomène sonore à un instant précis ;

S. à raies : les composantes sont discrètes et multiples du fondamental* ;

S. continu : toutes les fréquences peuvent être des composantes (spectre d'enveloppe).

Spectrogramme : document montrant la configuration fréquentielle d'un son ou d'une séquence de sons pendant une durée déterminée ; le `spectrogramme` permet de visualiser les caractéristiques de hauteur*, de timbre* et d'intensité* des sons*.

Suprasegmental : relatif aux unités « supérieures » au phonème* (syllabe, groupe rythmique, phrase). Ce terme, emprunté à l'anglais, n'est guère apprécié par les chercheurs français, qui lui préfèrent prosodique*. Les deux formes sont employés indifféremment. V. `Prosodie`.

Syllabe : unité immédiatement supérieure au phonème*, la `syllabe` apparaît auditivement comme un groupement phonique autour d'un sommet (noyau) de sonorité ; elle peut être ouverte (terminée par une voyelle) ou fermée (terminée par une consonne) ; c'est la plus petit unité perceptive.

Symbole : ce qui, en vertu d'une convention arbitraire ou d'un lien naturel, correspond à une chose ou à une opération qu'il désigne. Le symbole de Bühler est un signe* arbitraire, alors que le symbole de Saussure suppose un lien motivé entre le signifié et le signifiant.

Symptôme : ce qui manifeste, révèle ou permet de prévoir un état ; la `fonction de symptôme` d'un signe* linguistique* réfère à une fonction par laquelle le signe indique un caractère du locuteur.

Syntagmatique : par opposition à paradigmatique*, concerne l'ordre des éléments de la parole.

Synthèse de la parole : ensemble de techniques qui produisent de la parole à partir de procédés mécaniques ou électriques.

Tension : phénomène encore mal connu qui concerne l'énergie neuro-musculaire dépensée pour produire la parole. Quand Delattre parle du « mode tendu » du français, il assimile la tension à la précision articulatoire. On fait même lorsqu'on oppose les sons tendus aux sons relâchés. – La complexité de cette notion, qui peut être analysée selon les aspects physiologique, acoustique et perceptif, invite à la prudence dans la comparaison de son entre eux. Un phonème peut être réalisé selon différents degrés de tension. Par exemple, une voyelle est normalement plus tendue sous l'accent, une consonne l'est davantage à l'initial qu'en finale.

Timbre : qualité qui distingue des sons de même hauteur* subjective, de même intensité et de même durée et qui est déterminée par l'intensité relative des harmoniques* ; le timbre peut notamment être sombre ou clair selon la répartition spectrale favorable aux fréquences graves ou aiguës. Dans cette thèse, le terme `timbre` est utilisé dans un sens globale, désignant la `qualité de la voix` (ex. : la voix cassée et tremblante).

Ton : hauteur de la voix à un moment donné.

T. laryngien : syn. de `Voix` et de `Flux laryngé`.

Langue à ton : chaque syllabe y est dotée d'une hauteur mélodique à fonction distinctive* ;

Trémolo : tremblement de la voix.

Valence : dans la psychologie de l'émotion, la valence concerne la positivité d'une émotion donnée. L'axe de `valence`, dont les deux pôles sont l'émotion `positive` et l'émotion `négative`, est l'une des trois dimensions majeures de l'émotion avec l'axe d'`activation`* et l'axe de `puissance`*.

Voix : complexe sonore dont le support est fourni par le ton* laryngien et auquel peuvent être s'ajouter d'autres signaux apparentés aux bruit* pour constituer la parole.

Contenu du CD

1. Thèse de Soo-Jin CHUNG (2000)

2. Résumé (en français & en anglais)

2.1. Résumés –Sept chapitres

2.2. Résumé 1p

2.3. Résumé 6p

2.4. Summary 1p

2.5. Summary 6p

3. Données d'analyse (fichiers de son)

3.1. Corpus coréen

3.2. Corpus anglais

4. Présentation de la soutenance (Powerpoint)

5. Articles relatifs à la thèse

5.1. Soo-Jin CHUNG (1998)

5.2. Soo-Jin CHUNG (1999)

5.2. Soo-Jin CHUNG (1997)

5.2. Soo-Jin CHUNG (1995)

6. Curriculum Vitae

6.1. Français

6.2. Anglais