

N° d'ordre 25/ 2012 – M / ELN

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediene
Faculté d'Electronique et d'Informatique



MEMOIRE

Présenté pour l'obtention du diplôme de **MAGISTER**

En : **ELECTRONIQUE**

Spécialité : **Communication Parlée**

Par

Mme AKAK Malika

THÈME

**DETECTION DES EMOTIONS DU LOCUTEUR
A PARTIR DES CARACTERISTIQUES
VOCALES EXTRA-LINGUISTIQUES**

Soutenu publiquement le 17/10/2012, devant le jury composé :

Mme BOUDRAA Malika	Professeur à L'USTHB	Présidente
Mr SAYOUD Halim	Professeur à L'USTHB	Directeur de mémoire
Mr TEFFAHI Hocine	Professeur à L'USTHB	Examineur
Mme TALHA Malika	Maître de Conférences/A à L'USTHB	Examinatrice

Résumé :

Dans cette étude nous nous intéressons à la détection des émotions du locuteur à partir des caractéristiques vocales extra-linguistiques extraites des signaux de parole. Pour cela nous avons utilisé une base de données universelle de paroles émotionnelles allemande ORATOR .

Dans l'élaboration de ce travail nous avons basé notre étude sur 11 paramètres prosodiques, testés séparément pour la détection de l'émotion du locuteur suivis par des tests de signification de chaque paramètre.

Les résultats obtenus ont montré l'efficacité de certains paramètres (P4, P5, P6, P8, P11) dans la détection de certaine émotion du locuteur.

Résumé de Mémoire de Magister
AKAK Malika*

Intitulé

**DETECTION DES EMOTIONS DU LOCUTEUR
A PARTIR DES CARACTERISTIQUES VOCALES EXTRA-LINGUISTIQUES**

RESUME

La manifestation des émotions est un domaine particulièrement complexe de la communication humaine, et concerne des sciences pluridisciplinaires telles que la psychologie, la cognition, la philosophie, le traitement du signal et les sciences computationnelles orientées émotion. Cette pluridisciplinarité est due à la haute variabilité du comportement humain qui influence à la fois la production et la perception de ses émotions. L'affect (*les sentiments et leurs changements physiques associés*), la cognition, la personnalité, la culture et l'éthique sont les principales propriétés des émotions décrites dans la littérature. Bien qu'il existe plus d'une centaine de termes qui leur sont liés, six émotions primaires sont largement acceptées dans la littérature : peur, colère, joie, ennui, tristesse et dégoût. Plutchik postule que les émotions plus complexes seraient des états dérivés ou mixés, et apparaîtraient comme des combinaisons, des mélanges ou composés des émotions primaires. D'autres types d'émotions spécifiques (*appelées traits de la personnalité*) sont aussi étudiés dans cette thèse de Magister.

Ceci étant, notre travail de recherche s'intéresse à la détection de ces états émotionnels, uniquement par analyse du signal vocal de l'intéressé. Il s'agit d'identifier certaines émotions d'un locuteur (joie, colère, bonheur, tristesse, anxiété) et certains traits de sa personnalité à partir des caractéristiques de sa parole. Pour ce faire, nous avons essayé plusieurs paramètres prosodiques : Pitch, Durée, Energie, ... et leurs combinés statistiques (maximum, minimum, variation, moyenne,...) en essayant d'employer différents types de seuillage.

Les tests d'évaluation sont faits sur une base de données parlée universelle, appelée ORATOR. Celle-ci a été réalisée par des acteurs professionnels, parlant en Allemand. Elle comporte plusieurs acteurs exprimant différentes attitudes émotionnelles : tristesse, joie, colère, etc. ; ainsi que différents types de personnalités.

Les applications sont très diverses, dont on cite : la détection de mensonge dans les affaires judiciaires, la détection de fatigue chez les pilotes d'avion et l'estimation de l'attitude d'un interlocuteur parlant par téléphone (applications multimédia).

Dans le but d'évaluer la consistance de nos résultats (*consistency measure*), des tests de χ^2 avec mesure de probabilité de consistance liée à l'indépendance des variables de l'émotion, ont été entrepris sur chaque résultat obtenu. Ces tests ont permis de déceler les paramètres pertinents pour chaque type d'émotion.

Par ailleurs, nous avons divisé l'analyse des émotions en 2 types : analyse des locuteurs masculins et analyse des locuteurs féminins (*locutrices*). Cette séparation nous a montré que les hommes ont un comportement différent des femmes quant à l'expression des émotions et de leur détection.

Finalement, cette étude nous a éclairé sur différents moyens et techniques pouvant détecter l'émotion et la personnalité du locuteur. Les taux de reconnaissance obtenus dépassent les 80% de bonne identification dans beaucoup de cas.

*Directeur de thèse : Dr H. SAYOUD, Professeur à l'USTHB.

Table des matières

	Pages
INTRODUCTION GENERALE	1
 Chapitre I : GENERALITES SUR LE SIGNAL PAROLE	
I.1. Le signal de la parole	6
I.2. Description du signal parole	7
I.3. prosodie	8
I.4. La structure prosodique	8
I.5. Le pitch et la fréquence fondamentale.....	9
I.6. L'énergie	9
I.7. Les descripteurs de qualité de voix	9
I.8. Débit de parole	11
I.8. Conclusion	11
 Chapitre II: DEFINITION ET IDENTIFICATION DE L'EMOTION	
II.1. Préliminaire	12
II.2. Notion d'émotion	12
II.2.1. Le phénomène émotionnel	12
II.2.2. Définition de l'émotion	13
II.3. Catégories d'émotions	13
II.4. Les cinq caractéristiques fondamentales d'émotions	13
II.5. Types d'émotions	14

II.5.1. L'émotion vécue	15
II.5.2. L'émotion induite	16
II.5.3. L'émotion simulée	16
II.6. Etat de l'Art sur la Reconnaissance Automatique de l'Emotion (RAE)	17
II.6.1. Modèles Psychologiques des Emotions	18
II.6.2. Critère de Reconnaissance de l'Emotion	18
II.6.3. Travaux Effectués selon les Quatre Critères	19
II.7. Conclusion	26

Chapitre III. CARACTERISTIQUES PROPOSEES POUR L'IDENTIFICATION DE L'EMOTION

III.1. Introduction	27
III.2. Sélection des Caractéristiques	28
III.3. Méthodes de Sélection Citées dans la Littérature	28
III.4. Fonctions d'Evaluation	30
III.5. Nos Paramètres Prosodiques Proposés	31
III.6. Extraction des Caractéristiques	37
III.7. Conclusion	38

Chapitre IV: BASES DE DONNEES D'EVALUATION

IV .1 Introduction	39
IV.2. Base de Données ORATOR	39
IV.3. Présentation de notre base de données universelle	40

IV.4. Evaluation de degré émotionnel	40
IV.5. Influence de l'information lexicale sur l'identification émotionnelle	40
IV.6. Répartition des données en fonction du sexe des locuteurs	41
IV.7. Conclusion	44

Chapitre V: RESULTATS ET DISCUSSIONS

V.1. Introduction	45
V.2. Représentation Synoptique	46
V.3. Fréquence Fondamentale et le Contour de Pitch	47
V.4. Plages de variation des paramètres	53
V.5. Fausse Acceptation et le Faux Rejet	54
V.6. Taux d'erreur égale EER.....	55
V.7. Estimation du seuil EER	56
V.8. Problématique	58
V.9. Récapitulatif des seuils correspondant aux EER	60
V.10 Récapitulatif des taux de reconnaissance maximaux et d'erreur EER	63
V.11. Interprétation et Discussion des Résultats	66
V.12. Mesure et Test de Signification	71
V.12.1 Mesure de Dépendance entre les variables	71
V.12.2 Mesure de dépendance	71
V.12.3. Représentation des deux variables	71
V.12.4 Méthode de calcul de la valeur globale de la χ^2	72

Tables de matières

V.12.5. Calcul de la probabilité P	73
V.12.6. Estimation de la dépendance	73
V.12.7 Récapitulatif des résultats de la probabilité P et de la χ^2	97
V.13. Rapprochement des résultats d'identification émotionnelle et ceux de la dépendance	81
V.14. Conclusion	82
Chapitre VI: CONCLUSION GENERALE	83
REFERENCES BIBLIOGRAPHIQUES	84

Liste des figures

	pages
Figure 1 : Modèle simple de production de la parole	6
Figure 2.1 : Types d'unités d'analyse étudiés dans le domaine de la RAE ...	19
Figure 2.2 : Approche pseudo-phonétique pour la reconnaissance automatique des émotions	22
Figure 3.1 : Principe d'extraction des paramètres prosodiques	32
Figure 5.1 : Schéma synoptique représentant les différentes étapes	46
Figure 5.2 : Contour du pitch d'un signal de parole de locuteur N°13 pour l'émotion agitation	47
Figure 5.3 : Contour du pitch d'un signal de parole de locutrice N°7 pour l'émotion agitation	48
Figure 5.4 : Contour du pitch d'un signal de parole de locuteur N°69 pour l'émotion colère	48
Figure 5.5 : Contour du pitch d'un signal de parole de locutrice N°6 pour l'émotion colère	49
Figure 5.6 : Contour du pitch d'un signal de parole de locuteur N°80 pour l'émotion confiance	49
Figure 5.7 : Contour du pitch d'un signal de parole de locutrice N°56 parole de pour l'émotion confiance	50
Figure 5.8 : Contour du pitch d'un signal de parole de locuteur N°113 pour l'émotion force (de caractère)	50
Figure 5.9 : Contour du pitch d'un signal de parole de locutrice N°7 pour l'émotion force (de caractère)	51

Figure 5.10 : Taux d'erreur égal (EER) pour l'émotion: colère, avec le paramètre 'P1', 'sexe : (M)'	56
Figure 5.11 : Estimation du seuil EER l'émotion : colère avec le paramètre 'P1 (variation du signal)', sexe : (M)	59

Liste des tableaux

	pages
Tableau 4.1 : Répartition des données en fonction du sexe des locuteurs	41
Tableau 4.2 : Répartition de l'émotion " agitation " en fonction du sexe des locuteurs	41
Tableau 4.3 : Répartition de l'émotion " colère " en fonction du sexe des locuteurs.....	42
Tableau 4.4 : Répartition de l'émotion" confiance " en fonction du sexe des locuteurs... ..	42
Tableau 4.5 : Répartition de l'émotion" fort (de caractère) " en fonction du sexe des locuteurs.....	43
Tableau 5.1 : Répartition des plages de variation des paramètres en fonction du sexe des locuteurs.....	53
Tableau 5.2 : Seuils EER obtenus pour les locuteurs masculins	60
Tableau 5.3 : Seuils EER obtenus pour les locuteurs féminins	61
Tableau 5.4 : Taux d'erreur EER et taux de reconnaissances maximaux obtenus, pour chaque émotion dans le cas locuteurs masculins	63
Tableau 5.5 : Taux d'erreur EER et taux de reconnaissances maximaux obtenus, pour chaque émotion dans le cas locuteurs féminin.....	65
Tableau 5.6 : Représentant les deux variables	72
Tableau 5.7 : Tableau de contingence pour l'émotion agitation et le paramètre P1, ainsi que les valeurs de χ^2 et P (sexe : M).....	73
Tableau 5.8 : Tableau de contingence pour l'émotion agitation et le paramètre P1, ainsi que les valeurs de χ^2 et P (sexe : F).....	73
Tableau 5.9 : Tableau de contingence pour l'émotion agitation et le paramètre P2, ainsi que les valeurs de χ^2 et P (sexe : M).....	74
Tableau 5.10 : Tableau de contingence pour l'émotion agitation et le paramètre P2, ainsi que les valeurs de χ^2 et P (sexe : F)	74
Tableau 5.11 : Tableau de contingence pour l'émotion agitation et le paramètre P3, ainsi que les valeurs de χ^2 et P (sexe : M).....	74
Tableau 5.12. : Tableau de contingence pour l'émotion agitation et le paramètre P3, ainsi que les valeurs de χ^2 et P (sexe : F)	74
Tableau 5.13. : Tableau de contingence pour l'émotion agitation et le paramètre P4, ainsi que les valeurs de χ^2 et P (sexe : M).....	75

Tableau 5.14. : Tableau de contingence pour l'émotion agitation et le paramètre P4, ainsi que les valeurs de χ^2 et P (sexe : F)	75
Tableau 5.15. : Tableau de contingence pour l'émotion agitation et le paramètre P5, ainsi que les valeurs de χ^2 et P (sexe : M)	75
Tableau 5.16. : Tableau de contingence pour l'émotion agitation et le paramètre P5, ainsi que les valeurs de χ^2 et P (sexe : F).....	75
Tableau 5.17. : Tableau de contingence pour l'émotion agitation et le paramètre P6, ainsi que les valeurs de χ^2 et P (sexe : M)	76
Tableau 5.18. : Tableau de contingence pour l'émotion agitation et le paramètre P6, ainsi que les valeurs de χ^2 et P (sexe : F).....	76
Tableau 5.19. : Tableau de contingence pour l'émotion agitation et le paramètre P7, ainsi que les valeurs de χ^2 et P (sexe : M)	76
Tableau 5.20. : Tableau de contingence pour l'émotion agitation et le paramètre P7, ainsi que les valeurs de χ^2 et P (sexe : F)	76
Tableau 5.21. : Tableau de contingence pour l'émotion agitation et le paramètre P8, ainsi que les valeurs de χ^2 et P (sexe : M)	77
Tableau 5.22. : Tableau de contingence pour l'émotion agitation et le paramètre P8, ainsi que les valeurs de χ^2 et P (sexe : F)	77
Tableau 5.23. : Tableau de contingence pour l'émotion agitation et le paramètre P9, ainsi que les valeurs de χ^2 et P (sexe : M)	77
Tableau 5.24. : Tableau de contingence pour l'émotion agitation et le paramètre P9, ainsi que les valeurs de χ^2 et P (sexe : F)	77
Tableau 5.25. : Tableau de contingence pour l'émotion agitation et le paramètre P10, ainsi que les valeurs de χ^2 et P (sexe : M)	78
Tableau 5.26. : Tableau de contingence pour l'émotion agitation et le paramètre P10, ainsi que les valeurs de χ^2 et P (sexe : F)	78
Tableau 5.27. : Tableau de contingence pour l'émotion agitation et le paramètre P11, ainsi que les valeurs de χ^2 et P (sexe: M)	78
Tableau 5.28. : Tableau de contingence pour l'émotion agitation et le paramètre P11, ainsi que les valeurs de χ^2 et P (sexe: F)	78
Tableau 5.29 : Récapitulatif des résultats la probabilité P et χ^2 pour le sexe masculin ...	79
Tableau 5.30 : Récapitulatif des résultats la probabilité P et χ^2 pour le sexe féminin ...	80

Liste des abréviations

ERR	Equal Error Rate (taux d'erreur égal).
Taux reco max	taux de reconnaissance maximal d'émotions bien reconnues .
MD	taux de détections manquées des émotions.
FA	taux de fausses alarmes.
P1	variation de la fréquence fondamentale F_0 .
P2	maximum de la dérivée de la fréquence fondamentale F_0 .
P3	minimum de la dérivée de la fréquence fondamentale F_0 .
P4	puissance du signal de parole.
P5	maximum de puissance d'un signal de parole
P6	variation du signal de parole.
P7	perturbation de la fréquence fondamentale "jitter".
P8	nombre de pauses.
P9	durée totale des pauses.
P10	durée maximale des pauses d'un signal de parole.
P11	moyenne des pauses d'un signal de parole.
P	probabilité d'indépendance.
χ^2	Contingence entre deux variables
F0	Fréquence fondamentale (pitch)
EDA	Estimation of Distribution Algorithm
CES	Best-First Search and Correlation-based Feature Selection
VAL	Vérification Automatique du locuteur
RAE	Reconnaissance Automatique des Emotions
ICT	Information à Court Terme
ILT	Information à Long Terme
MFCC	Mel Frequency Cepstral Coefficients.

Liste des abréviations

LPC	Linear Predictive Coefficients
LPCC	Linear Predictive Cepstral Coefficients
LDC	Linguistic Data Consortium Emotional Prosody
LFPC	Log Frequency Power Coefficients
PLP	Perceptual Linear Prediction
HMM	Hidden Markov Model
GMM	Gaussian Mixture Model
SVM	Machine à Vecteur de Support



Introduction Générale

Introduction Générale

La parole est un moyen de communication extrêmement riche et complexe. Elle véhicule non seulement l'information linguistique mais aussi, entre autres, l'information sur la personnalité et sur l'état émotionnel du locuteur.

La manifestation des émotions est un domaine particulièrement complexe de la communication humaine. Elle concerne des sciences pluridisciplinaires telles que la psychologie, la cognition, la sociologie, la philosophie. Cette pluridisciplinarité, qui est due à la haute variabilité du comportement humain, influence à la fois la production et la perception de ces émotions.

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 (Chung, 2000).

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 émotionnelle (émotion vocale). Ce fait peut être attribué au formalisme des sciences modernes et à la difficulté méthodologique de l'étude de l'émotion vocale.

Les chercheurs considéraient la variation vocale due à l'émotion comme une variable aléatoire, non-systématique, qui ne mérite pas d'analyse scientifique. 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.

La plupart des systèmes de la synthèse et de la reconnaissance vocale ont été développés sans considération de la variation émotionnelle.

Récemment, les chercheurs réalisent de plus en plus l'importance de la connaissance sur la parole émotionnelle pour le développement des recherches dans les domaines pratique et 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.

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 technique de la parole, l'ajout des traits personnels et émotionnels est essentiel pour augmenter le caractère naturel de la parole synthétique.

Les reconnaissances automatiques de la parole et du locuteur pourront largement bénéficier du développement du système qui peut reconnaître la parole et le locuteur dans leurs différents états émotionnels.

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. Ces données sont censées être représentatives de celles qu'on éprouve dans la vie quotidienne (Chung, 2000).

Dans le cadre du développement des technologies de la parole, il est apparu qu'il est nécessaire de prendre en compte les modifications vocales associées aux émotions et aux attitudes des locuteurs afin d'améliorer d'une part, la performance des systèmes de reconnaissance de la parole et du locuteur et d'autre part, l'acceptabilité des systèmes de synthèse vocale. L'information sur l'état émotionnel du locuteur, véhiculée par le signal acoustique dans un contexte réel, est utilisée pour :

- La détection des problèmes de compréhension entre un locuteur et un système de dialogue homme-machine ;

- La détection automatique des appels problématiques de centres d'appels téléphoniques ; qui permet de prendre en charge la relation client à distance (dans le cas de la formation, de la consultation, etc ...); afin de répondre plus rapidement et plus économiquement aux requêtes des clients, et éviter leur colère etc... À titre d'exemple, un système de détection des émotions peut servir au développement de systèmes à interaction

homme-machine efficace, naturelle et sensible au comportement de l'utilisateur. Utilisé dans un contexte d'enseignement à distance, un tel système serait capable de savoir si l'utilisateur est ennuyé, découragé par la matière enseignée, il pourra donc changer le style et le niveau de la matière concernée, fournir une compensation et un encouragement émotionnel ou accorder une pause à l'utilisateur (Li, 2007) (Zhu, 2007) :

- le savoir conserver la satisfaction des clients qui n'ont pas réussi à exécuter la tâche désirée (faisant l'objet de leur appel); ou qui l'ont exécuté difficilement ;

- la détection des mensonges dans les enquêtes policières ;

- la détection précoce d'un appel problématique d'un client en colère permettrait à la machine d'entreprendre plusieurs stratégies de gestion de l'échec de l'appel, parmi celles-ci :

1. restreindre et guider le dialogue ;

2. s'excuser et apaiser l'usager par une voix douce et des remarques plaisantes ;

3. traiter l'appel en priorité en le redirigeant vers un opérateur humain dans le cas où le problème persiste ;

4. détecter les bris de communication afin de rappeler les clients mécontents qui, découragés par la tournure du dialogue, y ont mis fin prématurément ;

5. détecter la colère, qui est un avertisseur qui renseigne sur l'état d'un client faisant face à des difficultés induites par l'utilisation du système; afin de l'éloigner du danger (cas des usines, des zones industrielles, etc ...).

La reconnaissance automatique des émotions peut être utilisée dans beaucoup d'autres contextes applicatifs, elle peut :

• servir à détecter la fatigue et l'influence de l'alcool chez un conducteur automobile afin d'activer des routines de sécurité (Schuller, 2002);

• permettre au système équipant la voiture de fournir au conducteur un support et un encouragement au cours d'une pénible expérience de conduite (Jones, 2007);

• détecter la présence d'émotions extrêmes, principalement la peur, dans le cadre de la surveillance dans les lieux publics (Clavel, 2007);

- prioriser automatiquement les messages cumulés dans la boîte vocale selon différents axes affectifs tels que l'urgence, la valence (heureux ou triste) et l'excitation (calme ou excité) pour alerter le propriétaire du compte et lui permettre d'écouter les messages importants en premier (Inanoglu, 2005);
- utiliser les traits spéciaux véhiculés par les émotions pour le développement de systèmes de vérification automatique du locuteur (VAL) plus robustes et précis (Panât, 2008);
- évaluer l'urgence d'un appel afin de prendre une décision, dans le cadre d'un centre d'appel médical offrant un service de conseils médicaux aux patients (Devillers, 2007);
- améliorer le service à la clientèle

Objectif de ce Mémoire

Afin de contribuer à la résolution de ces problèmes, le présent travail consiste en une série d'études statistiques prosodiques des énoncés émotionnels. Cette étude vise à démontrer :

- 1 - Comment l'émotion du locuteur est-elle exprimée dans sa parole ?
- 2 - Quels sont les meilleurs indices prosodiques pour repérer l'excitation émotionnelle?

Notre algorithme permet de mettre au point la détection de ces états émotionnels, uniquement par analyse du signal vocal de l'intéressé. Il s'agit d'identifier certaines émotions d'un locuteur et certains traits de sa personnalité à partir des caractéristiques de sa parole. Pour ce faire, nous avons essayé plusieurs paramètres prosodiques : Pitch, Durée, Energie, ... et leurs combinés statistiques (maximum, minimum, variation, moyenne,...) en essayant d'employer différents types de seuillage.

Les tests d'évaluation sont faits sur une base de données parlée universelle, appelée ORATOR contenant 150 extraits de parole émotionnée. Celle-ci a été réalisée par des acteurs professionnels, parlant en Allemand. Elle comporte des paroles exprimant différentes attitudes émotionnelles : tristesse, joie, colère, etc. ; ainsi que différents types de personnalités.

Nous nous sommes intéressés exclusivement aux quatre émotions exprimant l'agitation, la colère, fort (de caractère) et la confiance, vu leur importance dans, la vie quotidienne, le milieu industriel et d'autres domaines.

Dans le but d'évaluer la consistance de nos résultats, des calculs de la χ^2 avec mesure de probabilité de consistance liée à l'indépendance des variables de l'émotion, ont été entrepris pour chaque résultat obtenu. Les résultats de ces calculs ont permis de déceler les paramètres pertinents pour chaque type d'émotion.

Par ailleurs, nous avons divisé l'analyse des émotions en 2 types : analyse des locuteurs masculins et analyse des locuteurs féminins (locutrices). Cette séparation nous a montré que les hommes ont un comportement différent des femmes quant à l'expression des émotions et de leur détection.

Finalement, cette étude nous a éclairé sur différents moyens et techniques pouvant détecter l'émotion et la personnalité du locuteur. Les taux de reconnaissance obtenus dépassent les 80% de bonne identification dans beaucoup de cas.

Plan du Mémoire

Le présent mémoire se compose d'une introduction générale suivie de **cinq chapitres** :

- **Le premier chapitre** traite des généralités sur le signal parole et les caractéristiques de la prosodie ;
- **Le deuxième chapitre** cite quelques notions théoriques sur l'émotion, suivies d'une discussion concernant les trois types d'émotions, et d'un aperçu des études précédentes dans ce domaine de recherche ;
- **Le troisième chapitre** concerne les paramètres prosodiques proposés ;
- **Le quatrième chapitre** présente la base de données d'évaluation, 'ORATOR' ;
- **Le cinquième chapitre** étale les expériences, les résultats et les discussions ; suivis d'une étude comparative avec les travaux précédents.

Et nous terminons par une conclusion générale et des perspectives.

Chapitre I : Généralités sur le Signal de Parole

Chapitre I. Généralités sur le Signal de Parole

I.1. Signal de Parole

Le signal de parole n'est pas un signal ordinaire. Il est le vecteur d'un phénomène complexe de la communication parlée. Il contient plusieurs types d'informations. En premier lieu, le signal de parole véhicule un message linguistique qui sert à la communication entre individus. Mais il transporte également des informations sur l'identité et l'état émotionnel de l'individu ayant prononcé le message. Les humains se servent de ces informations pour identifier les personnes qu'ils connaissent et leurs états émotionnels, même lorsqu'ils ne peuvent pas voir leur interlocuteur, au téléphone par exemple.

D'un point de vue mathématique, il est difficile de modéliser le signal de parole, compte tenu de sa variabilité.

Le signal de parole $s(n)$ est le résultat de l'excitation $u(n)$ du conduit vocal par un train d'impulsion ou un bruit avec un gain G donnant lieu respectivement aux sons voisés et non voisés (la figure 1). Dans le cas des sons voisés, l'excitation est une vibration périodique des cordes vocales suite à la pression exercée par l'air provenant de l'appareil respiratoire (Triki, 2007).

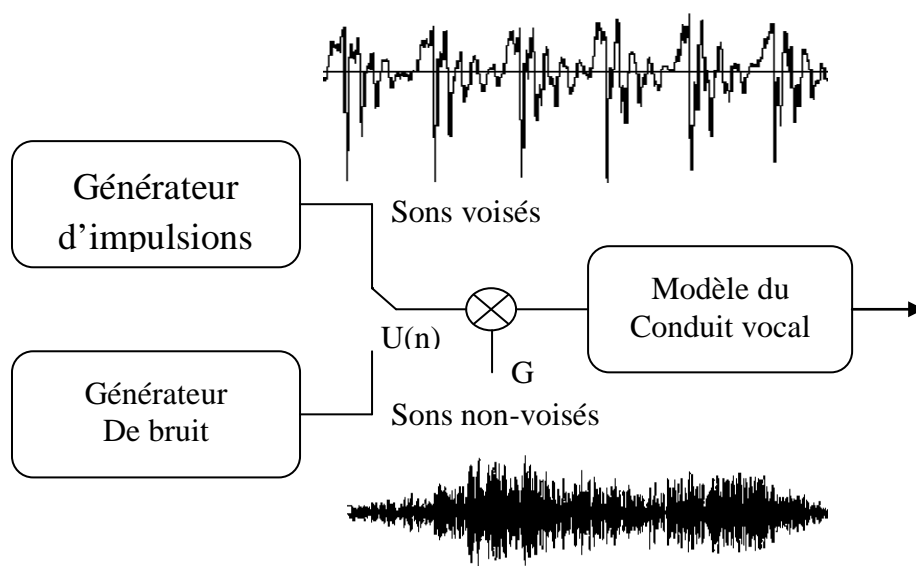


Figure 1 : Modèle simple de production de la parole (Triki, 2007).

Ce mouvement vibratoire correspond à une succession de cycles d'ouverture et de fermeture de la glotte. Le nombre de ces cycles par seconde correspond à la fréquence fondamentale F_0 . Quant aux signaux non-voisés, l'air passe librement à travers la glotte (du moins pas dans tout le conduit vocal) sans provoquer de vibration des cordes vocales.

La fréquence fondamentale du signal de parole peut varier selon le genre (masculin ou féminin) et l'âge du locuteur:

- de 80 à 200 Hz pour une voix masculine ;
- de 150 à 450 Hz pour une voix féminine ;
- de 200 à 600 Hz pour une voix d'enfant (Dutoit, 2005).

I.2. Description du Signal de Parole

▪ Redondance

Le signal de parole est extrêmement redondant. Cette grande redondance lui confère une robustesse à certains bruits.

▪ Variabilité intra-locuteur

Elle identifie les différences dans le signal produit par la même personne.

Cette variation peut résulter de :

- l'état physique (maladie des voies respiratoires) ;
- l'état moral (l'humeur ou l'émotion du locuteur).

▪ Variabilité interlocuteur

La cause principale des différences entre interlocuteurs est de nature physiologique (Triki, 2007) :

- Longueur de la corde vocale ;
- Forme et longueur du conduit vocal ;
- Genre (homme, femme) ;
- Age.

- **Stationnarité**

Le signal de parole n'est jamais rigoureusement stationnaire ce qui explique sa complexité ; afin de simplifier son étude on doit le considérer comme étant localement stationnaire (quasi-stationnaire) (Beller, 2009).

I.3. Prosodie

En parole, nous avons un précieux dicton: « isn't what you said; it's how you said it! » (Huang, Acero, 2001). Qui veut dire “la prosodie n'est pas ce que vous avez dit, c'est la façon dont vous l'avez dit ».

La prosodie est un canal parallèle au contenu sémantique du message parlé, dans les conversations quotidiennes, à travers lequel l'auditeur peut percevoir les intentions et l'état émotionnel du locuteur.

C'est à travers la prosodie également que le locuteur peut donner à l'énoncé un ton de déclaration, d'une question, ou d'une commande

L'intonation joue des rôles multiples dans le langage de tous les jours. Elle distingue une question d'une réponse. Elle désambigüise des séquences telles que « Je ne veux pas mourir traître » (la prosodie doit préciser lequel des deux est traître, le locuteur ou l'interlocuteur. Elle exprime des attitudes, des émotions, etc.) (Huang, 2001).

« Un accent expressif ou émotif renseigne sur l'état d'esprit du locuteur. Il sert ainsi à transmettre des sentiments d'impatience, de colère, de doute, d'incertitude, d'amour ou de haine, etc. C'est le domaine de l'expression des émotions » (Galarneau, web)

Les prononciations d'un même mot peuvent avoir des prosodies différentes sans affecter l'identité du mot: les phones peuvent être longs ou courts, forts ou doux et avoir des fréquences fondamentales variées.

I.4. Structure Prosodique

La structure prosodique s'intéresse à la relation qui lie la durée, l'énergie et le pitch au son.

Perceptivement, l'énergie et son évolution, et le débit, mais aussi les pauses et les silences, permettent la compréhension d'informations au-delà des mots prononcés.

I.5. Le pitch (la fréquence fondamentale)

Le pitch est le phénomène prosodique le plus expressif. En parlant, nous varions systématiquement notre fréquence fondamentale pour exprimer nos sentiments ou pour diriger l'attention de l'auditeur vers un aspect important de notre message parlé. Un paragraphe prononcé avec un pitch constant et uniforme paraîtra peu naturel (Huang, 2001).

Le pitch caractérise les parties voisées du signal de parole et est liée à la sensation de hauteur de la voix (aigüe ou grave).

Les parties voisées ont une structure pseudo-périodique et sur ces portions, le signal est généralement modélisé, comme la somme d'un signal périodique (de période T) et d'un bruit blanc.

La fréquence fondamentale est l'inverse de la période T, ($F_0 = 1/T$) (Calliope, 1997).

I.6. L'énergie

L'énergie est un bon indice prosodique car elle est très dépendante du contexte phonétique. En effet, une voyelle est beaucoup plus énergétique qu'une plosive (présence d'un silence avant l'explosion). Ce paramètre permet de fournir une mesure de la force sonore de la voix (faible ou forte) en décibel (dB).

I.7. Descripteurs de Qualité de Voix

(Bulut, 2002) indique très clairement que la prosodie ne suffit pas à décrire les émotions, il manque la qualité vocale. En ce qui concerne les émotions, la qualité vocale est souvent traduite par des mots comme : « soufflé », « crispé », « détendu », qui découlent directement de l'état physiologique de la personne. La qualité de voix, commence à être utilisée pour caractériser les variations émotionnelles.

Les impulsions naturelles glottales ne sont pas réellement périodiques, mais présentent des perturbations appelées tremblement (Shimmer) et vacillement (Jitter). Elles sont entre autres la conséquence des modifications physiologiques provoquées par les changements émotionnels du locuteur.

- **Perturbation de Fo (vacillement) et perturbation d'intensité (tremblement)**

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.

Le vacillement est décrit comme une déviation de la fréquence fondamentale; il permet de modéliser ces oscillations autour de la fréquence fondamentale de la voix. Il représente les micro-variations du Fo trame par trame dans les périodes de Fo suivant la formule (1.1).

Le tremblement représente les variations de l'intensité cycle par cycle dans les périodes de l'énergie. Il modélise, la modulation d'amplitude suivant la formule (1.2).

Les voix normales ont un vacillement de 0.5 à 1.0% et un tremblement de l'ordre de 0.04 à 0.21%, ce qui représente un niveau assez bas pour qu'il soit directement perceptible. Bien que le tremblement et le vacillement soient deux concepts différents, ils sont légèrement corrélés (Huang, 2001).

$$\text{Vacillement (\%)} = \frac{100N \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{(N-1) \sum_{i=1}^N T_i} \dots\dots\dots(1.1)$$

Où T_i , représente l'inverse de la fréquence fondamentale de la $i^{\text{ème}}$ trame et N le nombre total de trames voisées dans l'énoncé (Boersma, 2005).

$$\text{Tremblement (\%)} = \frac{100N \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{(N-1) \sum_{i=1}^N A_i} \dots\dots\dots(1.2)$$

Où A_i , représente la valeur du sommet de l'amplitude de la $i^{\text{ème}}$ trame et N le nombre total de trames voisées dans l'énoncé (Boersma, 2005).

I.8. Débit de Parole

Contrairement à la fréquence fondamentale, le débit est une notion difficilement mesurable, car il n'y a pas de consensus au niveau de sa définition.

On distingue trois types de débits réguliers (lent, moyen, rapide) et deux types de changements de débit (accélération et ralentissement).

Le débit est défini par (Yacoub, 2003) par la durée de segments audibles: Si une portion de signal est, d'énergie plus faible que 1% de son maximum, alors cette portion de signal est considérée comme inaudible. Les segments inaudibles sont ainsi reliés aux pauses silencieuses. Cette définition est préconisée par (Bänziger, 2004) qui a dû passer par une segmentation manuelle au préalable (étape assez longue et fastidieuse qui ne permet pas l'étude de vastes corpus).

I.9. Conclusion

Dans ce chapitre nous avons défini quelques notions sur les caractéristiques du signal de parole ainsi que sur celles de la prosodie : le pitch, l'énergie, les perturbateurs de fréquence et d'intensité, etc ...

Dans le chapitre suivant nous allons présenter quelques notions théoriques sur l'émotion et nous présenterons un éventail d'études, en fonction de quatre critères : le type d'unité d'analyse utilisé : énoncé, phonème etc ...; la nature de l'information para-linguistique employée comme traits caractéristiques qui peut être soit de type prosodique ou spectrale ; la portée temporelle des traits utilisée, qui peut être soit de type à court terme ou à long terme ; le type d'approche choisi pour concevoir le classifieur (dynamique, statique, floue).

Chapitre II : Définition et Identification de l'Emotion.

Chapitre II. Définition et Identification de l'Emotion

II.1. Préliminaire

Les systèmes de reconnaissance automatique de la parole donnent à la machine les capacités de transformer le signal sonore en une suite de mots. Le domaine du traitement automatique du langage permet d'accéder au sens de cette suite de mots.

Partant de ces outils relativement efficaces, il est nécessaire d'aller plus loin : la question n'est plus uniquement de savoir ce qui est dit mais aussi de connaître le contexte de prononciation de la phrase. C'est à ce niveau qu'intervient la dimension émotionnelle.

Si on ne prend pas en compte l'intonation de la phrase, il est difficile de faire la différence entre une question et une affirmation. De la même façon, selon l'émotion, et la latitude, mais aussi selon la personnalité du locuteur, une même phrase peut avoir un sens différent.

Dans le courrier électronique et les messages des utilisateurs de groupes de discussion, les émoticônes sont la représentation graphique d'un visage humain exprimant le sourire, l'étonnement, la colère, etc...; et permettent de nuancer le sens d'un texte écrit souvent rapidement, dans un langage proche du langage parlé.

II.2. Notion d'Emotion

II.2.1. Phénomène Emotionnel

En quoi consiste le phénomène émotionnel ?

La réponse à cette question est un sujet de discussion sur lequel se sont penchés de nombreux psychologues. L'article de (Kleinginna, 1981) retrace les 100 années d'études faites par la communauté des psychologues et conclut qu'il est difficile d'aboutir à un consensus sur la définition du phénomène émotionnel.

Les distinctions les plus couramment utilisées dans la littérature pour circonscrire le sens du terme émotion sont : (émotions primaires et secondaires (complexe)) ou (émotions basiques et dérivées).

Le nombre d'émotions basiques varie d'une étude à l'autre. La plupart du temps les émotions basiques incluent la peur, la colère, la joie et la tristesse.

Plutchik postule que les émotions plus complexes (mixtes) seraient des états dérivés ou mixés, et apparaîtraient comme des combinaisons, des mélanges ou composés des émotions primaires (Plutchik, 1980).

II.2.2. Définitions de l'Emotion

L'émotion est un sentiment intense qui se produit sous l'influence d'un événement extérieur perturbateur, par exemple : la réaction à une situation de stress, d'agression verbale ou de violence peut engendrer des états affectifs dont l'intensité varie d'un individu à l'autre.

L'émotion est au cœur de l'être humain, elle est au centre de toute relation et la complexité des rapports humains vient également du fait que nous ne sommes pas égaux face aux émotions (Attabi, 2008).

II.3. Catégories d'Emotions

Deux types de catégories d'émotions existent:

- 1- Négatives** : catégorie générale d'émotions désagréables, liées au mal-être (colère, tristesse et peur).
- 2- Positives** : catégorie générale d'émotions agréables, liées au bien-être (amour, joie).

II.4. Les cinq caractéristiques fondamentales d'émotions

Selon Campos et Barrett :

Il existe une continuité des émotions au cours de la vie, exemple : plusieurs situations peuvent créer la peur chez l'enfant comme chez l'adulte. Les motifs changent, ainsi que les comportements servant à exprimer cette émotion qui présente des caractéristiques de base invariantes.

Le développement entraîne un changement dans les relations entre l'expression émotionnelle et l'expérience émotionnelle. Exemple, au début de la vie, l'enfant sourit mais sans vraiment éprouver de la joie; plus tard, son sourire sera le reflet de la manifestation de la joie.

L'enfant développe de nouveaux moyens pour faire face aux émotions qui vont modifier ou contrer le sentiment lui-même. Exemple, un enfant qui est autonome dans ses déplacements pourra, s'il le veut contrer (affronter) une personne étrangère venant vers lui et dont la présence est dérangeante pour lui-même.

La complexité des émotions change au cours du développement surtout sur le plan cognitif. L'enfant devient capable d'évaluer les multiples aspects d'une situation en fonction de critères sociaux et moraux. Par exemple on verra des émotions telles que la culpabilité et la fierté apparaître à ce moment.

La réaction aux émotions d'autrui change au cours du développement. L'enfant, à ce moment, devient capable de réagir émotionnellement en accord avec l'expression d'autrui (gaieté, peur sur le visage d'une autre personne) (Campos, 1984).

II.5. Types d'Emotions

Afin de mettre en valeur les données de notre travail, qui sont des extraits de la parole émotionnelle simulée, une discussion sur les différents types d'émotion est indispensable.

Dans l'étude de l'émotion, trois types de données peuvent être distingués selon la manière de production :

- 1- les émotions vécues : exprimées par des locuteurs qui se trouvent réellement dans de tels états émotionnels ;
- 2- les émotions simulées : produites par des acteurs qui imitent de tels états émotionnels ;
- 3- les émotions induites : provoquées expérimentalement, par des moyens artificiels.

Un des obstacles majeurs dans l'étude des émotions vécues est le problème méthodologique de l'expérimentation, car ce 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 la notion d'émotion simulée.

Les émotions vécues sont provoquées par des expériences réelles et se produisent dans des situations naturelles, dans des milieux quotidiens, d'où vient la parole spontanée, tandis que les émotions simulées sont exprimées par des acteurs professionnels ou non professionnels et sont faites dans des cadres artificiels, tels que le théâtre et le laboratoire.

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; car le processus de l'expression de l'émotion vécue est enregistré avec le moins d'intervention possible. Mais, l'expression de l'émotion simulée est généralement sous le contrôle du locuteur.

Vu la difficulté de l'acquisition et de l'analyse des données spontanées, qui exclue la possibilité d'avoir les mêmes phrases à travers différentes conditions, ce qui est faisable avec l'émotion simulée, les chercheurs analysent souvent des données répétées et enregistrées dans un laboratoire. Ils les considèrent comme représentatives de la parole ou de l'émotion exprimée dans un milieu naturel.

II.5.1. Emotion Vécue

Les données des émotions vécues reflètent les meilleures caractéristiques de l'émotion naturelle, leur utilisation est assez limitée par des problèmes méthodologiques; car l'acquisition et l'expérimentation de ce genre de données sont extrêmement difficiles. En plus des difficultés d'enregistrement en situation naturelle, celui-ci est souvent de courte durée, très limité en nombre de locuteurs, de mauvaise qualité.

Parmi les émotions vécues, celles enregistrées en cas de catastrophe, (par exemple, la chute d'un avion ou un 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.

II.5.2. Emotion Induite

Diverses méthodes sont inventées, mais elles sont souvent problématiques du point de vue psychologique, elles sont produites par deux types de provocations :

- **La provocation attendue :** Souvent utilisée par les psychologues, dans un cadre de laboratoire. L'expérimentateur expose les locuteurs à une provocation par des moyens artificiels qui sont censés être sous un grand contrôle sur la production des émotions bien définies et une bonne acoustique d'enregistrement. Par exemple le suivi d'un documentaire vidéo concernant des séquences d'un tremblement de terre terrible.
- **La provocation inattendue :** Dans un milieu naturel, hors du laboratoire, et dont la procédure est la même que la provocation attendue, sauf que les personnes ne savent pas que leurs paroles sont enregistrées. Par exemple : un signal d'alarme présenté en plein cours, ce qui rend les étudiants anxieux et stressés; leurs voix émotionnelles sont enregistrées discrètement.

Les émotions vécues sont plus naturelles que les émotions provoquées par des stimuli, et elles reflètent mieux la communication des émotions naturelles.

II.5.3. Emotion Simulée

Souvent utilisée dans la recherche de l'expression émotionnelle, en raison de la facilité de la manipulation systématique des données.

Les chercheurs déterminent d'abord les émotions à étudier, et puis décrivent aux locuteurs professionnels le contexte de chaque émotion, qu'ils sont censés provoquer. Les acteurs imitent les différentes émotions, en modifiant des traits prosodiques de la voix. Le contenu

verbal est contrôlé dans le sens qu'il est, soit constant à travers les différentes émotions, ou différent.

Ce genre de données est d'un haut degré de contrôle mais peu spontané. Vu que l'enregistrement a lieu dans un studio insonorisé, dans lequel la qualité acoustique est généralement bonne, ce qui est essentiel pour les mesures exactes des traits prosodiques.

Certains chercheurs préfèrent les données d'émotions simulées par un locuteur non professionnel à celles simulées par un acteur professionnel. 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, 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 simulées 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 leur spontanéité est considérablement diminuée.

Les acteurs utilisent différentes techniques pour exprimer des émotions, comme le rappel des expériences personnelles, ce qui fait que les expressions émotionnelles varient largement d'un acteur à un autre (Chung, 2000).

II.6. Etat de l'Art sur la Reconnaissance Automatique de l'Emotion (RAE)

L'objectif de présenter un état de l'art dans le domaine de la reconnaissance automatique de l'émotion, est de bien connaître et comprendre les différentes évolutions techniques et technologiques dans ce domaine.

Un des problèmes majeurs rencontrés dans le domaine de la psychologie est l'absence d'un consensus autour de la définition de l'émotion et ses différents types.

II.6.1. Modèles Psychologiques des Emotions

Deux théories ont fortement déterminé le passé de la recherche dans le domaine des modèles émotionnels, à savoir, la théorie de l'émotion discrète et la théorie dimensionnelle. Un nouveau modèle d'émotion est apparu par la suite (Attabi, 2008).

- **Théorie de l'émotion discrète :** Les théoriciens proposent l'existence d'un nombre, compris entre 9 et 14, d'émotions de base ou fondamentales caractérisées par des modèles de réponse très spécifiques en physiologie, ainsi que dans les expressions faciales et vocales.

La plupart des études qui se sont intéressées à l'effet vocal des émotions ont utilisé ce modèle et ont choisi d'examiner les effets de la joie, de la tristesse, de la peur, de la colère et de la surprise (Attabi, 2008).

- **Théorie dimensionnelle :** Elle s'intéresse principalement à la description verbale des sentiments subjectifs. Les différents états émotionnels sont cartographiés dans un espace de deux dimensions principales consistant en la dimension de valence (agréable-désagréable) et en une dimension de l'activité (actif / passif).

Les travaux relatifs à l'effet de l'émotion sur la voix, de ce modèle se limitent souvent à l'étude des différences qui existent entre l'état émotionnel (positif et négatif) ou (actif et passif).

- **Modèle d'émotion à composant :** C'est un nouveau modèle d'émotion , son champ d'intérêt ne se limite pas à l'étude des sentiments subjectifs (telle que la théorie dimensionnelle) ni au nombre supposé d'émotions de base (comme c'est le cas de la théorie discrète). Ce modèle met l'accent sur la variabilité des différents états émotionnels. Il offre également la possibilité de modéliser les différences qui existent entre les membres de la même famille d'émotion, telle que la colère forte, la colère froide et le mépris.

II.6.2. Critères de Reconnaissance de l'Emotion

Une multitude de méthodes et de techniques ont été expérimentées pour aborder la RAE, suivant les quatre critères :

1. le type **d'unité** d'analyse utilisé pour la reconnaissance des émotions ;
2. la **nature de l'information** para-linguistique employée comme traits caractéristiques qui peut être de type prosodique ou spectrale ;
3. la **portée temporelle** des traits utilisée, qui peut être de type à court terme ou à long terme ;
4. le type **d'approche** choisi pour concevoir le classifieur (dynamique, statique, floue).

II.6.3. Travaux Effectués selon les Quatre Critères

▪ Travaux selon le type d'unité d'analyse

L'unité d'analyse représente le segment de données de base extrait de l'énoncé soumis au classifieur pour déterminer sa classe d'émotion. Dans les travaux réalisés à ce jour, six types d'unités ont été expérimentés (figure 2.1):

1- Enoncé

C'est une unité d'analyse sur laquelle est basée la quasi-totalité des travaux effectués. Les vecteurs de traits de la totalité de l'énoncé sont extraits et soumis en une seule entrée au classifieur pour déterminer la catégorie de l'émotion.

Parmi les travaux qui se sont basés sur cette unité, nous citons les travaux de El Ayadi (El Ayadi, 2007).

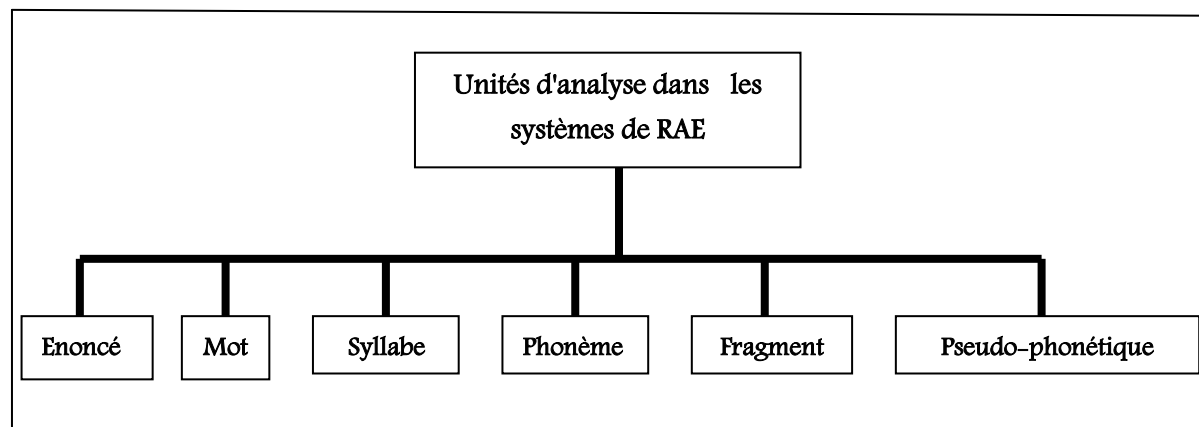


Figure 2.1: Types d'unités d'analyse étudiés dans le domaine de la RAE (Attabi, 2008).

2- Phonème

Le phonème représente la plus petite unité de son d'une langue. Le choix de cette unité est motivé par l'hypothèse que l'état émotionnel d'un locuteur affecte les phonèmes d'un énoncé avec des intensités très variables. L'énoncé est alors segmenté en phonèmes et chaque classe de phonèmes est modélisée séparément.

Afin de vérifier cette hypothèse, (Lee, 2004) a réalisé deux expériences :

- dans la première, un classificateur HMM (Hidden Markov Model ou modèles de Markov cachés) émotionnel générique est utilisé. Ce classificateur est entraîné en utilisant les données d'apprentissage de toutes les classes de phonèmes ;
- dans la deuxième expérience, des HMM par classe de phonèmes sont utilisés.

Le taux de reconnaissance obtenu avec le modèle HMM générique est de 64.77% alors que les résultats pour les modèles HMM par classe de phonèmes sont respectivement de 72.16%, 54.86%, 47.43%, 44.89% et 55.11% pour les classes voyelles, semi-voyelles, nasales, consonnes occlusives et fricatives.

Les résultats d'une classification basée sur la combinaison des modèles des cinq classes de phonèmes atteignent 75.57%.

Ces résultats montrent d'une part que les classes des phonèmes ne véhiculent pas, dans les mêmes proportions, la même charge émotionnelle. Les voyelles sont émotionnellement plus saillantes que les autres classes. D'autre part, une classification basée sur une modélisation par classe de phonèmes offre de meilleures performances que la classification à partir d'un modèle générique.

3- Syllabe

Pour la même raison ayant conduit à l'expérimentation de l'unité phonème, Schuller et ses collègues ont procédé à une segmentation de l'énoncé basée sur l'unité syllabe, nous citons les travaux de Schuller (Schuller, 2007b). Les performances du système basé sur l'unité syllabique sont inférieures à ceux du système basé sur l'énoncé.

4- Mot

L'unité d'analyse mot a été testée et comparée avec les performances d'un système basé sur l'unité énoncé, nous citons les travaux de Schuller (Schuller, 2007a). D'après les résultats obtenus, l'unité mot est préférable à l'énoncé à condition qu'un système de segmentation par mot efficace soit disponible.

5- Fragment

L'autre unité d'analyse expérimentée est le fragment. L'énoncé est segmenté automatiquement en fragments, en fonction des propriétés acoustiques de l'énoncé, nous citons les travaux de Schuller (Schuller, 2007b).

Les résultats obtenus sur le corpus de données utilisé montrent que les performances du système basé sur l'unité fragment sont meilleures que celles obtenues avec l'unité syllabe, mais qu'elles restent en dessous de celle de l'énoncé.

6- Pseudo-phonétique

C'est une unité similaire au phonème et est obtenue par une segmentation du signal de parole en zones stationnaires, combinée avec un détecteur de voyelles. Le processus de segmentation est indépendant de la langue et ne vise pas à l'identification exacte des phonèmes. Les segments obtenus lors de cette phase sont alors qualifiés d'unités pseudo-phonétiques (figure 2.2). Cette méthode a été introduite pour la reconnaissance automatique des langues.

Ces unités (pseudos-phonétiques) ont été exploitées dans la reconnaissance automatique d'émotions. Bien que la quantité d'informations portées par les segments 'voyelle' et 'consonne' soit bien inférieure à celles des segments voisés et non voisés, les résultats obtenus par ces unités sont supérieurs sur les plans acoustique et prosodique. Les segments phonétiques (ou pseudo-phonétiques) apparaissent plus appropriés que les segments voisés pour l'extraction d'informations.

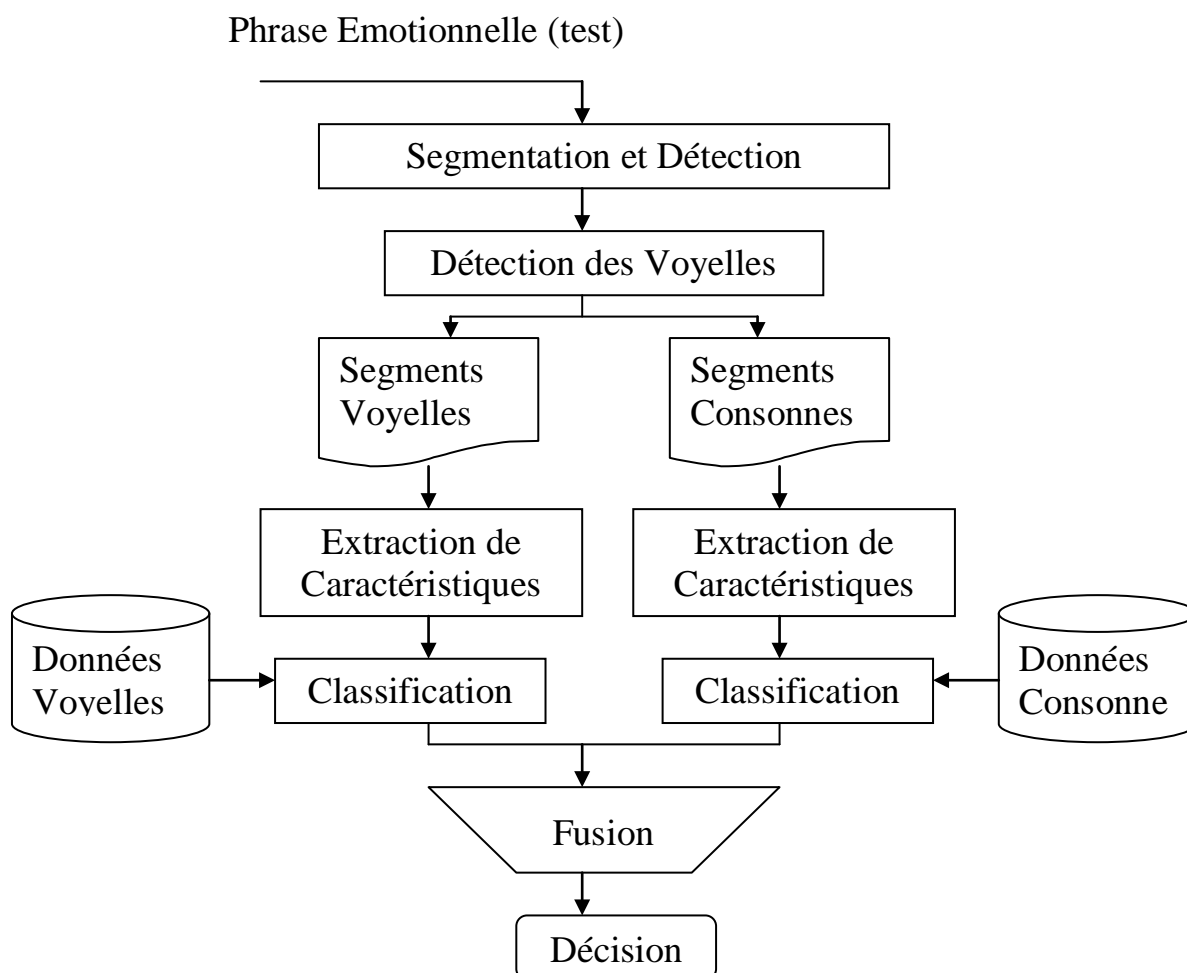


Figure 2.2 : Approche pseudo-phonétique pour la reconnaissance automatique des émotions (Ringeval, 2008).

Sur ce principe est basé le travail de la RAE qui à été réalisé par Ringeval et Chetouani. Ils proposent par conséquent d'intégrer les unités pseudo-phonétiques dans les systèmes de reconnaissance des émotions pour améliorer leur performance (Ringeval, 2008).

- **Travaux selon la portée des traits caractéristiques et la nature de l'information**

En se basant sur la portée de l'information utilisée comme traits caractéristiques. Cette information est classée dans l'une des deux classes : l'information à court terme ou l'information à long terme :

1. Information à Court Terme (ICT)

C'est une information locale qui s'étale généralement sur un intervalle de temps, appelé trame, allant de 10 ms à 30 ms, cadencé à chaque 10 ms. Chaque trame constitue un vecteur de traits caractéristiques. La séquence des vecteurs de l'énoncé véhicule sa structure temporelle.

Les deux types d'information à court terme utilisés dans les systèmes de la RAE : l'information spectrale, véhiculée en général par les coefficients spectraux, et l'information prosodique (temporelle) représentée par l'énergie, la fréquence fondamentale (F_0) etc ...

Le vecteur de traits à court terme le plus utilisé est composé des coefficients Mel-Frequency Cepstral Coefficients (MFCC), leurs dérivées premières et leurs dérivées secondes.

Parmi les travaux qui se sont basés sur ce type de vecteur, nous retrouvons nous citons les travaux de Gao (Gao, 2007; Pirker, 2007).

D'autres paramètres spectraux ont été également expérimentés, en combinaison avec les coefficients MFCC nous citons des travaux de Pao (Pao, 2005). Ces paramètres sont issus des techniques de compression de la parole telles que le codage par prédiction linéaire (Linear Predictive Coefficients, LPC), Linear Prédiction Cepstral Coefficients (LPCC), Log Frequency Power Coefficients (LFPC) et Perceptual Linear Prédiction (PLP).

L'information prosodique a été également utilisée en tant que traits à court terme, parfois seule (Huang, 2006; Sethu, 2007) ou combinée avec les coefficients cepstraux tels les MFCC et les LPCC, nous citons les travaux de Li (Li, 2007).

2. Information à Long Terme (ILT)

L'information à long terme caractérise l'énoncé dans sa globalité. Elle est représentée sous forme de valeurs statistiques pour une séquence temporelle de valeurs. Ce sont les variables de type prosodique qui sont les plus couramment utilisées, telles que la fréquence fondamentale, l'énergie, le débit de la parole, le nombre et la durée des silences, le rapport de la durée de la région voisée à la région non voisée et celles de type phonétique telles que les

formants et leurs bandes passantes, ainsi que des variations de paramètres, dont le tremblement (Shimmer) et le vacillement (Jitter), la moyenne, l'écart type, le maximum, le minimum, la déviation, la pente, le taux de passage par zéro sont des exemples de fonctions statistiques généralement utilisées.

Parmi les travaux qui se sont basés sur l'utilisation des statistiques de la prosodie, nous citons des travaux de Giripunje (Giripunje, 2007).

Certains chercheurs ont combiné, dans leurs vecteurs de traits de type à long terme, entre les statistiques de la prosodie et les statistiques des coefficients spectraux, qu'on retrouve par exemple dans les travaux de (Vlasenko, 2007). Rares sont les travaux où seules les statistiques des coefficients cepstraux sont utilisées comme traits caractéristiques à l'instar des travaux de Beritelli (Beritelli, 2007).

Les types de traits caractéristiques qu'on a cités jusqu'ici appartiennent tous à la catégorie d'information para-linguistique. Notons que certains auteurs ont introduit, dans leurs modélisations, l'information linguistique, notamment celle au niveau lexical et au niveau du discours (les actes de dialogue) en combinaison avec l'information para-linguistique (Devilleers, 2007). L'introduction de cette information linguistique a permis de réaliser un certain gain en performance, mais au coût de perdre la propriété d'indépendance du système de RAE de tout langage, qui est une caractéristique des systèmes basés uniquement sur l'information para-linguistique.

▪ **Travaux selon l'approche de classification**

Un autre critère qui distingue les différents travaux effectués est le type d'approche utilisé pour la classification des émotions. La plupart des méthodes utilisées s'inscrivent dans l'une des trois approches suivantes :

1- L'approche dynamique

Les classifieurs HMM sont utilisés dans ce type d'approche. L'avantage des modèles HMM est qu'ils permettent de modéliser la structure temporelle des énoncés. Parmi les travaux basés sur cette approche, on peut citer les travaux de Pirker (Pirker, 2007).

2- L'approche statique

Dans cette approche des classifieurs de type statique, tels que les GMM, SVM, LDC et les réseaux de neurones sont utilisés pour la modélisation. Avec ce type d'approche, il est possible également de récupérer l'information sur la structure temporelle de l'énoncé en optant pour l'information de portée à long terme comme traits caractéristiques. Quand les performances de plusieurs modèles de classificateurs sont testées par un chercheur, c'est en général le classificateur GMM qui offre de meilleures performances.

Parmi les travaux où ce type de classifieur a été expérimenté, on retrouve les travaux de Schuller (Schuller, 2007b).

3- L'approche logique floue

La troisième approche utilisée est basée sur un système d'inférence flou. Ce choix est motivé par les incertitudes qui caractérisent les émotions particulièrement, afin de résoudre le problème de l'absence de frontières claires entre les définitions des différentes catégories d'émotions ainsi que celui du chevauchement des classes dans la réalisation acoustique des émotions.

Plusieurs travaux ont examiné cette approche, nous citons les travaux de Giripunje (Giripunje, 2007).

En résumé, on peut caractériser un système de reconnaissance automatique des émotions par le type d'unité d'analyse, la portée et le type de traits caractéristiques utilisés, ainsi que le type d'approche utilisé pour la modélisation des classes d'émotions.

Il est difficile d'établir une comparaison entre les différents systèmes et d'en choisir le meilleur, pour les raisons suivantes. D'abord, notons l'absence d'un corpus de données commun des émotions qui servirait de test de performance. Aussi, les différents systèmes de RAE diffèrent dans la quantité et la qualité (le nom) de classes d'émotions utilisées dans les expérimentations. Enfin, l'utilisation de protocoles d'expérimentation dissemblables (test dépendant du locuteur et test indépendant du locuteur) rend la comparaison de performance des différents systèmes une opération hasardeuse, voire impossible.

II.7. Conclusion

Dans ce chapitre nous avons présenté des notions sur l'émotion et leurs différents types ainsi qu'un historique des études précédentes sur les différents aspects de l'émotion. A la fin nous avons terminé par une comparaison entre les différents systèmes de RAE.

Dans le chapitre suivant nous allons présenter nos onze caractéristiques proposées pour l'identification de l'émotion.

Chapitre III : Caractéristiques proposées pour l'identification

Chapitre III. Caractéristiques Proposées pour l'Identification de l'Emotion

III.1. Introduction

Les méthodes d'extraction de caractéristiques usuellement utilisées reposent sur des paramètres acoustiques et prosodiques, regroupés dans un vecteur de caractéristiques. Les paramètres acoustiques sont dérivés du traitement de la parole (MFCC, LPCC), tandis que la prosodie est caractérisée par des mesures statistiques de ses principales composantes (pitch, énergie et durée des sons produits) calculés sur les segments voisés et non voisés.

Chaque paramètre prosodique varie d'un locuteur à un autre et les différences peuvent être particulièrement importantes entre un homme et une femme; ainsi que le déséquilibre, entre le nombre de femmes par rapport au nombre d'hommes, et celui de la répartition de chaque classe en fonction du sexe du locuteur. L'étude de leurs variations en fonction de l'émotion peut être fortement biaisée. C'est pourquoi, la première étape, indispensable, consiste à séparer l'étude des paramètres selon le sexe du locuteur. Les caractéristiques permettent de réduire la quantité d'information contenue dans un signal de parole échantillonné.

Nous avons vu qu'il existait, dans le domaine de la RAE, deux types d'information pour distinguer les classes d'émotions : l'information linguistique et l'information paralinguistique. Dans ce mémoire nous allons considérer uniquement l'information paralinguistique, étant donné que nous voulons établir une étude statistique indépendante de la langue. L'expérience est réalisée en utilisant des énoncés de sens inconnu. Nous nous intéresserons dans ce qui suit, à la description de l'information prosodique.

III.2. Sélection des Caractéristiques

Les caractéristiques permettent de distinguer une forme appartenant à une classe par rapport aux formes des autres classes. Dans certains cas, il existe des caractéristiques redondantes ou non pertinentes, elles sont nuisibles en termes de calculs ou carrément inutiles (perte de temps dans l'exécution) dans la classification de l'objet d'où l'importance de la phase de sélection des caractéristiques.

L'objectif principal de la sélection des caractéristiques est de réduire le plus possible le nombre de caractéristiques utilisées tout en essayant d'améliorer significativement la performance de la classification du système de reconnaissance. Cet objectif est défini par une fonction objective à optimiser. Nous relevons deux critères essentiels dans la sélection des caractéristiques : les méthodes de sélection et les fonctions d'évaluation.

III.3. Méthodes de Sélection Citées dans la Littérature

Afin de rechercher le plus petit ensemble efficace de traits, différentes méthodes de sélection sont proposées dans la littérature de la reconnaissance des émotions. Nous dressons ci-après, la liste des méthodes utilisées dans certains travaux :

- ✓ Algorithme Relief-F (Yu, 2004) : C'est un algorithme sert à localiser le plus proche voisin de chaque échantillon aléatoire pour la même classe et la classe opposée. les valeurs des plus proches voisins des caractéristiques sont comparées aux échantillons pour mettre à jour un grand nombre des caractéristiques pertinents.
- ✓ Forward sélection (Huang, 2008) : C'est un algorithme sert à renforcer l'identification des émotions et l'élimination d'une grande partie des paramètres non pertinents ou redondants.

Backword sélection (Xie, 2007) : C'est un algorithme de sélection débute par une ou peu de caractéristiques sélectionnées selon des critères spécifiques de l'algorithme. Les caractéristiques sont itérativement ajoutées les uns après les autres pour une meilleure sélection jusqu'à l'arrêt automatique de l'algorithme.

- ✓ Algorithme génétique (Beritelli, 2007) : C'est un algorithme se base sur le calcul de chromosome et le calcul de gène.

- ✓ Facteur de corrélation (Vlasenko, 2007) : Le facteur de corrélation entre les différentes caractéristiques de la parole et les émotions prouve une certaine dépendance des caractéristiques et les émotions.

- ✓ Estimation of Distribution Algorithm (EDA) (Alvarez, 2007): C'est une nouvelle branche d'algorithmes remplace les algorithmes de sélection arbitraire des échantillons des distributions utilisées et exploite l'évaluation de la probabilité de sélection d'une distribution parmi plusieurs distributions.

- ✓ Méthode de la variance (Seppänen, 2003) : C'est une méthode permet de comparer les moyennes à une ou plusieurs variables dépendantes obtenues, dans des conditions différentes (expérimentales ou pas), sur les mêmes individus ou non (variables indépendantes intra ou inter-individuelle), en contrôlant ou pas certaines variables externes (covariables). Elle permet aussi de prédire l'appartenance à un groupe en fonction de plusieurs prédicteurs par une analyse discriminante.

- ✓ Best-First Search et Correlation-based Feature Selection (CES) (Vogt, 2005): Son but est d'avoir une meilleure extraction des caractéristiques avant la phase d'apprentissage. Cette méthode est capable de faire face contre a un grand

nombre de cas ou de données. Elle sert à réduire l'ensemble de données redondantes, non pertinentes ou du bruit en éliminant certaines colonnes et lignes. L'évaluation de la qualité de la réduction des paramètres est basée sur des protocoles métriques.

III.4. Fonctions d'Evaluation

Les fonctions d'évaluation mesurent la qualité des paramètres extraits en vue de trouver l'ensemble de traits optimal. L'ensemble de traits optimal est toujours relatif à la fonction d'évaluation utilisée, car l'utilisation d'une autre fonction d'évaluation peut générer un autre ensemble optimal.

Dash et Liu (Dash, 1997) ont proposé une classification des fonctions d'évaluation en cinq catégories en fonction du type de mesure utilisé :

1. la mesure de distance, telle que la distance euclidienne, permet de mesurer la capacité de discrimination entre les classes ;
2. la mesure d'information, qui permet d'estimer le gain d'information à partir d'une caractéristique ;
3. la mesure de dépendance qui permet de quantifier la corrélation qui existe entre les caractéristiques et leurs classes ou entre les caractéristiques elles-mêmes ;
4. la mesure de consistance de l'ensemble des caractéristiques ;
5. la mesure du taux d'erreur de la classification.

III.5. Nos Paramètres Prosodiques Proposés

Le but de notre recherche est de prouver que le changement d'état émotionnel, d'un locuteur affecte nos paramètres prosodiques proposés, avec des proportions variables et de mesurer le taux de pertinence de chaque paramètre en fonction d'une émotion.

Face à une émotion colère les variations de F0, de vitesse (dérivée de la fréquence fondamentale), et d'énergie deviennent plus élevées, que celles d'une émotion neutre. Une personne stressée met beaucoup de longues pauses et parle lentement, par contre une personne en colère met moins de pauses et parle rapidement.

Les émotions agissent fortement sur la vitesse, l'énergie (et leurs maximales et minimales) et la sur qualité de la parole. Une personne triste basse sa voix et la qualité de sa parole sera différent à celle d'une personne heureuse.

L'instabilité et les perturbations de nos onze paramètres prosodiques en fonction du changement d'état et de charge émotionnel justifient l'importance de notre choix.

Nous proposons nos onze paramètres prosodiques, sur lesquels est basée notre étude statistique qu'ils sont illustrés dans la figure 3.1.

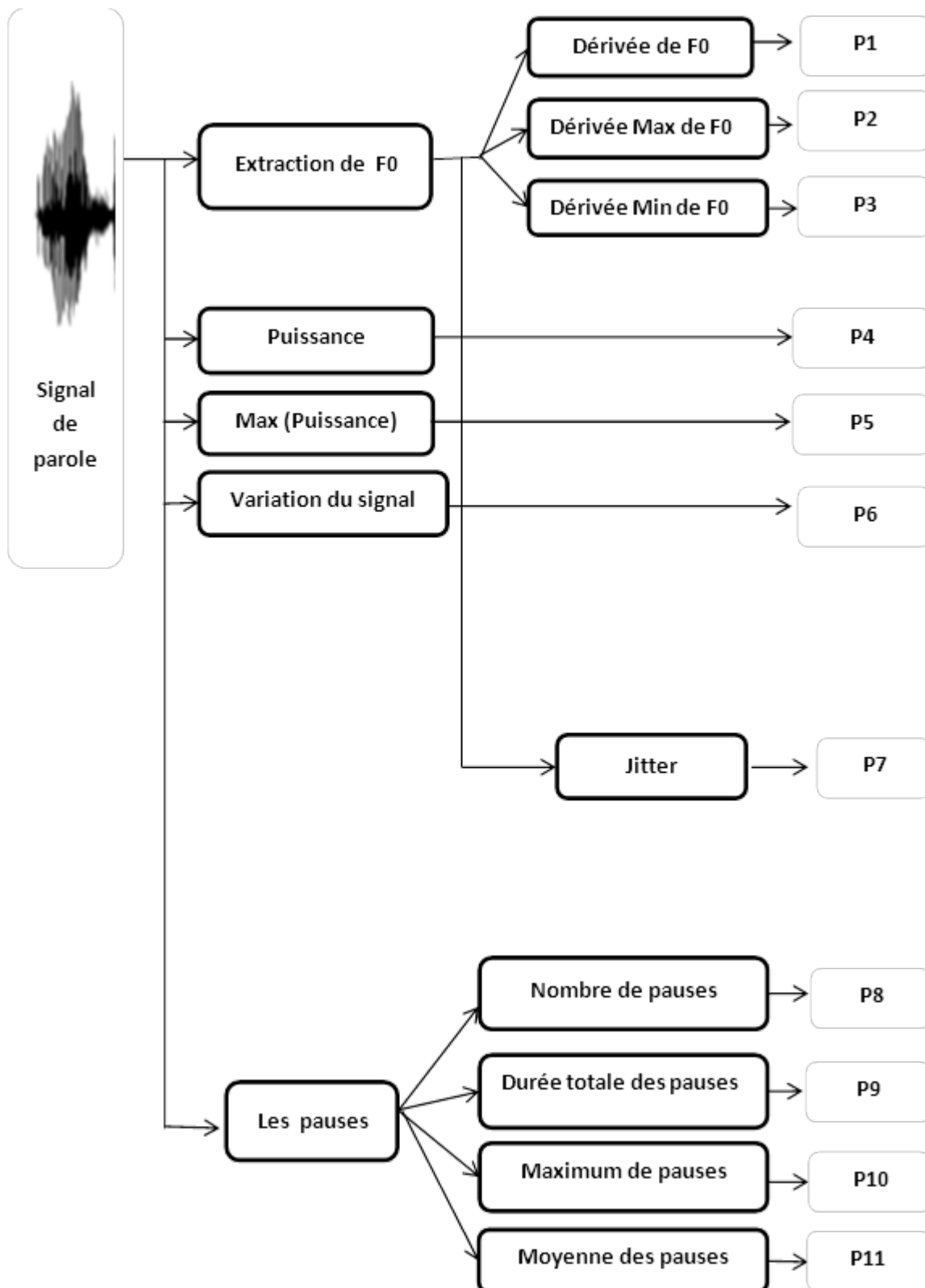


Figure 3.1 : Principe d'extraction des paramètres prosodiques.

1) Premier paramètre : 'Variation de Fo'

Il représente la différence entre les deux valeurs : maximale et minimale de Fo symbolisé par '**P1**'. Il est calculé, suivant la formule (3.1) :

$$\mathbf{P1} = \mathbf{max (Fo) - min (Fo) \dots\dots\dots (3.1)}$$

Tels que max(Fo) et min(Fo) sont respectivement les valeurs maximale et minimale de Fo du signal.

2) Deuxième paramètre : 'Maximum de dérivée de Fo'

Il représente la valeur maximale de la dérivée de Fo. Il est symbolisé par '**P2**' et calculé suivant la formule (3.2).

$$\mathbf{P2} = \mathbf{max (Fo(i) - Fo(i+1)) \dots\dots\dots (3.2)}$$

((Fo(i)-Fo(i+1)) : c'est la variation de deux valeurs adjacentes successive des Fo).

3) Troisième paramètre : 'Minimum de dérivée'

Il représente la valeur minimale de la dérivée de Fo symbolisé par '**P3**' et calculé suivant la formule (3.3).

$$\mathbf{P3} = \mathbf{min (Fo(j) - Fo(i)) \dots\dots\dots (3.3)}$$

4) Quatrième paramètre : 'Puissance'

Il représente la valeur globale de puissance du signal symbolisé par '**P4**' et donné par la formule (3.4).

$$P4 = 20 \log \left(\frac{\sum_{i=1}^N |S(i)|}{N} \right) \dots\dots\dots (3.4)$$

Avec N est la taille de la fenêtre d'analyse de 10ms dans laquelle le signal de parole s(i) est considéré comme stationnaire.

5) Cinquième paramètre : 'Maximum de puissance'

Il représente la valeur maximale de la puissance symbolisé par '**P5**', et donné par la formule (3.5).

$$P5 = \max \left(20 \log \left(\frac{\sum_{i=1}^N |S(i)|}{N} \right) \right) \dots\dots\dots (3.5)$$

6) Sixième paramètre : 'Variation du signal'

Il représente la différence, en décibels, entre les deux valeurs maximale et minimale du signal symbolisé par '**P6**', calculé par la formule (3.6).

$$P6 = 20 \log \left(\max \left(\sum_{i=1}^N |S(i)| \right) - \min \left(\sum_{i=1}^N |S(i)| \right) \right) \dots\dots (3.6)$$

N est la taille du signal de parole.

7) Septième paramètre : 'Jitter moyen'

Il représente le rapport de la moyenne de la dérivée de la période sur la moyenne de la période symbolisé par 'P7' et calculé par la formule (3.6 bis).

$$\text{Jitter (\%)} = \frac{N \sum_{i=2}^N |T_i - T_{i-1}|}{(N - 1) \sum_{i=1}^N T_i} \dots\dots\dots(3.6 \text{ bis})$$

Où T_i , représente l'inverse de la fréquence fondamentale de la $i^{\text{ème}}$ trame ($F_i = 1 / T_i$) et N le nombre total de trames voisées dans l'énoncé.

Avant d'entamer les autres paramètres concernant les pauses silencieuses ; il faut éliminer le silence de début et de la fin du signal. Puis segmenter le signal en deux parties : silence et parole.

Plus la charge émotionnelle de tristesse ou du stress est élevée plus la durée, la maximale et la moyenne des pauses seront importantes. Le contraire est vrai, d'où l'importance des paramètres suivants pour l'identification émotionnelle.

8) Huitième paramètre : 'Nombre de pauses'

Il représente le nombre des pauses, ce paramètre est symbolisé par '**P8**'.

9) Neuvième paramètre : 'Durée totale des pauses'

Il représente la durée totale des pauses. Il est symbolisé par '**P9**'.

10) Dixième paramètre : 'Maximum de pauses'

Il représente la durée maximale des pauses, symbolisé par '**P10**'.

11) Onzième paramètre : 'Moyenne des pauses'

Il représente la durée moyenne des pauses. Il est symbolisé par '**P11**'.

Notre travail permet la sélection et l'évaluation des paramètres les plus pertinents et les plus significatifs. Il s'articule sur deux étapes :

1-l'extraction des onze paramètres, séparément pour chaque classe, en éliminant ceux qui ne sont pas pertinents, en se basant sur le taux de reconnaissance des émotions et le taux d'erreur EER ;

2-l'évaluation de la consistance de nos résultats des tests de la χ^2 avec mesure de probabilité de consistance liée à l'indépendance des variables de l'émotion, ont été entrepris sur chaque résultat obtenu. Ces tests vont permettre d'éliminer les paramètres qui ne sont pas significatifs pour chaque type d'émotion.

III.6. Extraction des Caractéristiques

Nous distinguons deux types d'information utilisés comme traits caractéristiques : l'information à court terme et l'information à long terme. Les paramètres statistiques tels que 'puissance', 'variation de Fo', 'maximum de dérivée de Fo', 'minimum de dérivée de Fo' et 'variation du signal' sont utilisés comme fonctions pour obtenir l'information à long terme. Cependant, ces valeurs ne reflètent pas parfaitement les variations locales du contour de Fo et celui de l'énergie. C'est pourquoi nous utiliserons les paramètres comme 'jitter', 'maximum de puissance' et 'les pauses' pour obtenir l'information à court terme.

Pour l'extraction des 11 paramètres, on utilise les logiciels "Praat" (Boersma, 2005), le logiciel "WaveSurfer (version 1.8)" et "Matlab (version 7.1)". La fréquence fondamentale est extraite en utilisant le logiciel WaveSurfer1.8.

La fréquence fondamentale est la base des calculs des quatre paramètres (les trois premiers et le septième). Les quatre derniers sont basés sur les pauses silencieuses: nombre de pauses et leurs différentes durées, valeur maximale des pauses, somme totale des pauses, valeur moyenne des pauses. D'après nos résultats : un segment du signal est considéré comme une pause si sa durée est supérieure ou égale à 0.1s. Car s'il est inférieur à 0.1s et que son énergie est inférieure à 25db, il est considéré comme temps de respiration.

III.7. Conclusion

Dans ce chapitre nous avons défini notre propre contribution constituée de 11 caractéristiques proposées pour l'identification de l'émotion à partir des signaux de parole.

Dans le chapitre suivant nous allons présenter notre base de données universelle 'ORATOR'.

Chapitre IV. Base de données d'évaluation

Chapitre IV. Base de Données d'Evaluation

IV. 1. Introduction

Les bases de données de paroles émotionnelles dans de nombreuses langues, sont toujours orientées vers une application précise; elles se distinguent par les catégories d'émotions qu'elles regroupent, par leurs tailles, par la nature des données (discours, phrases, conversations; émotions vécues ou simulées). La plupart sont en anglais ou en chinois. La distinction la plus forte entre les bases de données, après la langue, provient de la manière dont sont générés les extraits de parole émotionnelle.

Dans le but d'évaluer notre programme de reconnaissance d'émotion nous avons besoin d'une base de données universelle qui nous permet de comparer nos résultats avec ceux obtenus par les autres chercheurs du domaine de la reconnaissance des émotions. Notre choix s'est porté sur une partie des 150 énoncés de la base de données très intéressante appelée ORATOR qui sont évalués par vingt juges, par des tests de perception.

IV.2. Base de Données ORATOR

ORATOR est un ensemble de données de la parole émotionnelle allemande simulée par des acteurs professionnels et des non acteurs, les émotions simulées sont : la colère, l'agitation, la force de caractère, la confiance, la joie, l'admiration. Son contenu verbal est un texte constant qui utilise une variété de huit phrases différentes, (4 longues et 4 courtes), telles que :

- (1) In der Vergangenheit ist schon einiges eine guter Vorarbeit geleistet worden.
- (2) Die Ziele, die wir jetzt verfolgen, sind die gleichen müssen auch auf die gleiche Weise behandelt werden.
- (3) Unsere Aufgabe ist nun, noch einmal die Zeiteinteilung durchzusehen.
- (4) Sie überprüfen dann das Weitere.

- (5) Bitte notieren Sie die Punkte, die Sie herausuchen, und tragen Sie uns diese vor!
- (6) Wir erledigen alles andere.
- (7) Glauben Sie, dass das Schaffen ?
- (8) Gut!

Leur format original est 48kHz mono WAV, ils sont codés sur 16 bits. Les orateurs sont venus d'une large gamme de milieux. Les enregistrements ont été effectués dans différents endroits : lieu de travail, maison, etc...).

Aussi, ORATOR utilise une grande variété de locuteurs.

IV.3. Présentation de notre Base de Données Universelle

Elle est composée de 150 énoncés extraits des monologues re-échantillonnés à 44,1kHz et révélant une bonne qualité acoustique de l'enregistrement. La durée de chaque énoncé varie de 1 à 11 secondes et leur durée totale est d'environ 13 minutes et 18 secondes.

IV.4. Evaluation du Degré d'Excitation Emotionnel

Les 150 énoncés sont évalués par vingt auditeurs provenant de différentes cultures (américains et allemands), rapportant leurs impressions sur l'excitation émotionnelle des signaux de parole par des scores entre [-2 à +2], qui sont rarement égaux à -2 ou +2.

IV.5. Influence de l'Information Lexicale sur l'Identification Emotionnelle

L'information lexicale n'influe pas sur l'identification émotionnelle pour les raisons suivantes :

- les mêmes phrases ont été utilisées ;
- la langue d'expression nous est inconnue (allemand).

IV.6. Répartition des Données en fonction du Sexe des Locuteurs

Tableau 4.1 : Répartition des données en fonction du sexe des locuteurs.

	Femme	Homme
Nombre des locuteurs	59	91

Le nombre de données correspondant aux hommes est presque deux fois plus élevé que celui correspondant aux femmes (tableau 4.1).

La répartition de chaque classe d'émotion en fonction du sexe des locuteurs est présentée dans les tableaux suivants :

Tableau 4.2 : Répartition de l'émotion 'agitation' en fonction du sexe des locuteurs.

	Femme	Homme
agitation	20	13
non agitation	39	78

Le tableau 4.2 indique que le nombre de données correspondant aux femmes :

- pour la classe 'agitation' est environ deux fois plus élevé que celui correspondant aux hommes ;
- pour la classe 'non agitation', la tendance s'inverse.

Tableau 4.3 : Répartition de l'émotion 'colère' en fonction du sexe des locuteurs.

	Femme	Homme
colère	7	8
non colère	52	83

Nous observons dans le tableau 4.3 que le nombre de données correspondant aux locuteurs femmes est quasiment égal à :

- celui correspondant aux locuteurs hommes pour la classe d'émotion 'colère' ;
- la moitié de celui correspondant aux locuteurs hommes, pour la classe d'émotion 'non colère'.

Tableau 4.4 : Répartition de l'émotion 'confiance' en fonction du sexe des locuteurs.

	Femme	Homme
confiance	30	36
non confiance	29	55

Nous remarquons dans le tableau 4.4 que le nombre de données se rapportant aux femmes est pratiquement égal à :

- celui se rapportant aux hommes, pour la classe d'émotion 'confiance' ;
- la moitié de celui se rapportant aux hommes, pour la classe d'émotion 'non confiance'.

Tableau 4.5 : Répartition de l'émotion 'fort (de caractère)' en fonction du sexe des locuteurs.

	Femme	Homme
fort (de caractère)	13	26
faible (de caractère)	46	65

Le tableau 4.5 montre que le nombre des signaux de parole propres aux hommes représente :

- le double de celui propre aux femmes, pour la classe d'émotion 'Force (de caractère)';
- les deux tiers de celui propre aux femmes, pour la classe d'émotion 'Faible (de caractère)'.

IV.7. Conclusion

Dans ce chapitre nous avons présenté la base de données de la parole émotionnelle allemande 'ORATOR' simulée par des acteurs professionnels et des non acteurs.

Dans le chapitre suivant nous allons détailler les expériences réalisées ainsi que les résultats obtenus et leurs interprétations.

Chapitre V : Résultats et discussions

Chapitre V. Résultats et Discussions

V.1. Introduction

La technique d'extraction de caractéristiques proposée est appliquée sur la totalité des énoncés du corpus allemand reposant sur des paramètres prosodiques, et de qualité de voix regroupés dans des vecteurs de caractéristiques qui sont : variation de F_0 , maximum et minimum de dérivée de F_0 , l'énergie et son maximum, variation du signal, perturbation de F_0 ('jitter'), et le nombre de pauses, leurs durées totales, le maximum et la moyenne des durées.

La tâche principale de cette étude est de préciser à quel point ces paramètres contribuent à la reconnaissance de l'excitation émotionnelle (colère, agitation, force de caractère et confiance) dans la parole, par le calcul de leurs taux de reconnaissance maximaux, et d'erreur EER.

L'importance, la signification et la consistance des paramètres dans la reconnaissance émotionnelle sont confirmées en fonction du calcul de la χ^2 et de la probabilité P de dépendance qui est déterminée en fonction de la χ^2 .

Le degré d'excitation d'émotion de chaque énoncé sera décrit en ayant recours au jugement de vingt auditeurs. On considère la valeur de la moyenne des scores de 0.65 comme un seuil de référence d'émotion pour chaque paramètre. Si la moyenne des scores est supérieure ou égale à ce seuil, le signal de parole est émotionné (ou fortement émotionné). Si non, il est neutre (ou très faiblement émotionné).

V.2. Représentation Synoptique des différentes étapes d'identification

Nous résumons notre recherche sous forme d'un schéma synoptique illustré ci-dessous :

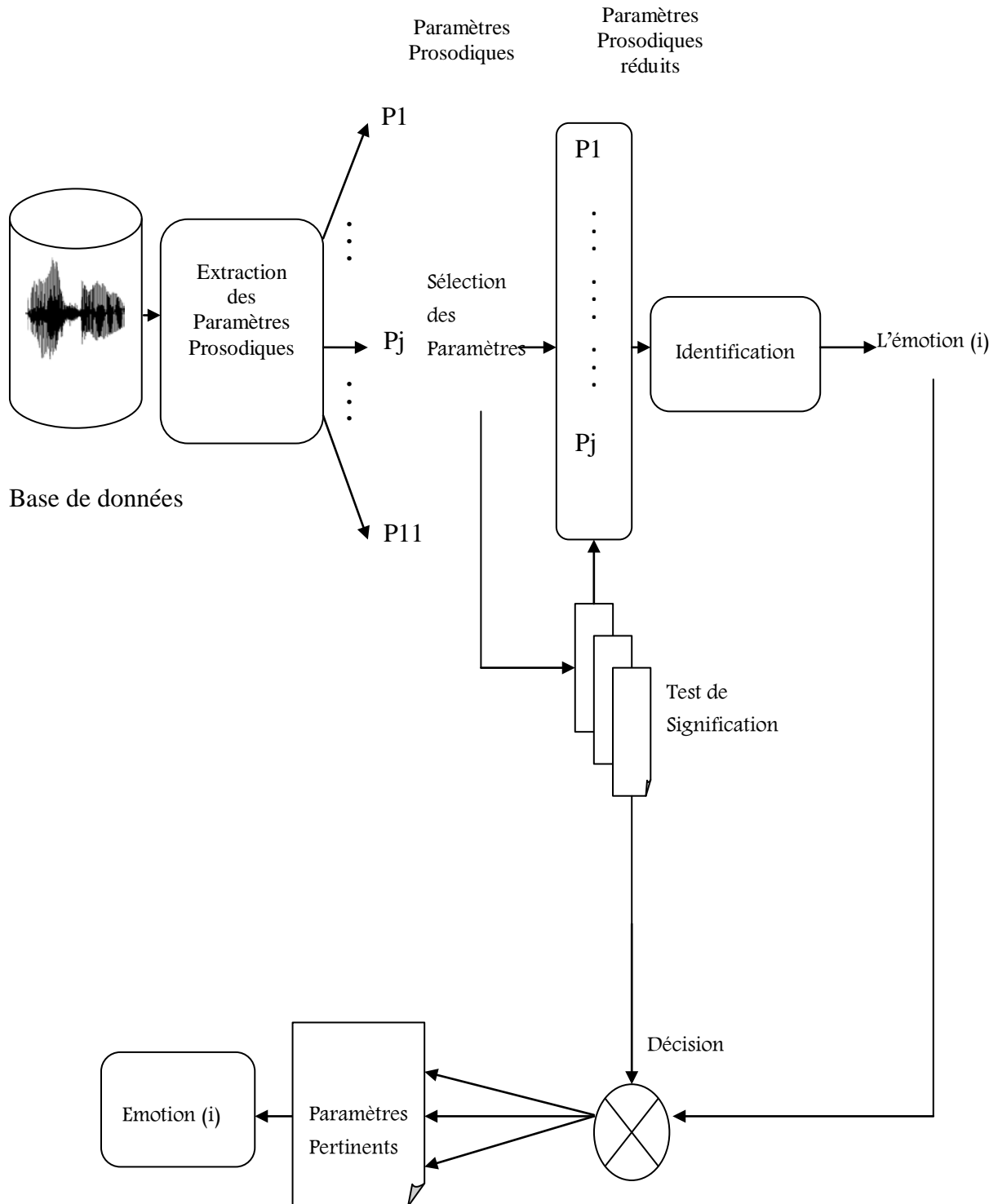


Figure 5.1 : Schéma synoptique des différentes étapes d'identification.

V.3. Fréquence Fondamentale et le Contours de Pitch

Les séquences audio d'ORATOR ont une fréquence d'échantillonnage de 44.1kHz et possèdent ainsi une bande passante de 22,05 kHz. Chaque échantillon est codé sur 16 bits. Cette qualité d'enregistrement est largement suffisante pour traiter des signaux de parole sachant qu'une oreille humaine performante perçoit des fréquences comprises entre 0Hz et 20kHz.

Nous avons tracé tous les contours de pitch ainsi que les signaux de parole pour toute la base de données. Nous avons l'intention d'étudier l'évolution de la fréquence fondamentale à partir de contours des pitch mais cela demandera beaucoup plus de temps.

Les figures 5.2 à 5.9 représentent les schémas de contours de pitch et la forme acoustique des signaux de parole des huit émotions

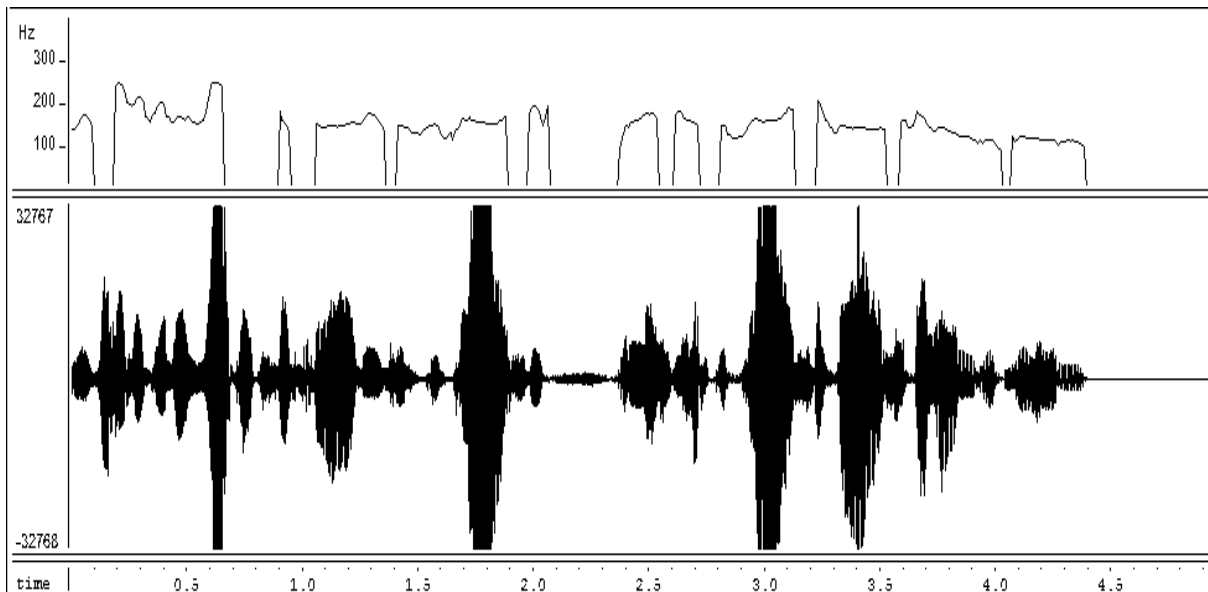


Figure 5.2: Contour du pitch du signal de parole du locuteur N°13 pour l'émotion agitation.

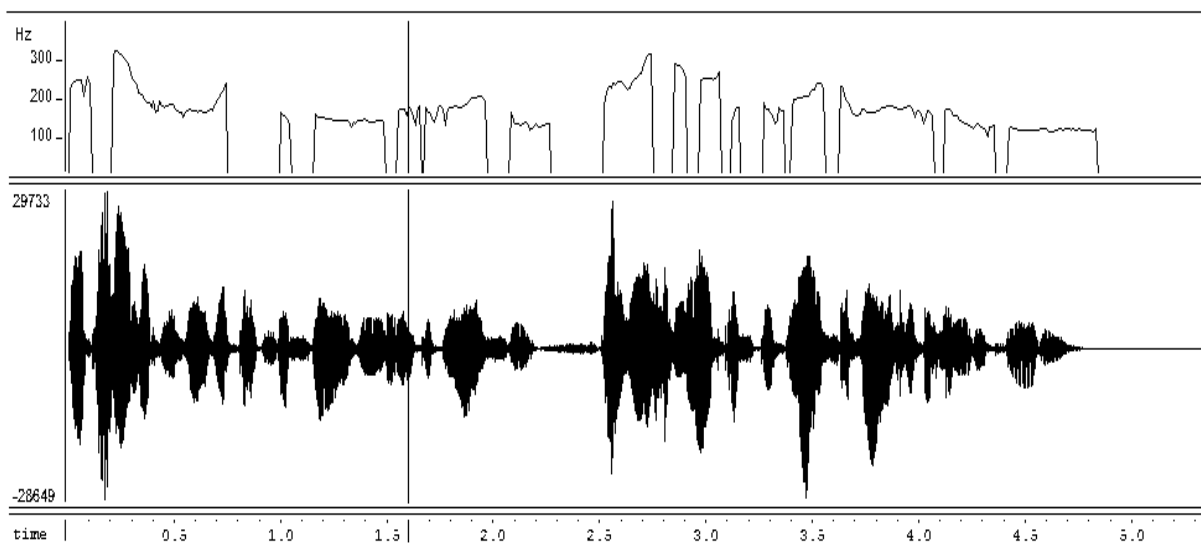


Figure 5.3 : Contour du pitch du signal de parole de la locutrice N°7 pour l’émotion agitation.

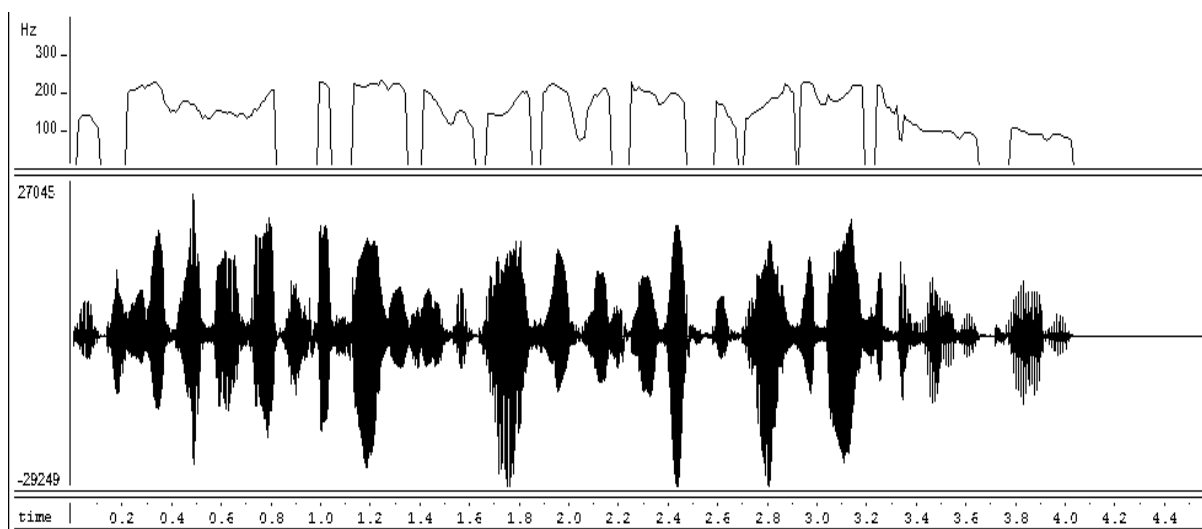


Figure 5.4 : Contour du pitch du signal de parole du locuteur N°69 pour l’émotion colère.

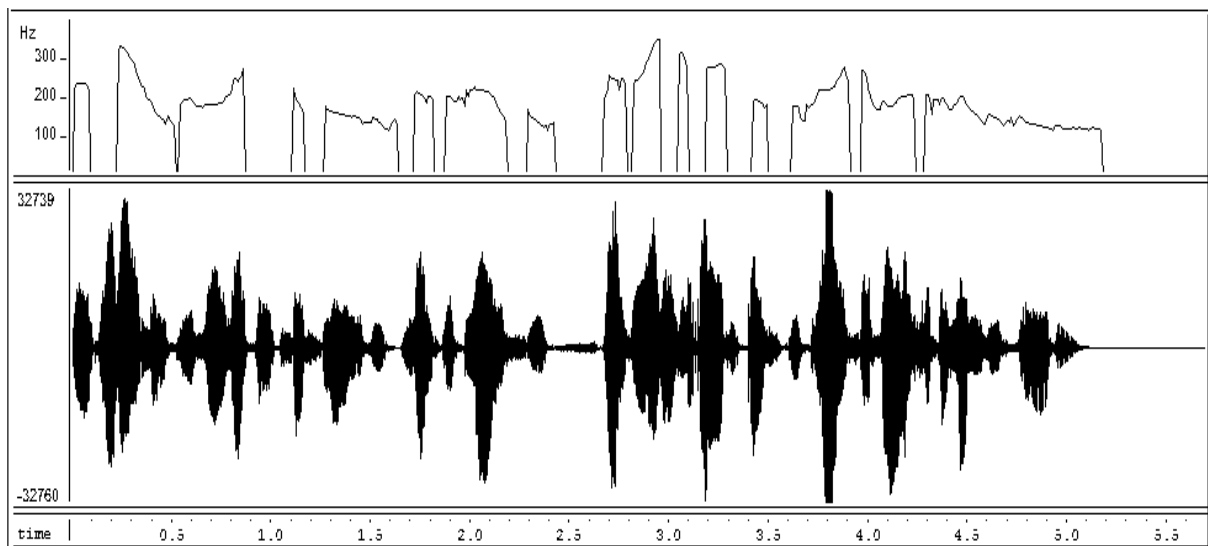


Figure 5.5 : Contour du pitch du signal de parole de la locutrice N°6 pour l'émotion colère.

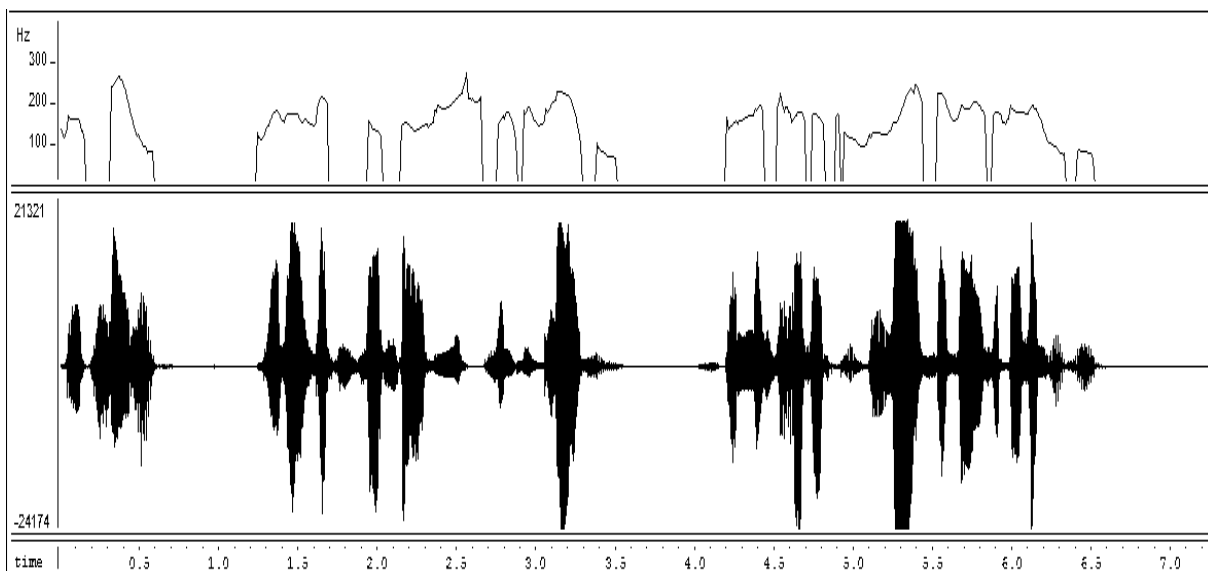


Figure 5.6 : Contour du pitch du signal de parole du locuteur N°80 pour l'émotion confiance.

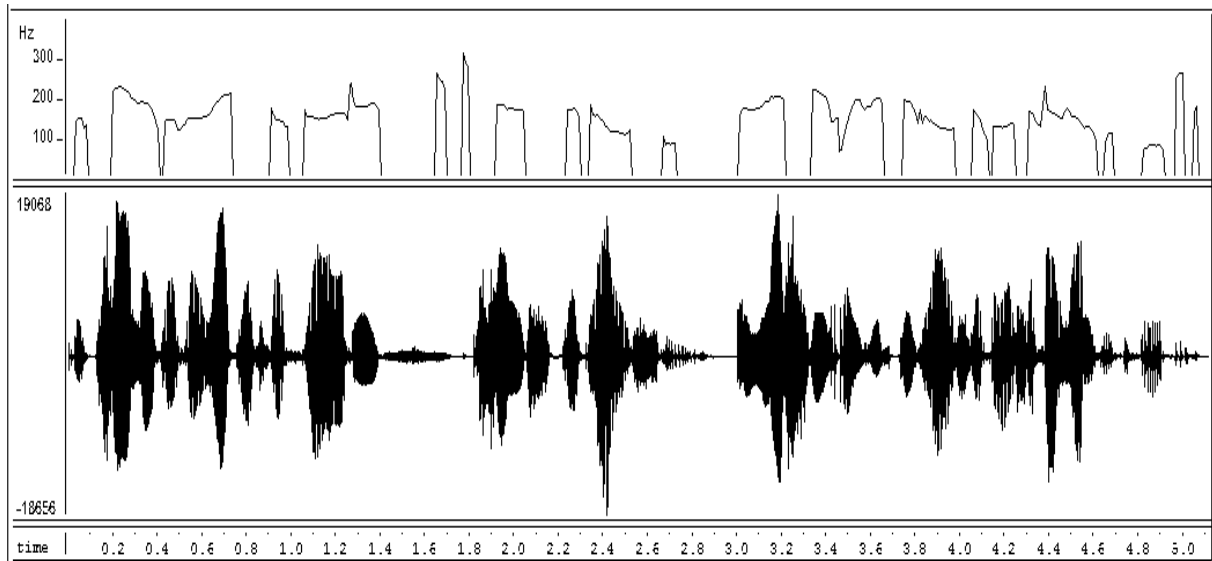


Figure 5.7 : Contour du pitch du signal de parole de la locutrice N°56 parole de pour l'émotion confiance.

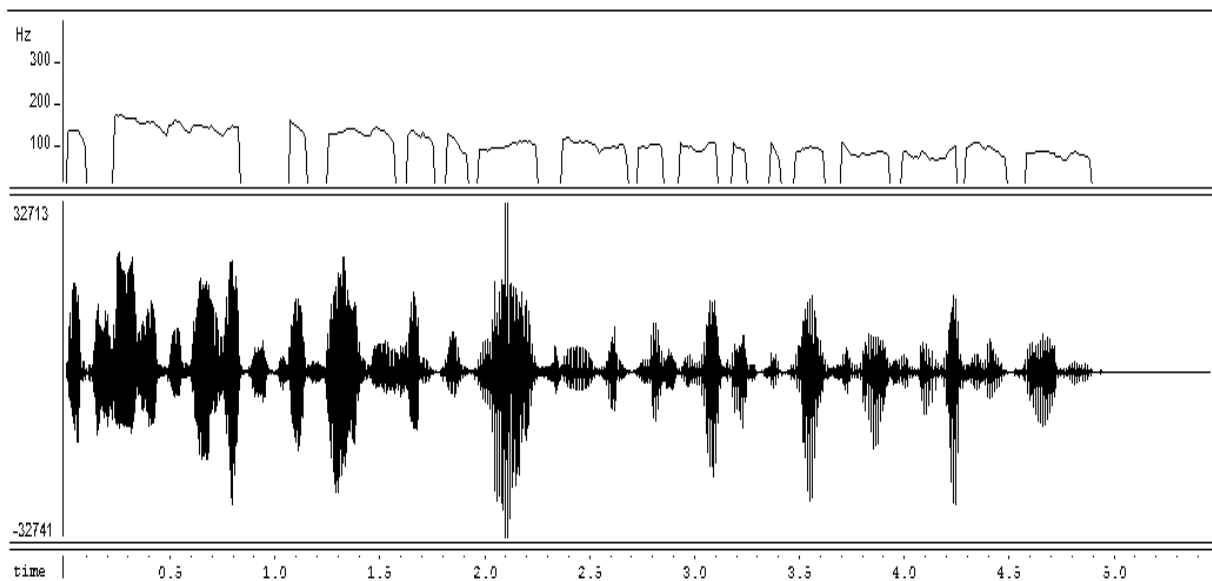


Figure 5.8 : Contour du pitch du signal de parole du locuteur N°113 pour l'émotion force (de caractère).

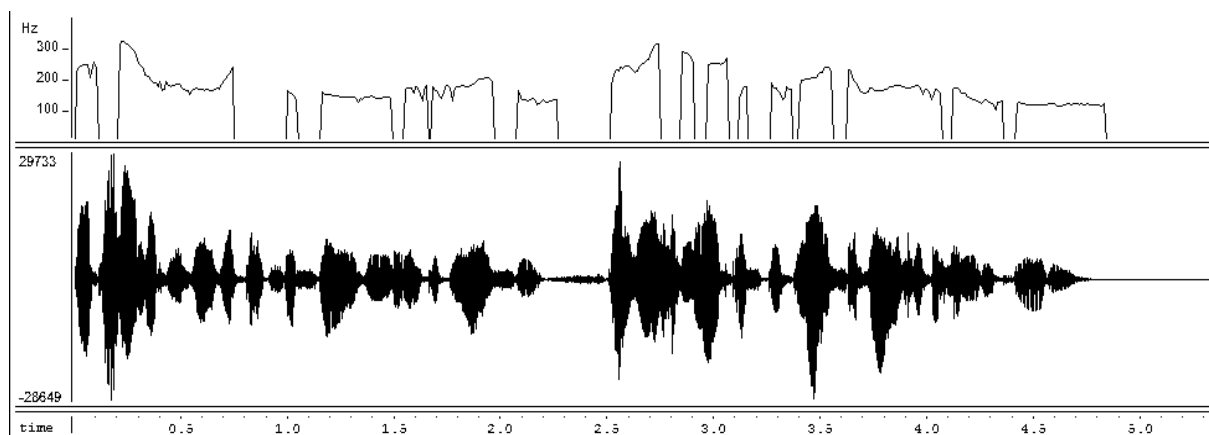


Figure 5.9 : Contour du pitch du signal de parole de la locutrice N°7 pour l'émotion force (de caractère).

Nous remarquons que l'évolution de la fréquence fondamentale F_0 au cours du temps décrit des contours. Une phrase est généralement composée d'une suite de contours qui ne suivent pas nécessairement la même orientation de pente. Nous observons cependant une déclinaison générale qui correspond à un abaissement de F_0 du début à la fin de l'énoncé.

La hauteur la plus basse correspond donc à la fin de cet énoncé et constitue ainsi un bon indice de segmentation.

Une déclinaison est remarquée après chaque pause importante ou respiration.

Ce phénomène à priori universel est de nature physiologique, mais il est géré par le locuteur à des fins linguistiques; il permet de délimiter la fin d'une phrase syntaxique. Il faut remarquer que l'on ne peut évaluer cette fréquence fondamentale que sur les segments voisés (voyelles et quelques consonnes...). En parlant, nous varions systématiquement notre

fréquence fondamentale pour exprimer nos sentiments ou pour diriger l'attention de l'auditeur vers un aspect important de notre message parlé. Un paragraphe prononcé avec un pitch constant et uniforme paraîtra peu naturel (Huang, 2001).

Par contre pour les segments non voisés, le signal est considéré comme un bruit et la mesure de F0 est égale à zéro. Cela veut dire qu'en sortie, nous obtenons une courbe par morceaux de la fréquence fondamentale estimée seulement sur les segments voisés.

Toutes les figures (5.2 à 5.9) de contour de pitch indiquent la déclinaison générale due à l'expiration (moins d'air dans les poumons en fin de phrase) et quelques variations intrinsèques des phones dues à la prononciation du locuteur. Les parties non voisées correspondent bien aux consonnes (plosives, sifflantes, fricatives...) (Beller, 2005)

De plus, La hauteur de la voix étant fondamentalement différente selon le locuteur, on ne peut associer aux niveaux décrits précédemment des valeurs de fréquence fixes. De manière à comparer les groupes prosodiques entre eux, les contours de F0 sont normalisés en durée et en hauteur (Audibert, 2008).

Nous constatons ainsi, que, la fréquence fondamentale est statistiquement plus élevée pour les femmes et les enfants que pour les hommes. Le volume des poumons, la position du larynx, par exemple, varient fortement avec le sexe, l'âge et l'individu. Notamment le larynx est plus élevé chez la femme ce qui aura des conséquences lors de la production vocalique.

V.4. Plages de Variation des Paramètres

Le calcul des 11 paramètres extraits, a permis d'obtenir les plages de variations de chacun d'eux.

Tableau 5.1 : Répartition des plages de variation des paramètres en fonction du sexe des locuteurs.

Symbole	paramètre	Femme [Min - Max]	Homme [Min - Max]
P1	'variation de Fo'	[112,75 - 268,57]	[27,37 - 295,21]
P2	'maximum de dérivé (Fo)'	[19,11 - 154,87]	[7,66 - 161,44]
P3	'minimum de dérivé (Fo)'	[21,89 - 173,90]	[9,21 - 121,53]
P4	'puissance'	[111,86 - 155,97]	[105,02 - 156,55]
P5	' maximum de puissance'	[74,27 - 94,75]	[70,71 - 99,13]
P6	'variation du signal'	[200,18 - 221,80]	[216,17 - 221,80]
P7	''jitter''	[1,58 - 3,97]	[1,94 - 4,73]
P8	'nombre de pauses'	[0 - 11]	[0 - 11]
P9	'durée totale des pauses'	[0 - 25]	[0 - 40]
P10	'maximum des pauses'	[0 - 10]	[0 - 17]
P11	'moyennes pauses'	[0 - 8]	[0 - 7,5]

Nous remarquons que ces plages varient d'un locuteur à un autre suivant : le sexe (masculin ou féminin), et l'âge de chacun d'eux (tableau 5.1).

Ainsi, ces paramètres prosodiques changent suivant l'état émotionnel et l'état de santé du locuteur.

Le nombre de pauses diffère d'un locuteur à un autre, cela est dû : à la différence qui existe entre les humains, l'état émotionnel, l'état de santé du locuteur et les expressions émotionnelles qui varient largement d'un locuteur à un autre.

Les marges de variation de tous les paramètres sont plus remarquables pour les hommes que pour les femmes (à l'exception de P8 et P11 qui sont égales), cela est dû au fait que les hommes ne savent pas cacher leur colère et agitation, alors que les femmes sont douées; et le font de manière naturelle, elles sont capables de cacher leurs colères, leurs agitations et leurs préoccupations, cela fait parti des coutumes et traditions, aussi, la femme cherche toujours à entretenir sa réputation et sa place au sein de la société.

Par contre l'homme, lorsqu'il se révolte (colère ou agitation), il ne se contrôle pas, la force de son caractère et de sa confiance se manifeste dans sa voix grave, tandis que la voix de la femme est aigue, en effet elle cache sa force de caractère et sa confiance.

V.5. Fausse Acceptation et Faux Rejet

En biométrie, une acceptation fautive d'un système de sécurité (par exemple) vérifiant ou identifiant une personne non autorisée, est considérée comme la plus grave des erreurs de sécurité biométriques de type II. Car le système donne l'accès aux utilisateurs non autorisés aux systèmes alors qu'il faut les empêcher d'entrer.

Un faux rejet du système de sécurité pour vérifier ou identifier une personne autorisée est considéré comme une erreur de type I, car le système refuse l'accès aux utilisateurs autorisés.

Le taux de fautive acceptation est la mesure de la probabilité que le système de sécurité biométrique accepte, de manière incorrecte, une tentative d'accès par un utilisateur non autorisé.

Le taux de faux rejet est la mesure de la probabilité que le système de sécurité biométrique rejette, de manière incorrecte, une tentative d'accès par un utilisateur autorisé (Griaule biometrics.com) (webopedia.fr).

V.6. Taux d'Erreur Egal (Equal Error Rate, EER)

Un système de sécurité biométrique prédétermine des valeurs de seuil pour les fausses acceptations et les faux rejets. Et quand les taux sont égaux, la valeur commune est désignée par les chercheurs sous le nom du taux d'erreur égal (Equal Error Rate EER).

Cette valeur indique que le nombre de fausses acceptations est égale au nombre de faux rejets. Plus la valeur du taux d'erreur égal (EER) est faible, plus l'exactitude du système biométrique est importante. (figure 5.10) (Griaule biometrics.com) (webopedia.fr) (Boite, 2000) (GloCCal.com) (Dorizzi, 2004)

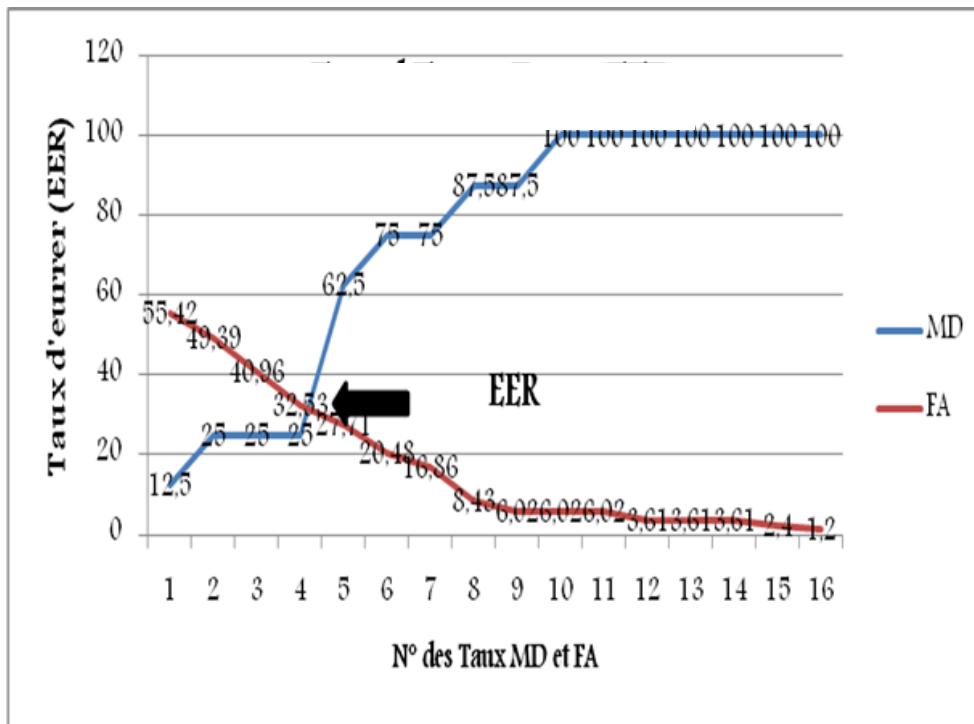


Figure 5.10 : Taux d'erreur égal (EER) pour l'émotion: colère, avec le paramètre 'P1', 'sexe : (M')

V.7. Estimation du Seuil EER

En optant provisoirement pour un ‘seuil’ initial pour chaque émotion qui semble le plus proche de la frontière de cette dernière, puis on sélectionne des valeurs autour de chaque seuil. Chaque paramètre est alors comparé à un seuil au-dessus duquel l’émotion est reconnue.

Le choix de ce seuil est crucial parce qu’il influe sur le bon fonctionnement du système de détection, si ce dernier est trop élevé le taux de fausse acceptation sera trop important, alors que s’il est trop bas le taux de de faux rejet sera inacceptable.

Par conséquent, les performances des systèmes de reconnaissance ou de vérification sont souvent exprimées en terme de taux d’erreurs égal (EER) correspondant au point de fonctionnement (seuil de décision) pour lequel le taux de fausses acceptations est égal au taux de faux rejets.

Le seuil EER est souvent estimé en supposant que les distributions (normalisées) sont gaussiennes, et en calculant le seuil EER résultant. Si ce seuil EER est estimé sur l’ensemble d’entraînement, il sera souvent appelé seuil a priori. Alors qu’il sera appelé seuil a posteriori s’il a été estimé sur l’ensemble du test. Evidemment, dans les systèmes réels, le seuil a posteriori n’est pas acceptable (et il faudra supposer que la base d’entraînement est suffisamment représentative des conditions réelles), les performances EER seront alors souvent remplacées (estimées) par la moitié de la somme des taux de faux rejet et de fausse acceptation.

Ainsi, nous définissons 3 taux pour l’évaluation des résultats :

1. Taux de reconnaissance : définit par le nombre d’émotions bien reconnues sur le nombre de phrase de test, désigné par : ‘ Taux Reco ’, ce taux est exprimé par la formule (5.1) suivante :

$$\text{Taux Reco \%} = 100 - \frac{\text{nombre d'erreurs}}{\text{nombre de locuteurs}} * 100 \dots\dots (5.1)$$

2. Taux de détections manquées des émotions, (Missed Detections), désigné par MD, et exprimée selon la formule (5.2) :

$$\text{MD \%} = \frac{\text{nombre de cas émotionnés non détectés}}{\text{nombre total des locuteurs émotionnés}} * 100 \dots(5.2)$$

3. Taux de fausses alarmes (false alarms), désigné par FA suivant la formule (5.3):

$$\text{FA \%} = \frac{\text{nombre de cas détectés émotionnés alors qu'ils ne le sont pas}}{\text{nombre total des locuteurs émotionnés}} * 100 \dots(5.3)$$

V.8. Problématique

En cherchant à avoir les valeurs les plus basses du taux d'erreur égal (EER), Puisqu'on ne peut pas avoir l'égalité des taux (MD, FA), on choisit les seuils qui correspondent à la moyenne des valeurs les plus proches des taux (MD \approx FA), on les désigne par seuil_{EER},

La figure 5.11 montre le rapprochement des taux (MD \approx FA). La parole est émotionnée si la valeur d'un paramètre considéré est supérieure ou égale au seuil_{EER}.

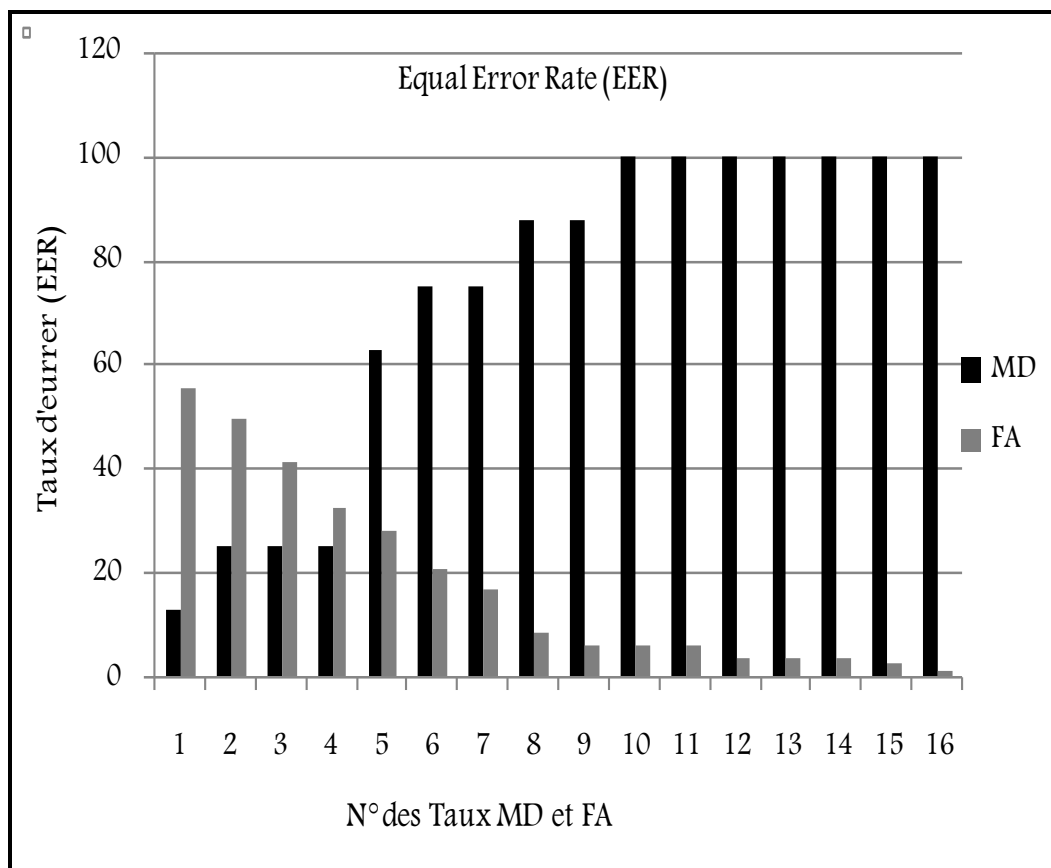


Figure 5.11 : Estimation du seuil EER l'émotion : colère avec le paramètre 'P1 (variation du signal)', sexe : (M).

V.9. Récapitulatif des Seuils Correspondant aux EER (seuil_{EER})

Les valeurs des seuils (seuil_{EER}), trouvés pour chaque émotion sont résumés dans les tableaux 5.2 et 5.3 suivants

Tableau 5.2 : Seuils EER obtenus pour les locuteurs masculins.

paramètre	Seuil de l'émotion			
	Agitation	Colère	Confidence	Fort (de caractère)
P1	149	150	140	140
P2	33	/	46	/
P3	33	/	/	/
P4	145	145	140	144
P5	89	90	88	89
P6	218	218.5	216.8	217
P7	3	3	3	3
P8	2	1	3	2
P9	3	2	6	5
P10	2	2	4	3
P11	1.5	1	2	2

/ : indique que le paramètre correspondant n'apporte rien à l'identification de l'émotion.

Tableau 5.3 : Seuils EER obtenus pour les locuteurs féminins.

Paramètre	Seuil de l'émotion			
	Agitation	Colère	Confidence	Fort (de caractère)
P1	195	200	180	200
P2	61	/	60	/
P3	54	/	/	/
P4	141	148	136	141
P5	87	89	85	88
P6	217	218	216,5	217,5
P7	2.5	2.5	2.5	2.5
P8	2	1	2	1
P9	3	2	4	3
P10	2	1	3	2
P11	2	1	2	2

/ : indique que le paramètre correspondant n'apporte rien à l'identification de l'émotion.

Concernant les seuils correspondant à l'erreur EER obtenus pour chaque émotion, ils sont différents dans la plupart des cas d'une variable à l'autre pour le locuteur féminin et masculin à l'exception du paramètre P7 pour les deux sexes (tableaux 5.2 et 5.3).

Ainsi, les valeurs des taux EER et des taux de reconnaissance maximaux trouvés pour chaque émotion sont résumés dans les tableaux 5.4 et 5.5 suivants :

V.10. Récapitulatif des Taux de Reconnaissance Maximaux et d'Erreur EER

Tableau 5.4 : Taux d'erreurs EER et taux de reconnaissance maximaux obtenus, pour chaque, émotion dans le cas de locuteurs masculins.

Paramètre	Agitation		Colère		Confidence		Fort (de caractère)	
	EER (%)	Taux Reco max (%)	EER (%)	Taux Reco max (%)	EER (%)	Taux Reco max (%)	EER (%)	Taux Reco max (%)
P1	30,76	81,32	28	90,10	28,91	72,52	34,99	70,32
P2	44,86	83,51	/	/	43,33	65,93	/	/
P3	39,09	85,71	/	/	/	/	/	/
P4	24,35	86,81	20,10	92,30	22,92	79,12	19,99	83,51
P5	24,99	86,81	24,54	92,30	42,11	78,02	30,76	73,62
P6	25,63	89,01	24,54	90,10	29,33	70,32	32,3	74,72
P7	57,04	81,11	61,37	84,44	54,51	61,11	58,45	52,22
P8	25,63	85,71	19,72	91,20	33,45	67,03	31,53	80,67
P9	25,63	85,71	23,94	91,2	38,53	65,93	34,99	82,41
P10	25,63	85,71	28,16	91,20	40,78	62,63	34,22	82,41
P11	28,84	85,71	19,72	91,20	51,83	62,63	33,84	82,41

Nous rappelons que :

- le taux de reconnaissance maximal est le taux d'émotions bien reconnues dans le meilleur cas ;
- le taux EER (Equal Error Rate) est le taux d'erreur égal.

Les résultats obtenus (tableau 5.4) montrent les points suivants :

- pour l'émotion agitation, les taux de reconnaissance maximaux restent toujours supérieurs à 81% pour tous les paramètres ;
- pour l'émotion colère, les taux de reconnaissance maximaux dépassent 90% pour les paramètres des quatrième jusqu'au onzième sauf pour le septième paramètre 84,44%. Le deuxième et troisième paramètres n'apportent rien à la reconnaissance de cette émotion, ainsi que pour le paramètre P1;
- pour l'émotion confiance, le score de reconnaissance maximal varie entre 61,11% et 79,12% pour tous les paramètres à l'exception du troisième paramètre qui n'apporte rien à la reconnaissance de cette émotion ;
- pour l'émotion force de caractère les taux maximaux atteignent les 80,67% à 83,51% jusqu'à 83% du huitième jusqu'au onzième paramètre ainsi que pour le quatrième paramètre. Ils varient entre 52,22% et 74,72% du cinquième au septième paramètre de même pour le premier paramètre.
- Le deuxième paramètre n'apporte rien à la reconnaissance des émotions colère et force de caractère ;
- Le troisième paramètre aussi n'apporte rien à la reconnaissance des émotions colère, confiance et force de caractère.

On observe aussi les points suivants :

- pour l'émotion agitation, le taux d'erreur EER est entre 24,35% et 57,04% ;
- pour l'émotion colère, le taux EER est 19,72% pour le huitième et le onzième paramètre. Ce taux est 20,10% pour le quatrième paramètre. Il est de 23,94% et 61,37% pour les autres paramètres sauf pour le deuxième et quatrième ;
- pour l'émotion confiance, le taux d'erreur est 22,92% et 54,51% pour tous les paramètres à l'exception du deuxième ;
- pour l'émotion force de caractère, le taux d'erreur est le plus bas possible pour le quatrième paramètre et il varie entre 30,76% et 58,45% pour les autres paramètres sauf pour le deuxième et le troisième paramètre.

Tableau 5.5 : Taux d'erreurs EER et taux de reconnaissance maximaux obtenus, pour chaque, émotion dans le cas de locuteurs féminins.

Paramètre	Agitation		Colère		Confidence		Force de caractère	
	EER (%)	Taux Reco max (%)	EER (%)	Taux Reco max (%)	EER (%)	Taux Reco max (%)	EER (%)	Taux Reco max (%)
P1	34,16	67,79	31,59	84,74	32,24	71,18	25,66	79,66
P2	50,64	49,15	/	/	54,25	50,84	/	/
P3	50,57	69,49	/	/	/	/	/	/
P4	22,75	81,35	17,71	84,74	32,24	74,57	10,89	83,05
P5	16,75	84,74	16,75	89,83	30,74	69,49	23,49	84,74
P6	37,88	76,27	14,83	86,44	38,90	67,79	21,31	88,13
P7	53,07	62,71	34,47	84,74	47,35	57,62	53,59	69,49
P8	41,98	69,49	32,96	88,13	27,35	64,40	42,72	77,96
P9	44,35	69,49	22,52	88,13	25,51	74,57	49,16	77,96
P10	35,38	69,49	32,96	88,13	22,18	77,96	41,55	77,96
P11	39,23	69,49	32,96	88,13	35,51	67,79	39,88	77,96

D'après le tableau 5.5, nous constatons que :

Les meilleurs taux de reconnaissance maximums (Taux Reco max) ont été obtenus en utilisant le cinquième paramètre pour les deux émotions agitation et colère 84,74% et 89,83% respectivement et pour le dixième paramètre pour l'émotion colère 88,13% ainsi pour le sixième paramètre pour l'émotion force de caractère 88,13%.

Les plus bas taux d'erreurs EER (Equal Error Rate) ont été obtenus avec le cinquième paramètre pour l'émotion agitation 16,75% ainsi que pour le quatrième, cinquième et le sixième paramètres pour l'émotion colère 17,71%, 16,75%, 14,83% respectivement et pour le dixième paramètre pour l'émotion confiance 22,18% et avec le quatrième paramètre pour l'émotion force de caractère 10,89 %.

Le deuxième n'apporte rien à la reconnaissance des émotions colère et force de caractère. Le troisième paramètre aussi n'apporte rien à la reconnaissance des émotions colère, confiance et force de caractère.

V.11. Interprétation et Discussions des Résultats

▪ Observation pour le paramètre P1 (Variation de Fo)

La variation de Fo renseigne sur les points suivants :

- l'incertitude pour reconnaître les quatre émotions pour le sexe masculin puisque l'erreur EER est considérable, malgré leurs taux de reconnaissance maximaux qui sont assez bons pour l'agitation et la colère (tableau 5.4) ;
- l'impuissance pour identifier l'émotion agitation dans le cas de locuteur féminin (tableau 5.5) ;
- l'incertitude pour détecter les trois émotions colère, confiance et force de caractère du sexe féminin puisque l'erreur EER est élevée, malgré que le taux de reconnaissance maximale est appréciable (tableau 5.5).

▪ Observation pour le paramètre P2 (Maximum de dérivée de Fo)

On observe les points suivants :

- ce paramètre n'apporte rien pour la reconnaissance des deux émotions colère et force de caractère pour les deux sexes masculin et féminin (tableaux 5.4 et 5.5) ;
- il est non fiable pour détecter l'émotion agitation pour le sexe masculin puisque l'erreur EER est considérable, malgré que le taux de reconnaissance maximal est assez bon (tableau 5.4) ;
- il est défaillant pour identifier l'émotion confiance pour les deux sexes, ainsi que, pour l'émotion agitation dans les cas féminins (tableau 5.4 et 5.5).

▪ **Observation pour le paramètre P3 (Minimum de dérivée)**

- Les trois émotions colère, confiance et force de caractère ne sont pas reconnues par le paramètre pour les deux sexes (tableaux 5.4 et 5.5) ;
- concernant la reconnaissance de l'émotion agitation pour le sexe masculin nous remarquons la non fiabilité de ce paramètre, puisque l'erreur EER est considérable 39,19%, bien que le taux de reconnaissance maximal est de 85,71% (tableau 5.4) ;
- une défaillance est remarquée quant à la détection de l'émotion agitation sexe féminin dûe aux valeurs EER =50,57% et Taux Reco Max = 69,49% (tableau 5.5).

▪ **Observation pour le paramètre P4 (puissance)**

Nous remarquons les points suivants :

- la pertinence de ce paramètre pour l'identification des deux émotions colère et force de caractère pour le cas de locuteur masculin avec un taux de reconnaissance de 92,30 % et 83,51% et un taux EER 20,10% et 19,99% respectivement, de même pour le cas de locuteur féminin avec un taux de reconnaissance de 84,74 % et 83,05% respectivement et un taux EER 17,71 % et 10,89% respectivement (tableaux 5.4 et 5.5) ;

Nous remarquons aussi la non fiabilité de ce paramètre pour la détection des deux émotions restantes (agitation et confiance) :

- dans le cas de locuteur masculin vu les appréciables taux de reconnaissance maximaux de 86,81% et 79,12% puisque l'erreur EER est considérable, elle est de l'ordre 24,35% et 22,92% respectivement ;
- dans le cas de locuteur féminin vu que les erreurs EER sont élevées 22,75% et 32,24%, en contrepartie les taux de reconnaissance maximaux sont moyens 81,35% et 74,57% respectivement.

▪ **Observation pour le paramètre P5 (Maximum de puissance)**

Le maximum d'énergie (puissance) dans un signal a permis de faire les taches suivantes :

- discriminer les deux émotions agitation et colère pour le sexe féminin, leurs taux de reconnaissance maximaux correspondants sont : 84,74%, 89,83% respectivement, et leurs taux d'erreur EER sont : les plus bas 16,75%, 16,75% respectivement (tableau 5.5) ;
- montrer son insuffisance pour identifier les quatre émotions du sexe masculin dû à leurs erreurs EER qui sont inacceptables, bien que leurs taux de reconnaissance maximaux sont acceptables (tableau 5.4) ;
- témoigner son insuffisance pour détecter l'émotion force de caractère du sexe féminin (tableau 5.5) ;
- prouver sa défaillance pour reconnaître l'émotion confiance sexe féminin a cause de EER =30,74% et taux Reco Max =69,49% (tableau 5.5).

▪ **Observation pour le paramètre P6 (Variation du signal)**

Le paramètre de variation du signal montre les points suivants :

- ce paramètre assure une identification pertinente pour l'émotion colère dans le cas de locuteur féminin avec un taux de reconnaissance maximal qui est de 86,44% et un taux d'erreur EER de 14,83% (tableau 5.5) ;
- ce paramètre donne une identification incertaine pour les émotions agitation et force de caractère du sexe féminin (tableau 5.5) ;
- il donne des résultats d'identification non fiables pour les quatre émotions dans le cas de locuteur masculin, puisque l'erreur EER est considérable (tableau 5.4) ;
- il donne des résultats défailants pour l'émotion confiance sexe féminin (tableau 5.5).

▪ Observation pour le paramètre P7 (Jitter)

Pour les locuteurs féminins ce paramètre indique les observations suivantes :

- ce paramètre est inefficace pour reconnaître l'émotion colère avec un taux de reconnaissance maximal de 84,74% et un taux EER de 34,47% (tableau 5.5) ;
- il est déficient pour identifier l'émotion agitation, confiance et force de caractère du sexe féminin (tableau 5.5) ;

Pour les locuteurs masculins ce paramètre présente une défaillance totale vu les taux EER qui sont entre 54,51% et 61,37% (tableau 5.4).

▪ Observation pour le paramètre P8 (Nombre de pauses)

Le nombre de pauses dans un signal a révélé les points suivants :

- ce paramètre est capable d'identifier l'émotion colère pour les locuteurs masculins avec un taux de reconnaissance maximal qui est de 91,20% et un taux d'erreur EER de 19,72% (tableau 5.4) ;
- le paramètre P8 est incapable de reconnaître l'émotion confiance pour les locuteurs masculins avec EER de 33,45% et Taux Reco Max de 67,03% (tableau 5.4) ;
- il est incapable de reconnaître les émotions agitation et fort (de caractère) sexe masculin avec EER de 25,63% et 31,53% respectivement, et Taux Reco Max de 85,71% et 80,67% respectivement (tableau 5.4) ;
- il est défaillant pour la détection des émotions agitation, confiance et force de caractère dans le cas de locuteur féminin vu leur grand taux d'erreur EER de 41,98%, 27,35% et 42,72% respectivement et leur faible taux de reconnaissance maximaux de 69,49%, 64,40% et 77,96% respectivement (tableau 5.5) ;
- il est non fiable pour identifier l'émotion colère du sexe féminin vu le taux d'erreur EER de 32,96% (tableau 5.5) ;

▪ **Observation pour le paramètre P9 (Durée totale des pauses)**

Le taux d'erreur EER a montré les qualités suivantes :

- le taux d'erreur EER varie de 23,94% et 34,99% il nous informe que les émotions agitation, colère et force (de caractère) sont mal identifiées avec ce paramètre pour le cas de locuteur masculin (tableau 5.4) ;
- ce taux est égal à 38,53% et le taux de reconnaissance maximal de 65,93% nous confirme que l'émotion confiance n'est pas identifiée avec ce paramètre pour le cas de locuteur masculin (tableau 5.4) ;
- les taux d'erreur EER des deux émotions colère et confiance du sexe féminin (22,52% et 25,51%), nous assure la non fiabilité de leurs identifications par ce paramètre (tableau 5.5) ;
- les taux EER des deux émotions agitation et force (de caractère) du sexe féminin (44,35% et 49,16%) et les deux taux Reco Max (69,49% et 77,96%) affirme la défaillance de ce paramètre dans la reconnaissance de ces émotions (tableau 5.5).

▪ **Observation pour le paramètre P10 (Maximum de pauses)**

En ce qui concerne le paramètre maximum de pauses :

- il exprime sa non fiabilité pour l'identification des trois émotions agitation, colère et force (de caractère) pour le sexe masculin (tableau 5.4) ;
- il présente une défaillance pour la reconnaissance de l'émotion confiance sexe masculin ; de même pour les deux émotions agitation et force (de caractère) du sexe féminin (tableaux 5.4 et 5.5) ;
- il est non fiable pour la détection des deux émotions colère, confiance du sexe féminin (tableau 5.5).

- **Observation pour le paramètre P11 (Moyenne des pauses)**

La moyenne des pauses est un paramètre pertinent dans la discrimination de l'émotion colère sexe masculin avec un taux de reconnaissance maximal qui est de 91.20% et un taux d'erreur EER de 19.72% (tableau 5.4). Ce paramètre n'est pas fiable pour la reconnaissance des émotions agitation et force de caractère du sexe masculin ; ainsi pour les deux émotions colère et force de caractère du sexe féminin (tableaux 5.4 et 5.5). Il est défaillant pour l'identification de l'émotion confiance sexe masculin de même pour les deux émotions agitation et confiance sexe féminin (tableaux 5.4 et 5.5).

V.12. Mesure et Test de Signification

V.12.1 Mesure de Dépendance entre les Variables

- **Notion d'Indépendance**

L'indépendance est un terme qui est souvent utilisé, et qui peut avoir plusieurs sens, on parle de dépendance ou d'indépendance statistique pour caractériser le lien entre deux (ou plusieurs) variables aléatoires (Allano, 2009).

Nos deux variables (paramètre, émotion) sont dites indépendantes lorsque la connaissance de l'une n'apporte rien sur la connaissance de l'autre. Autrement dit la connaissance d'un paramètre n'apporte rien à la détection d'une émotion.

Par le biais d'une mesure de dépendance (corrélation) appliquée à travers la mesure du niveau de consistance statistique de dépendance, qui permet de déterminer si les deux variables sont indépendantes ou non ; la connaissance de niveau de dépendance existant permet d'avoir la possibilité d'utiliser le paramètre (extrait d'un fichier de parole) pour renforcer la détection d'une émotion. Pour pouvoir estimer cette dépendance on définit la mesure de consistance de celle-ci.

V.12.2 Mesure de Dépendance

La mesure de dépendance entre deux variables est obtenue en effectuant les étapes suivantes :

- 1 - l'élaboration d'un tableau qui regroupe les deux variables ;
- 2 - le calcul du degré de liberté ;
- 3 - le calcul de la valeur globale de la χ^2 ;
- 4 - le calcul de la probabilité.

V.12.3. Représentation des deux Variables :

Les deux variables aléatoires doivent être représentées sous forme de tableau à deux entrées :

- l'émotion et sa négation sont indiquées en colonnes ;
- Le paramètre \geq au seuil_{EER} et $<$ au seuil_{EER} est représenté en ligne (tableau 5.6).

Tableau 5.6 : Représentant les deux variables.

Paramètre	Emotion		Totaux
	Confidence	Non confiance	
Paramètre \geq seuil _{EER}	A	B	C = A + B
Paramètre $<$ seuil _{EER}	D	E	F= D + E
Totaux	G= A+D	H=B + E	I=C+F= G+H

Les nombres de locuteurs sont repartis par classes d'émotion, selon qu'ils soient émotionnés ou non émotionnés; aussi bien que par leur position de valeur de paramètre, par rapport à la valeur du seuil correspondant au EER quelle soit supérieur ou égale au seuil, ou quelle soit strictement inférieur à ce seuil.

A : désigne le nombre entier de locuteurs émotionnés, et concrétise l'égalité ou la supériorité du paramètre étudié par rapport au seuil correspondant.

B : désigne le nombre entier de locuteurs non émotionnés et concrétise l'égalité ou la supériorité du paramètre étudié par rapport au seuil correspondant.

D : désigne le nombre entier de locuteurs émotionnés et concrétise l'infériorité du paramètre étudié par rapport au seuil correspondant.

E : désigne le nombre entier de locuteurs non émotionnés et concrétise l'infériorité du paramètre étudié par rapport au seuil correspondant (Contingency Table, web).

V.12.4. Méthode de Calcul de la Valeur Globale de la χ^2

Elle dépend des nombres particuliers de lignes et de colonnes. Elle est connue par convention sous le nom 'chi 2', dont la notation symbolique est χ^2 . Dans le cas des deux variables pour chaque colonne, on calcule la valeur de la composante individuelle de χ^2 . La valeur finale de χ^2 est obtenue par la somme des valeurs composantes individuelles de χ^2 (Inferential Statistics, web), (Distribution χ^2 , web) (Corrélation non paramétrique, web).

V.12.5. Calcul de la Probabilité

La probabilité entre deux variables correspondant à χ^2 est symbolisée par **P**. Plusieurs sites sur internet donnent directement les valeurs de la χ^2 et de la probabilité **P**, en introduisant le nombre de colonnes et de lignes et aussi les valeurs A, B, D et E définies précédemment.

V.12.6. Estimation de Dépendance

Pour pouvoir estimer la dépendance ou l'indépendance de deux variables, il faut s'articuler sur la valeur de la probabilité **P** et la comparer à une valeur critique de **P** choisie par les chercheurs qui est égale à 5%. Si la probabilité entre deux variables **P** est inférieure à cette valeur critique, les deux variables sont dites dépendantes ; sinon elles sont indépendantes.

Ainsi, nous exposons les résultats de la χ^2 et **P**, pour les différentes émotions en utilisant les 11 paramètres, dans les tableaux 5.7 à 5.84.

Tableau 5.7: Tableau de Contingence pour l'émotion agitation et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P1 \geq seuil_{EER}	9	24	33	7.13	0.0076
P1 < seuil_{EER}	4	54	58		
Totaux	13	78	91		

Tableau 5.8: Tableau de Contingence pour l'émotion agitation et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P1 \geq seuil_{EER}	13	13	26	5.38	0.0204
P1 < seuil_{EER}	7	26	33		
Totaux	20	39	59		

Rappelons que :

- P1 est la variation de F0.

Tableau 5.9: Tableau de Contingence pour l'émotion agitation et le paramètre P2, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P2 \geq seuil_{EER}	7	34	41	0.47	0.493
P2 < seuil_{EER}	6	44	50		
Totaux	13	78	91		

Tableau 5.10 : Tableau de Contingence pour l'émotion agitation et le paramètre P2, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P2 \geq seuil_{EER}	10	20	30	0.01	0.9203
P2 < seuil_{EER}	10	19	29		
Totaux	20	39	59		

Tableau 5.11 : Tableau de Contingence pour l'émotion agitation et le paramètre P3, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P3 \geq seuil_{EER}	9	37	46	2.12	0.1454
P3 < seuil_{EER}	4	41	45		
Totaux	13	78	91		

Tableau 5.12 : Tableau de Contingence pour l'émotion agitation et le paramètre P3, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P3 \geq seuil_{EER}	11	21	32	0.01	0.9203
P3 < seuil_{EER}	9	18	27		
Totaux	20	39	59		

Rappelons que :

- P2 est le maximum de dérivée de Fo ;
- P3 est minimum de dérivée.

Tableau 5.13 : Tableau de Contingence pour l'émotion agitation et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P4 \geq seuil_{EER}	10	20	30	13.26	0.0003
P4 < seuil_{EER}	3	58	61		
Totaux	13	78	91		

Tableau 5.14: Tableau de Contingence pour l'émotion agitation et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P4 \geq seuil_{EER}	15	8	23	16.5	<.0001
P4 < seuil_{EER}	5	31	36		
Totaux	20	39	59		

Tableau 5.15 : Tableau de Contingence pour l'émotion agitation et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P5 \geq seuil_{EER}	10	21	31	12.4	0.0004
P5 < seuil_{EER}	3	57	60		
Totaux	13	78	91		

Tableau 5.16 : Tableau de Contingence pour l'émotion agitation et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P5 \geq seuil_{EER}	16	5	21	26.03	<.0001
P5 < seuil_{EER}	4	34	38		
Totaux	20	39	59		

Rappelons que :

- P4 est la puissance ;
- P5 est le maximum de la puissance.

Tableau 5.17 : Tableau de Contingence pour l'émotion agitation et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P6 \geq seuil_{EER}	10	22	32	11.6	0.0007
P6 < seuil_{EER}	3	56	59		
Totaux	13	78	91		

Tableau 5.18 : Tableau de Contingence pour l'émotion agitation et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P6 \geq seuil_{EER}	11	12	23	3.26	0.071
P6 < seuil_{EER}	9	27	36		
Totaux	20	39	59		

Tableau 5.19 : Tableau de Contingence pour l'émotion agitation et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P7 \geq seuil_{EER}	6	47	53	1.02	0.3125
P7 < seuil_{EER}	7	30	37		
Totaux	13	77	90		

Tableau 5.20 : Tableau de Contingence pour l'émotion agitation et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P7 \geq seuil_{EER}	8	18	26	0.2	0.6547
P7 < seuil_{EER}	12	21	33		
Totaux	20	39	59		

Rappelons que :

- P6 est la variation du signal en db ;
- P7 est le Jitter (perturbation de Fo).

Tableau 5.21 : Tableau de Contingence pour l'émotion agitation et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P8 \geq seuil_{EER}	6	12	18	6.65	0.0099
P8 < seuil_{EER}	7	66	73		
Totaux	13	78	91		

Tableau 5.22 : Tableau de Contingence pour l'émotion agitation et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P8 \geq seuil	9	7	16	4.89	0.027
P8 < seuil	11	32	43		
Totaux	20	39	59		

Tableau 5.23 : Tableau de Contingence pour l'émotion agitation et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P9 \geq seuil	6	12	18	6.65	0.0099
P9 < seuil	7	66	73		
Totaux	13	78	91		

Tableau 5.24 : Tableau de Contingence pour l'émotion agitation et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P9 \geq seuil	9	7	16	4.89	0.027
P9 < seuil	11	32	43		
Totaux	20	39	59		

Rappelons que :

- P8 est le nombre de pauses ;
- P9 est la durée totale des pauses.

Tableau 5.25 : Tableau de Contingence pour l'émotion agitation et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe: M)

	Agitation	Non Agitation	Totaux	χ^2	P
P10 \geq seuil	6	12	18	6.65	0.0099
P10 < seuil	7	66	73		
Totaux	13	78	91		

Tableau 5.26 : Tableau de Contingence pour l'émotion agitation et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	Agitation	Non Agitation	Totaux	χ^2	P
P10 \geq seuil	9	7	16	4.89	0.027
P10 < seuil	11	32	43		
Totaux	20	39	59		

Tableau 5.27 : Tableau de Contingence pour l'émotion agitation et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Agitation	Non Agitation	Totaux	χ^2	P
P11 \geq seuil	6	12	18	6.65	0.0099
P11 < seuil	7	66	73		
Totaux	13	78	91		

Tableau 5.28 : Tableau de Contingence pour l'émotion agitation et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Agitation	Non Agitation	Totaux	χ^2	P
P11 \geq seuil	9	7	16	4.89	0.027
P11 < seuil	11	32	43		
Totaux	20	39	59		

Rappelons que :

- P10 est la durée maximale de pauses ;
- P11 est la durée moyenne des pauses.

Tableau 5.29 : Tableau de Contingence pour l'émotion colère et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Colère	Non colère	Totaux	χ^2	P
P1 \geq seuil_{EER}	6	27	33	5.69	0.0171
P1 < seuil_{EER}	2	56	58		
Totaux	8	83	91		

Tableau 5.30: Tableau de Contingence pour l'émotion colère et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	colère	Non colère	Totaux	χ^2	P
P1 \geq seuil_{EER}	5	18	23	3.52	0.0606
P1 < seuil_{EER}	2	34	36		
Totaux	7	52	59		

Tableau 5.31 : Tableau de Contingence pour l'émotion colère et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	colère	Non colère	Totaux	χ^2	P
P4 \geq seuil_{EER}	7	23	30	11.8	0.0006
P4 < seuil_{EER}	1	60	61		
Totaux	8	83	91		

Tableau 5.32 : Tableau de Contingence pour l'émotion colère et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	colère	Non colère	Totaux	χ^2	P
P4 \geq seuil_{EER}	6	17	23	7.29	0.0069
P4 < seuil_{EER}	1	35	36		
Totaux	7	52	59		

Rappelons que :

- P2 n'apporte rien pour la reconnaissance de l'émotion colère ;
- P3 n'apporte rien pour la reconnaissance de l'émotion colère.

Tableau 5.33: Tableau de Contingence pour l'émotion colère et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	colère	Non colère	Totaux	χ^2	P
P5 \geq seuil_{EER}	6	20	26	9.26	0.0023
P5 < seuil_{EER}	2	63	65		
Totaux	8	83	91		

Tableau 5.34 : Tableau de Contingence pour l'émotion colère et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	colère	Non colère	Totaux	χ^2	P
P5 \geq seuil_{EER}	6	10	16	13.8	0.0002
P5 < seuil_{EER}	1	42	43		
Totaux	7	52	59		

Tableau 5.35: Tableau de Contingence pour l'émotion colère et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	colère	Non colère	Totaux	χ^2	P
P6 \geq seuil_{EER}	6	20	26	9.26	0.0023
P6 < seuil_{EER}	2	63	65		
Totaux	8	83	91		

Tableau 5.36 : Tableau de Contingence pour l'émotion colère et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	colère	Non colère	Totaux	χ^2	P
P6 \geq seuil_{EER}	6	8	14	16.86	<.0001
P6 < seuil_{EER}	1	44	45		
Totaux	7	52	59		

Tableau 5.37 : Tableau de Contingence pour l'émotion colère et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	colère	Non colère	Totaux	χ^2	P
P7 \geq seuil_{EER}	3	50	53	1.66	0.1976
P7 < seuil_{EER}	5	32	37		
Totaux	8	82	90		

Tableau 5.38 : Tableau de Contingence pour l'émotion colère et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	colère	Non colère	Totaux	χ^2	P
P7 \geq seuil_{EER}	5	21	26	2.41	0.1206
P7 < seuil_{EER}	2	31	33		
Totaux	7	52	59		

Tableau 5.39: Tableau de Contingence pour l'émotion colère et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	colère	Non colère	Totaux	χ^2	P
P8 \geq seuil	6	12	18	16.86	<.0001
P8 < seuil	2	71	73		
Totaux	8	83	91		

Tableau 5.40 : Tableau de Contingence pour l'émotion colère et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	colère	Non colère	Totaux	χ^2	P
P8 \geq seuil	4	12	16	3.62	0.0571
P8 < seuil	3	40	43		
Totaux	7	52	59		

Tableau 5.41 : Tableau de Contingence pour l'émotion colère et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe: M)

	colère	Non colère	Totaux	χ^2	P
P9 \geq seuil	6	12	18	16.86	<.0001
P9<seuil	2	71	73		
Totaux	8	83	91		

Tableau 5.42: Tableau de Contingence pour l'émotion colère et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	colère	Non colère	Totaux	χ^2	P
P9 \geq seuil	4	12	16	3.62	0.0571
P9<seuil	3	40	43		
Totaux	7	52	59		

Tableau 5.43: Tableau de Contingence pour l'émotion colère et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe: M)

	colère	Non colère	Totaux	χ^2	P
P10 \geq seuil	6	12	18	16.86	<.0001
P10<seuil	2	71	73		
Totaux	8	83	91		

Tableau 5.44: Tableau de Contingence pour l'émotion colère et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	colère	Non colère	Totaux	χ^2	P
P10 \geq seuil	4	12	16	3.62	0.0571
P10<seuil	3	40	43		
Totaux	7	52	59		

Tableau 5.45: Tableau de Contingence pour l'émotion colère et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	colère	Non colère	Totaux	χ^2	P
P11 \geq seuil	6	12	18	16.86	<.0001
P11<seuil	2	71	73		
Totaux	8	83	91		

Tableau 5.46 : Tableau de Contingence pour l'émotion colère et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	colère	Non colère	Totaux	χ^2	P
P11 \geq seuil	4	12	16	3.62	0.0571
P11<seuil	3	40	43		
Totaux	7	52	59		

Tableau 5.47 : Tableau de Contingence pour l'émotion confiance et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P1 \geq seuil_{EER}	25	15	40	15.71	<.0001
P1< seuil_{EER}	11	40	51		
Totaux	36	55	91		

Tableau 5.48 : Tableau de Contingence pour l'émotion confiance et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P1 \geq seuil_{EER}	1	0	1	0	1
P1< seuil_{EER}	0	0	0		
Totaux	1	0	0		

Tableau 5.49: Tableau de Contingence pour l'émotion confiance et le paramètre P2, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P2 \geq seuil_{EER}	12	11	23	2.05	0.1522
P2 < seuil_{EER}	24	44	68		
Totaux	36	55	91		

Tableau 5.50: Tableau de Contingence pour l'émotion confiance et le paramètre P2, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P2 \geq seuil_{EER}	14	16	30	0.43	0.512
P2 < seuil_{EER}	16	13	29		
Totaux	30	29	59		

Tableau 5.51 : Tableau de Contingence pour l'émotion confiance et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P4 \geq seuil_{EER}	28	13	41	0.512	<.0001
P4 < seuil_{EER}	8	42	50		
Totaux	36	55	91		

Tableau 5.52: Tableau de Contingence pour l'émotion confiance et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P4 \geq seuil_{EER}	22	10	32	8.01	0.0047
P4 < seuil_{EER}	9	19	28		
Totaux	31	29	60		

Rappelons que :

- P2 n'apporte rien pour la reconnaissance de l'émotion confiance ;
- P3 n'apporte rien pour la reconnaissance de l'émotion confiance.

Tableau 5.53: Tableau de Contingence pour l'émotion confiance et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P5 \geq seuil_{EER}	30	14	44	29.19	<.0001
P5 < seuil_{EER}	6	41	47		
Totaux	36	55	91		

Tableau 5.54: Tableau de Contingence pour l'émotion confiance et le paramètre P5, ainsi que les valeurs de χ^2 et **P** sexe: F)

	Confidence	Non Confidence	Totaux	χ^2	P
P5 \geq seuil_{EER}	19	9	28	6.17	0.013
P5 < seuil_{EER}	11	20	31		
Totaux	30	29	59		

Tableau 5.55 : Tableau de Contingence pour l'émotion confiance et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P6 \geq seuil_{EER}	26	17	43	14.9	0.0001
P6 < seuil_{EER}	10	38	48		
Totaux	36	55	91		

Tableau 5.56 : Tableau de Contingence pour l'émotion confiance et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P6 \geq seuil_{EER}	17	10	27	2.92	0.0875
P6 < seuil_{EER}	13	19	32		
Totaux	30	29	59		

Tableau 5.57 : Tableau de Contingence pour l'émotion confiance et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P7 \geq seuil_{EER}	19	34	53	0.93	0.3349
P7 < seuil_{EER}	17	20	37		
Totaux	36	54	90		

Tableau 5.58 : Tableau de Contingence pour l'émotion confiance et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P7 \geq seuil_{EER}	14	12	26	0.17	0.6801
P7 < seuil_{EER}	16	17	33		
Totaux	30	29	59		

Tableau 5.59 : Tableau de Contingence pour l'émotion confiance et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P8 \geq seuil_{EER}	19	13	32	8.1	0.0044
P8 < seuil_{EER}	17	42	59		
Totaux	36	55	91		

Tableau 5.60: Tableau de Contingence pour l'émotion confiance et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P8 \geq seuil_{EER}	26	12	38	13.19	0.0003
P8 < seuil_{EER}	4	17	21		
Totaux	30	29	59		

Tableau 5.61: Tableau de Contingence pour l'émotion confiance et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P9 \geq seuil_{EER}	19	16	35	5.16	0.0231
P9 < seuil_{EER}	17	39	56		
Totaux	36	55	91		

Tableau 5.62: Tableau de Contingence pour l'émotion confiance et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	Confidence	Non Confidence	Totaux	χ^2	P
P9 \geq seuil_{EER}	28	12	40	18.23	<.0001
P9 < seuil_{EER}	2	17	19		
Totaux	30	29	59		

Tableau 5.63: Tableau de Contingence pour l'émotion confiance et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P10 \geq seuil_{EER}	10	8	18	2.4	0.1213
P10 < seuil_{EER}	26	47	73		
Totaux	36	55	91		

Tableau 5.64: Tableau de Contingence pour l'émotion confiance et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	Confidence	Non Confidence	Totaux	χ^2	P
P10 \geq seuil_{EER}	26	9	35	18.91	<.0001
P10 < seuil_{EER}	4	20	24		
Totaux	30	29	59		

Tableau 5.65 : Tableau de Contingence pour l'émotion confiance et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Confidence	Non Confidence	Totaux	χ^2	P
P11 \geq seuil_{EER}	0	18	18	91	<.0001
P11 < seuil_{EER}	73	0	73		
Totaux	73	18	91		

Tableau 5.66 : Tableau de Contingence pour l'émotion confiance et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Confidence	Non Confidence	Totaux	χ^2	P
P11 \geq seuil_{EER}	26	15	41	8.49	0.0036
P11 < seuil_{EER}	4	14	18		
Totaux	30	29	59		

Tableau 5.67: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P1 \geq seuil_{EER}	17	23	40	6.78	0.0092
P1 < seuil_{EER}	9	42	51		
Totaux	26	65	91		

Tableau 5.68: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P1, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P1 \geq seuil_{EER}	10	13	23	10.09	0.0015
P1 < seuil_{EER}	3	33	36		
Totaux	13	46	59		

Tableau 5.69 : Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P4 \geq seuil_{EER}	20	11	31	29.77	<.0001
P4 < seuil_{EER}	6	54	60		
Totaux	26	65	91		

Tableau 5.70 : Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P4, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P4 \geq seuil_{EER}	12	11	23	19.93	<.0001
P4 < seuil_{EER}	1	35	36		
Totaux	13	46	59		

Tableau 5.71: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P5 \geq seuil_{EER}	16	15	31	12.23	0.0005
P5 < seuil_{EER}	10	50	60		
Totaux	26	65	91		

Tableau 5.72 : Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P5, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P5 \geq seuil_{EER}	10	11	21	12.42	0.0004
P5 < seuil_{EER}	3	35	38		
Totaux	13	46	59		

Rappelons que :

- P2 n'apporte rien pour la reconnaissance de l'émotion force (de caractère) ;
- P3 n'apporte rien pour la reconnaissance de l'émotion force (de caractère).

Tableau 5.73: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P6 \geq seuil_{EER}	18	22	40	9.44	0.0021
P6 < seuil_{EER}	8	43	51		
Totaux	26	65	91		

Tableau 5.74: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P6, ainsi que les valeurs de χ^2 et **P** (sexe: F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P6 \geq seuil_{EER}	10	9	19	15.27	<.0001
P6 < seuil_{EER}	3	37	40		
Totaux	13	46	59		

Tableau 5.75: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P7 \geq seuil_{EER}	12	41	53	2.45	0.1175
P7 < seuil_{EER}	14	23	37		
Totaux	26	64	90		

Tableau 5.76: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P7, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P7 \geq seuil_{EER}	5	21	26	0.21	0.6468
P7 < seuil_{EER}	8	25	33		
Totaux	13	46	59		

Tableau 5.77: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P8 \geq seuil _{EER}	14	4	18	26.62	<.0001
P8 < seuil _{EER}	12	61	73		
Totaux	26	65	91		

Tableau 5.78: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P8, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P8 \geq seuil _{EER}	5	11	16	1.09	0.2965
P8 < seuil _{EER}	8	35	43		
Totaux	13	46	59		

Tableau 5.79 : Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P9 \geq seuil _{EER}	14	4	18	26.62	<.0001
P9 < seuil _{EER}	12	61	73		
Totaux	26	65	91		

Tableau 5.80: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P9, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P9 \geq seuil _{EER}	5	11	16	1.09	0.2965
P9 < seuil _{EER}	8	35	43		
Totaux	13	46	59		

Tableau 5.81: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P10 \geq seuil_{EER}	14	4	18	26.62	<.0001
P10 < seuil_{EER}	12	61	73		
Totaux	26	65	91		

Tableau 5.82: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P10, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P10 \geq seuil_{EER}	5	11	16	1.09	0.2965
P10 < seuil_{EER}	8	35	43		
Totaux	13	46	59		

Tableau 5.83 : Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : M)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P11 \geq seuil_{EER}	0	18	18	91	<.0001
P11 < seuil_{EER}	73	0	73		
Totaux	73	18	91		

Tableau 5.84: Tableau de Contingence pour l'émotion force (de caractère) et le paramètre P11, ainsi que les valeurs de χ^2 et **P** (sexe : F)

	Fort (de caractère)	Faible (de caractère)	Totaux	χ^2	P
P11 \geq seuil_{EER}	5	11	16	1.09	0.2965
P11 < seuil_{EER}	8	35	43		
Totaux	13	46	59		

V.12.7 Récapitulatif des Résultats de la Probabilité **P** et χ^2

Les valeurs de la probabilité **P** et χ^2 trouvées pour chaque émotion sont résumées dans les tableaux 5.29 et 5.30 suivants :

Tableau 5.29 : Récapitulatif des résultats de la probabilité **P** et χ^2 pour le sexe masculin.

	Agitation		Colère		Confidence		Fort (de caractère)	
	P %	χ^2	P %	χ^2	P %	χ^2	P %	χ^2
P1	0.76	7.13	1.71	5.69	<0.01	15.71	0.92	6.78
P2	49.3	0.47	/	/	15.22	2.05	/	/
P3	14.54	2.12	/	/	/	/	/	/
P4	0.03	13.26	0.06	11.8	<0.01	0.512	<0.01	29.77
P5	0.04	12.4	0.23	9.26	<0.01	29.19	0.05	12.23
P6	0.07	11.6	0.23	9.26	0.01	14.9	0.21	9.44
P7	31.25	1.02	19.76	1.66	33.49	0.93	11.75	2.45
P8	0.99	6.65	<0.01	16.86	0.44	8.1	<0.01	26.62
P9	0.99	6.65	<0.01	16.86	2.31	5.16	<0.01	26.62
P10	0.99	6.65	<0.01	16.86	12.13	2.4	<0.01	26.62
P11	0.99	6.65	<0.01	16.86	<0.01	91	<0.01	91

D'après le tableau 5.29, nous remarquons que :

Les paramètres donnant des valeurs de la probabilité **P** inférieures à la valeur critique de **P** (5%) dans le cas du locuteur masculin sont :

- P4 et P5 qui déclarent leur dépendance avec l'émotion agitation ;
- P8, P9, P10 et P11 qui déclarent leur dépendance avec l'émotion colère ;
- P1, P4, P5, P6 et P11 qui déclarent leur dépendance avec l'émotion confiance ;
- P4, P8, P9, P10 et P11 qui déclarent leur dépendance avec l'émotion force de caractère.

Tableau 5.30 : Récapitulatif des résultats de la probabilité **P** et χ^2 pour le sexe féminin

	Agitation		Colère		Confidence		Fort (de caractère)	
	P %	χ^2	P %	χ^2	P %	χ^2	P %	χ^2
P1	2.04	5.38	6.06	3.52	100	0	0.15	10.09
P2	92.03	0.01	/	/	51.2	0.43	/	/
P3	92.03	0.01	/	/	/	/	/	/
P4	<0.01	16.5	0.69	7.29	0.47	8.01	<0.01	19.93
P5	<0.01	26.03	0.02	13.8	1.3	6.17	0.04	12.42
P6	7.1	3.26	<0.01	16.86	8.75	2.92	<0.01	15.27
P7	65.47	0.2	12.06	2.41	68.01	0.17	64.68	0.21
P8	2.7	4.89	5.71	3.62	0.03	13.19	29.65	1.09
P9	2.7	4.89	5.71	3.62	<0.01	18.23	29.65	1.09
P10	2.7	4.89	5.71	3.62	<0.01	18.91	29.65	1.09
P11	2.7	4.89	5.71	3.62	0.36	8.49	29.65	1.09

D'après le tableau **5.30**, nous remarquons que :

Les paramètres donnant des valeurs de la probabilité **P** inférieures à la valeur critique de **P** (5 %) dans le cas du locuteur féminin sont :

- P4 et P5 qui déclarent leur dépendance avec l'émotion agitation ;
- P6 qui déclare sa dépendance avec l'émotion colère ;
- P8, P9, P10 qui déclarent leur dépendance avec l'émotion confiance ;
- P4, P5, P6 qui déclarent leur dépendance avec l'émotion force de caractère.

Rappelons que / : indique que le paramètre correspondant n'apporte rien à la détection de l'émotion.

V.13. Rapprochement des Résultats d'Identification Emotionnelle et ceux de la Dépendance

Les résultats d'identification émotionnelle et ceux de la dépendance, confirment la pertinence du paramètre suivants

- P4 (puissance) pour l'identification de l'émotion force (de caractère) dans le cas du locuteur masculin et féminin ;
- P5 (maximum de puissance) pour l'identification de l'émotion agitation sexe féminin ;
- P6 (variation du signal en db) pour l'identification de l'émotion colère dans le cas du locuteur féminin ;
- P8 (nombre de pauses) pour la reconnaissance de l'émotion colère sexe masculin ;
- P11 (durée moyenne des pauses) pour la détection de l'émotion colère sexe masculin.

A ce moment nous pouvons dire que les relations entre certains paramètres prosodiques du signal de parole et les émotions ne sont pas aléatoires et que le changement de l'état émotionnel du locuteur influe sur certains paramètres prosodiques.

V.14. Conclusion

Dans ce chapitre nous avons décrit les étapes réalisées, ainsi que les résultats obtenus au cours de notre recherche. Nous avons utilisé onze caractéristiques prosodiques pour l'identification de chaque émotion à partir des signaux de parole.

Les taux de reconnaissance et de l'erreur EER obtenus ainsi que les valeurs de χ^2 et de la probabilité P confirment l'identification de l'émotion agitation par le paramètre P5 pour les locuteurs femmes, et l'identification de l'émotion colère par le paramètre P6 pour les locuteurs femmes et par les paramètres P8 et P11 pour les locuteurs hommes. Ces taux obtenus confirment aussi l'identification de l'émotion force (de caractère) par le paramètre P4 pour les locuteurs femmes et hommes.

Chapitre VI : Conclusion Générale

Chapitre VI. Conclusion Générale

Ce travail de recherche a concerné la détection des états émotionnels par analyse du signal vocal du locuteur. Il s'agit d'identifier certaines émotions et certains traits de sa personnalité à partir des caractéristiques pertinentes de la parole. Pour ce faire, nous avons essayé plusieurs paramètres prosodiques : Pitch, Durée, Energie, ... et leurs combinés statistiques (maximum, minimum, variation, moyenne,...) en essayant d'employer différents types de seuillage.

Les tests d'évaluation ont été faits sur une base de données parlée universelle, appelée ORATOR et qui a été réalisée par des acteurs professionnels, parlant en Allemand.

Dans le but d'évaluer la consistance de nos résultats (*consistency measure*), des tests de χ^2 (*Chi-2*) avec mesure de probabilité de consistance liée à l'indépendance des variables de l'émotion, ont été entrepris sur chaque résultat obtenu.

Les résultats expérimentaux ont permis de déceler les paramètres pertinents pour chaque type d'émotion.

Par ailleurs, nous avons divisé l'analyse des émotions en 2 types : analyse des locuteurs masculins et analyse des locuteurs féminins (*locutrices*). Cette séparation nous a montré que les hommes ont un comportement différent des femmes quant à l'expression des émotions et de leur détection.

Les taux de reconnaissance obtenus dépassent les 80% de bonne identification dans beaucoup de cas ; ce qui est assez encourageant comme résultat.

Comme perspectives, nous proposons l'utilisation des traits du visage comme complément supplémentaire d'information, pouvant (*théoriquement*) améliorer la précision de détection des émotions dans de larges proportions.

Finalement, nous espérons que ce modeste travail contribuera au développement de la reconnaissance en général, et à la détection émotionnelle en particulier.

Références Bibliographiques

(Allano, 2009) L. Allano, « La Biométrie multimodale : stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles », thèse de Doctorat. Université des télécommunications D'EVRY- VAL D'ESSONNE, Paris Sud, 2009.

(Alvarez , 2007) A. Alvarez, I. Cearreta, J. M. Lopez, A. Arruti, E. Lazkano, B. Sierra et N. Garay, « A comparison using different speech parameters in the automatic emotion recognition using feature subset selection based on evolutionary algorithms ». In Proceedings IOth International Conference, TSD 2007, pp: 423-430. Pilsen, Czech Republic: The Institution of Engineering and Technology, 2007.

(Attabi, 2008) Y. Attabi, « Reconnaissance automatique des émotions à partir du signal acoustique », Montréal, thèse de Doctorat. Ecole de technologie supérieur. Université de Québec, 2008.

(Audibert, 2008) N. Audibert, « Prosodie de la parole expressive : dimensionnalité d'énoncés méthodologiquement contrôlés authentiques et actés », thèse de Doctorat. Institut polytechnique de Grenoble, 2008.

(Bänziger, 2004) T. Bänziger, D. Grandjean, P. J. Bernard, G. Klasmeyer & K. R. Scherer, « étude de l'encodage et du décodage », Université de Genève, 2004.
Tanja.Banziger@pse.unige.ch

(Beller, 2005) G. Beller, « étude et modèle génératif de l'expressivité dans la parole », Rapport de stage Master II, Université Pierre et Marie Curie, Paris VI, 2005.
<http://www.atiam.ircam.fr/Archives/Stages0405/Beller.pdf>

(Beller, 2009) G. Beller, « Analyse et modèle génératif de l'expressivité, application à la parole et à l'interprétation musicale », thèse de Doctorat. Université Paris VI Pierre et Marie Curie, 2009.

(Beritelli, 2007) F. Beritelli, C. Salvatore, R. Alessandra, S. Salvatore et E. Donato. « Speech emotion recognition using MFCCs extracted from a mobile terminal based on ETSI front end ». In International Conference on Signal Processing Proceedings, ICSP. Vol. 2, Guilin, China: Institute of Electrical and Electronics Engineers Inc. Piscataway, NJ 08855-1331, United States, 2007. <http://dx.doi.org/10.1109/ICQSP.2006.345670>

(Boite, 2000) R. Boite, H. Boulard, T. Dutoit, J. Hancq et H. Leich, « Traitement de la parole », 2000. <http://books.google.fr/books>.

(Boersma, 2005) P. Boersma et D. Weenink. Praat: « doing phonetics by computer », 2005, [computer program], <http://www.praat.org/>.
<http://www.fon.hum.uva.nl/praat/>

(Bulut, 2002) M. Bulut, S. Shrikanth, S.S. Narayanan, et A. K. Syrdal « Expressive speech synthesis using a concatenative synthesizer ». In ICSLP, ATT Labs-Research, Florham Park, NJ, 2002.

(Calliope, 1997) Calliope, « La parole et son traitement automatique » Dunod, 1997.

(Campos, 1984) J. Campos, et K. Barrett, « Toward a new understanding of emotions and their development » Cambridge University Press, 1984.

(Chung, 2000) S. Chung, « L'expression et la perception de l'émotion extraite de la parole spontanée », thèse de Doctorat, Université Paris III, Sorbonne Nouvelle, Institut de Linguistique et Phonétique Générales et Appliquées, 2000.

(Clavel, 2007) C. Clavel, « Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales », thèse de Doctorat École Nationale Supérieure des Télécommunications, 2007.
http://www.afcp-parole.org/doc/theses/these_CC07.pdf

(Dash, 1997) M. Dash, et H. Liu. « Feature sélection for classification ». Intelligent Data Analysis, vol. 1, n°3.

(Devillers, 2007) L. Devillers et L. Vidrascu, « Real-life emotion recognition in speech », Lecture Notes in Computer Science. Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, vol. 4441 NAI, pp: 34- 42, 2007.

(Dorizzi, 2004) B. Dorizzi, « Techniques et Usages Biométriques », GET/INT Evry2204 http://pari-stic.labri.fr/TUTORIAL/tutorial_BIO_1_PARISTIC_05.pdf
Bernadette.Dorizzi@int-evry.fr

(Dutoit, 2005) T. Dutoit « un bilan de développements récents de traitement automatique de la parole », Faculté polytechnique de Mons, 2005.

(El Ayadi, 2007) M. El Ayadi, S. Kamel et F. Karray, « Speech emotion recognition using Gaussian mixture vector autoregressive models ». In International Conference on Acoustics, Speech, and Signal Processing (IEEE Cat. No. 07CH37846), pp: 957-960. Honolulu, HI, USA: The Institution of Engineering and Technology, 2007.

(Gao, 2007) H. Gao, C. Shanguang et S. Guangchuan, « Emotion classification of mandarin speech based on TEO nonlinear features ». In Proceedings – SNPD 2007: Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. Vol. 3, pp: 394-398. Qingdao,China: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 2007.<<http://dx.doi.org/10.11Q9/SNPD.2007.267>>.

(Giripunje , 2007) S. Giripunje, et N. Bawane, « ANFIS based émotions recognition in speech ». In Knowledge-Based Intelligent Information and Engineering Systems: KES 2007 -WIRN 2007. Proceedings 11th International Conference, KES 2007. XVII Italian Workshop on Neural Networks. (Lecture Notes in Artificial Intelligence vol. 4692). Vol. I, pp: 77-84. Vietri sul Mare, Italy: The Institution of Engineering and Technology. 2007

(Huang, 2001) X. Huang, A. Alex et H. Hsiao-woen, « Spoken language processing: a guide to Theory, Algorithm, and System Development» United states of America: Prentice Hall PTR, 2001.

(Huang, 2006)Huang, Rongqing, et Changxue Ma, « Toward a speaker- independent real-time affect détection system ». In 18th International Conference on Pattern Recognition Hong Kong, China: The Institution of Engineering and Technology, 2006.

(Huang, 2008) Huang, Huiqin, Qi Luo et Aiqin Zhu, « Speech emofion recognition in web based service ». In International Conference on Communications, Circuits and Systems 2007, pp: 804-806. Kokura, Japan: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 2008.

(Inanoglu, 2005) Z. Inanoglu et C. Ron, « Emofive alert: HMM- based Emotion détection in voice mail messages », In International Conference on Intelligent User Interfaces, Proceedings IUI, pp: 251-253. San Diego, CA, United States: Associafion for Compufing Machinery, New York, NY 10036-5701, United States, 2005.

(Jones, 2007) C. M. Jones, et J. Ing-Marie, « Performance analysis of acoustic émotion recognition for in-car conversational interfaces », In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), PART 2, Vol. 4555 NCS, pp. 411-420. Beijing,China: Springer Verlag, Heidelberg, D-69121, Germany, 2007.

(Kleinginna, 1981) P.R. Kleinginna et A.M. Kleinginna, « A categorized list of emotion definitions with suggestions for a consensual definition. Motivation and Emotion», 5(4) pp:345–379, 1981.

(Lee, 2004) S. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, et S. Narayanan, « Emotion Recognition based on Phoneme Classes ». In ICSLP. Korea, 2004. < http://sail.usc.edu/publications/icslp04_chulminlee.pdf.

(Li, 2007) W. Li, Z. Yanhui et F. Yingzi, « Speech emotion recognition in E-learning System based on affective Computing ». In Proceedings - Third International Conference on Natural Computation, (ICNC), Vol. 5, pp: 809-813. Haikou, Hainan, China: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, 2007. <http://dx.doi.org/10.1109/IC.NC.2007.677>

Lieberman, P. (1961), «Perturbations in vocal pitch » J. Acoust. Soc. Am. 33, pp: 597-603.

(Panât, 2008) A. Panât, et V. Ingole, « Affective state analysis of speech for speaker verification: Experimental study, design and development », In Proceedings -International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2007. Vol. 1, pp: 255-261, Sivakasi, Tamil Nadu, India: Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 2008. <http://dx.doi.org/10.1109/ICCIMA.2007.59>.

(Pao, 2005) Pao, L. Tsang, C. Yu, J. Heng et W. Liao, « Detecting Emotions in Mandarin Speech ». Computational Linguistics and Chinese Language Processing, vol. 10, pp: 347-362, 2005.

(Pirker, 2007) H. Pirker, « Mixed feelings about using phoneme-level models in emotion recognition ». In Affective Computing and Intelligent Interaction Proceedings Second International Conference, AII 2007. (Lecture Notes in Computer Science vol. 4738), pp: 772-773. Lisbon, Portugal: The Institution of Engineering and Technology, 2007.

(Plutchik, 1980) R. Plutchik. « A General Psychoevolutionary Theory of Emotion », In Plutchik R. & Kellerman H. (eds.): Emotion: theory, research, and experience, New York: Academic, volume 1, pp: 3-33, 1980.

(Ringeval, 2008) F. Ringeval, M. Chetouani, « Une Approche Basée Voyelle pour la Reconnaissance d'Émotions Actées », Institut des Systèmes Intelligents et Robotique, 2008. fabien.ringeval@isir.fr mohamed.chetouani@upmc.fr

(Schuller, 2002) « Towards intuitive speech interaction by the integration of Emotional aspects », In IEEE International Conference on Systems, Man and Cybernetics. Conference Proceedings (Cat. No.02CH37349). Vol. 6, Yasmine Hammamet, Tunisia: IEEE, 2002. <<http://dx.doi.org/10.1109/ICSMC.2002.1175635>
http://www.uquebec.ca/edusante/mentale/imp_emotions.htm

(Schuller, 2007a) B. Schuller, D. Seppi, A. Batliner, A. Maier et S. Steidl, « Towards more reality in the recognition of emotional ». In IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE Cat. No. 07CH37846), pp: 941-944. Honolulu, HI, USA: The Institution of Engineering and Technology, 2007.

(Schuller, 2007b) B. Schuller, B. Vlasenko, R. Minguéz, G. Rigoll et A. Wendemuth, « Comparing one and two-stage acoustic modeling in the recognition of emotion in speech ». In IEEE Workshop on Automatic Speech Recognition and Understanding, pp: 596-600. Kyoto, Japan: The Institution of Engineering and Technology, 2007.

(Seppänen, 2003) T. Seppänen, V. Eero et T. Juhani, « Prosody-based classification of emotions in spoken Finnish ». In EUROSPEECH. Geneva, Switzerland, 2003.
<<http://www.mediateam.oulu.fi/publications/pdf/438.pdf>>.

(Sethu, 2007) V. Sethu, E. Ambikairajah et J. Epps, « Speaker normalisation for speech-based emotion detection ». In 15th International Conference on Digital Signal Processing, pp: 611-614. Cardiff, UK: The Institution of Engineering and Technology, 2007.

(Triki, 2007) A. Triki, « Développement d'un Système de Reconnaissance Robuste de la Parole », Mémoire de Magister, Institut d'électronique, Centre Universitaire Cheikh el-Arbi Tebessi Tebessa, 2007.

(Vlasenko, 2007) B. Vlasenko, B. Schuller, A. Wendemuth et G. Rigoll, « Frame vs. tum-level: emotion recognition from speech considering static and dynamic processing ». In Affective Computing and Intelligent Interaction. Proceedings Second International Conference, ACII 2007. (Lecture Notes in Computer Science vol. 4738), pp: 139-147. Lisbon, Portugal: The Institution of Engineering and Technology.

(Vogt, 2005) T. Vogt, et E. André, « Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition ». In IEEE International Conference on Multimedia and Expo. Amsterdam, Netherlands: The Institution of Engineering and Technology.

(Xie, 2007) B. Xie, C. Ling, C. Gen-Cai et C. Chun, « Feature selection for emotion recognition of mandarin speech ». Zhejiang Daxue Xuebao (Gongxue Ban)/Journal of Zhejiang University (Engineering Science), 2007. vol. 41, n° 11, pp. 1816-1822.

(Yacoub, 2003) S. Yacoub, S. Steve, L. Xiao et B. John, « Recognition of Emotions in Interactive Voice Response Systems ». In EUROSPEECH. Geneva, Switzerland, 2003.

(Yu, 2004) C. Yu et W. Allison, « Detecting user engagement in everyday conversations ». In In Proc. 8th International Conference on Spoken, 2004. Language Processing (ICSLP 2004). p. 1329-1332. Jeju Island; Korea.

(Zhu, 2007) A. Zhu et L. Qi, « Study on speech émotion récoognition system in E-learning », In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), PART 3, Vol. 4552 NCS, pp. 544-552. Beijing, China: Springer Verlag, Heidelberg, D-69121, Germany, 2007.

Sites web

(Contingency Table, web) <https://faculty.vassar.edu/lowry/newcs.html>

(Corrélation non paramétrique, web) <http://www.cons-dev.org/elearning/stat/stat7/st7.html>

(Distribution χ^2 , web) http://dmo.econ.msu.ru/teaching/stat/Khi_deux.htm

(GloCCal, web) <http://www.gloccal.fr/verification-et-identification-du-locuteur.html>

(Griaule biometrics, web) <http://www.griaulebiometrics.com/en-us/book/understanding-biometrics/evaluation/accuracy/matching/interest/equal>

(Inferential Statistics, web) <https://faculty.vassar.edu/lowry/webtext.html>

(Webopedia, web) http://www.webopedia.com/TERM/E/equal_error_rate.html

Dictionnaires

« Dictionnaire de la parole », A. Galarneau, P. Tremblay, and P. Martin.
<http://www.lli.ulaval.ca/labo2256/lexique/dico.html>. Laboratoire de Phonétique et Phonologie de l'Université Laval a Québec.

« 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 A. Rey, deuxième édition (1985), Le Robert, Paris.

ملخص

يهتم هذا البحث بدراسة الكشف عن عواطف و انفعالات المتحدث غضب قلق... من خلال خصائصه الصوتية من أجل هذا استخدمنا قاعدة البيانات عالمية تحوي تسجيلات باللغة الالمانية لإشارات صوتية للعواطف و الانفعالات مسجلة من طرف ممثلين ألمان.

لإنجاز هذا العمل اعتمدنا في دراستنا هذه على 11 من الوسائط العروضية والتي تم اختبارها كل على حدى لمعرفة عاطفة المتحدث متبوعا باختبار مدى ارتباط هذه الوسائط بعواطف و انفعالات المتحدث النتائج المتحصل عليها بينت نجاعة بعض الوسائط العروضية (مثل P4, P5, P6, P8, P11) في الكشف على بعض عواطف و انفعالات المتحدث.

Résumé

Dans cette étude nous nous intéressons à la détection des émotions du locuteur partir des caractéristiques vocales extra-linguistiques extraites de son signal de parole. Pour cela nous avons utilisé une base de données universelle de paroles émotionnelles allemande ORATOR simulées par des acteurs professionnels allemands.

Dans l'élaboration de ce travail nous avons basé notre étude sur 11 paramètres prosodiques, testés séparément pour la détection de l'émotion du locuteur suivis par des tests de signification de chaque paramètre.

Les résultats obtenus ont montré l'efficacité de certains paramètres (P4, P5, P6, P8, P11) dans la détection de certaine émotion du locuteur.

Abstract

This research work focuses on the detection of speaker emotion from his extra-linguistic vocal characteristics. For this purpose, we have used a universal database of German emotional speech "ORATOR" simulated by German professional actors.

In developing this work we have based our study on 11 prosodic parameters, tested separately for the detection of the speaker's emotion, followed by measurement and significance testing.

The results obtained showed the effectiveness of some parameters (P4, P5, P6, P8, P11) in the task of emotion detection.