
Étude des paramètres acoustiques de voix et de parole liés à la production orale d'émotions

Auteur : Simon, Marie

Promoteur(s) : Leclercq, Anne-Lise; Remacle, Angélique

Faculté : Faculté de Psychologie, Logopédie et Sciences de l'Éducation

Diplôme : Master en logopédie, à finalité spécialisée en voix

Année académique : 2023-2024

URI/URL : <http://hdl.handle.net/2268.2/19952>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Étude des paramètres acoustiques de voix et de parole liés à la production orale d'émotions

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION DU GRADE
DE MASTER EN LOGOPÉDIE

KHELIL ELLA & SIMON MARIE

PROMOTRICE : LECLERCQ ANNE-LISE

CO-PROMOTRICE : REMACLE ANGÉLIQUE

LECTRICES : HUBIN MARION & PIRON LÉONOR

Année académique 2023-2024

REMERCIEMENTS

Nous souhaitons tout d'abord remercier notre promotrice, Madame Anne-Lise Leclercq ainsi que notre co-promotrice Angélique Remacle pour leur encadrement et leur écoute bienveillante durant ces deux années de travail. Merci de nous avoir permis de réaliser notre mémoire dans de si bonnes conditions.

Merci également à Elodie Etienne de nous avoir permis de participer à son projet de thèse de doctorat et de nous avoir fait confiance dans la mise en place de cette collaboration.

Merci à nos lectrices, Mesdames Marion Hubin et Léonor Piron, d'avoir pris le temps de consulter ce travail et d'y avoir apporté vos réflexions et vos questionnements.

Merci également à toutes les personnes qui ont répondu volontiers à notre invitation pour participer au projet, mais aussi à celles qui l'ont partagée afin de nous aider à mettre sur pied nos séances de testings et d'enregistrement. Sans vous, ce mémoire n'aurait pas pu voir le jour. Nous souhaitons notamment remercier nos deux papas d'avoir accepté de réaliser les séances de pré-tests avec nous et d'avoir fait preuve de patience lors de celles-ci.

Merci à nos deux mamans, au papy d'Ella, ainsi qu'à Eric d'avoir pris le temps de relire et de commenter notre mémoire afin que sa forme soit la plus parfaite possible.

Nous souhaitons également nous remercier mutuellement, Marie & Ella, pour avoir réalisé ce projet ensemble de manière complémentaire, comme une véritable équipe. Nous nous remercions d'avoir pris le temps de nous écouter l'une et l'autre, et d'avoir rendu ce mémoire qui nous plait et nous correspond.

Enfin, il nous semblait indispensable de remercier Madame Dominique Morsomme, notre professeure de la finalité « voix », pour nous avoir tout appris sur la prise en soins de la voix parlée et chantée durant cette dernière année d'études. Merci de nous avoir donné le goût de ce métier et d'avoir éveillé en nous notre esprit critique et notre réflexion professionnelle qui, nous en sommes convaincues, nous ont permis de nous questionner et d'approfondir davantage nos connaissances pour réaliser ce mémoire.

Pour terminer, merci aussi à nos copines logopèdes : Sarah, Delphine, Caroline, Maëlle, Cécile, pour ne citer qu'elles, qui nous ont toujours soutenues et nous ont permis de rester motivées malgré les difficultés rencontrées durant ces cinq années d'études.

LISTE DES TABLEAUX ET FIGURES

Tableau 1. *Données issues de la littérature sur l'évolution des paramètres acoustiques avec les émotions de Paul Ekman.* **p.32**

Tableau 2. *Description de l'échantillon d'acteurs participant au projet « EVE ».* **p.38**

Tableau 3. *Liste des 10 phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020).* **p.39**

Tableau 4. *Définitions des 10 émotions intégrées dans la base de données émotionnelle « EVE » (Ekman, 2005 ; Geer & al., 2000 ; Hareli & al., 2018 ; Hess, 2003 ; Perry, 2011).* **p.42**

Tableau 5. *Description des intensités des 10 émotions intégrées dans la base de données émotionnelle « EVE » (Gournay & al., 2018).* **p.43**

Tableau 6. *Analyse delta de la médiane moyenne de la fréquence fondamentale lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.49**

Tableau 7. *Analyse delta de l'étendue moyenne de la fréquence fondamentale lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.50**

Tableau 8. *Analyse delta de la médiane moyenne de l'intensité lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.51**

Tableau 9. *Analyse delta de l'étendue moyenne de l'intensité lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.53**

Tableau 10. *Analyse delta du rapport L1-L0 moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.54**

Tableau 11. *Analyse delta du pourcentage de non-voisement moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.55**

Tableau 12. *Analyse delta du CPPs moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.57**

Tableau 13. *Analyse delta du débit moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).* **p.58**

Tableau 14. *Comparaison de nos résultats des deltas obtenus pour chaque paramètre acoustique étudié avec nos hypothèses initialement posées.* **p.62**

TABLE DES MATIÈRES

REMERCIEMENTS	1
LISTE DES TABLEAUX ET FIGURES	2
1. INTRODUCTION GÉNÉRALE	6
1.1 CONTEXTE DANS LEQUEL S'INSCRIT LE PROJET	6
1.2 DESCRIPTION DU PROJET GLOBAL DANS LEQUEL NOTRE MÉMOIRE S'INSCRIT	6
1.3 DESCRIPTION DE NOTRE MÉMOIRE DANS LE CADRE DE CE PROJET COLLABORATIF	7
2. INTRODUCTION THÉORIQUE	9
2.1 LES ÉMOTIONS	9
2.1.1 <i>Définition</i>	9
2.1.2 <i>Les 6 émotions universelles de Paul Ekman</i>	9
2.1.3 <i>L'expression des émotions au sein de la parole</i>	13
2.1.4 <i>L'identification des émotions</i>	15
2.1.5 <i>L'identification des émotions dans le signal de parole</i>	15
2.2 LA PRODUCTION VOCALE	17
2.2.1 <i>Définition</i>	17
2.2.2 <i>Les paramètres acoustiques de la voix</i>	17
2.3 LA FLUENCE DE LA PAROLE	20
2.3.1 <i>Définition</i>	20
2.3.2 <i>Les paramètres de la fluence de la parole</i>	21
2.4 L'IMPACT DES ÉMOTIONS SUR LES PARAMÈTRES DE LA VOIX ET DE LA FLUENCE	23
2.5 LES BASES DE DONNÉES ÉMOTIONNELLES	26
2.5.1 <i>Méthodologie de création des bases de données émotionnelles</i>	26
2.5.2 <i>Les bases de données émotionnelles existantes</i>	29
3. OBJECTIFS ET HYPOTHÈSES	31
3.1 OBJECTIFS	31
3.2 HYPOTHÈSES	31
3.2.1 <i>Hypothèses sur l'évolution de f_0 selon l'émotion produite</i>	33
3.2.2 <i>Hypothèses sur l'évolution de l'intensité avec la production d'émotions</i>	33
3.2.3 <i>Hypothèses sur l'évolution du LTAS avec la production d'émotions</i>	34
3.2.4 <i>Hypothèses sur l'évolution du CCPs avec la production d'émotions</i>	35
3.2.5 <i>Hypothèses sur l'évolution du débit de la parole avec la production d'émotions</i>	35
3.2.6 <i>Hypothèses sur l'évolution de la vitesse articulatoire avec la production d'émotions</i>	35
4. MÉTHODOLOGIE	37
4.1 PARTICIPANTS	37
4.2 CORPUS	38
4.3 MATÉRIEL	39
4.4 PROCÉDURE GÉNÉRALE	40
4.4.1 <i>Essais et définition des paramétrages avec des participants volontaires</i>	40
4.4.2 <i>Elaboration de la base de données avec les participants prenant part à l'étude</i>	40
4.4.3 <i>Description de la partie de la base de données analysée dans ce mémoire</i>	44
4.4.4 <i>Traitements des enregistrements</i>	44
4.4.5 <i>Analyse des paramètres acoustiques de la voix et de la parole des 4 participants</i>	45
5. RÉSULTATS	48
5.1 RÉSULTATS OBTENUS SUR L'ÉVOLUTION DES PARAMÈTRES DE LA VOIX AVEC LES ÉMOTIONS D'ÉKMAN	48
5.1.1 <i>La fréquence fondamentale</i>	48
5.1.2 <i>L'intensité</i>	51

5.1.3	<i>Long-Term Average Spectrum (LTAS)</i>	54
5.1.4	<i>Smoothed Cepstral Peak Prominence (CPPs)</i>	56
5.2	RÉSULTATS OBTENUS SUR L'ÉVOLUTION DES PARAMÈTRES DE LA FLUENCE AVEC LES ÉMOTIONS .	58
5.2.1	<i>Le débit</i>	58
5.2.2	<i>La vitesse articulatoire</i>	59
5.3	IDENTIFICATION DES PARAMÈTRES DE VOIX ET DE FLUENCE SPÉCIFIQUES À CHAQUE ÉMOTION ...	59
6.	DISCUSSION	63
6.1	INTERPRÉTATION DE NOS RÉSULTATS SELON LES HYPOTHÈSES INITIALEMENT POSÉES POUR LES PARAMÈTRES DE LA VOIX	63
6.1.1	<i>La fréquence fondamentale (f_0) - Hz</i>	63
6.1.2	<i>L'intensité - dB</i>	64
6.1.3	<i>Le LTAS - dB</i>	65
6.1.4	<i>Le CPPs - dB</i>	67
6.2	INTERPRÉTATION DE NOS RÉSULTATS SELON LES HYPOTHÈSES INITIALEMENT POSÉES POUR LES PARAMÈTRES DE LA FLUENCE	67
6.2.1	<i>Le débit de parole – syllabes/seconde</i>	67
6.2.2	<i>La vitesse articulatoire – syllabes/seconde (sans pauses)</i>	68
6.3	LIMITES MÉTHODOLOGIQUES	70
6.3.1	<i>Caractéristiques des acteurs</i>	70
6.3.2	<i>Caractéristiques de la méthodologie appliquée</i>	71
7.	CONCLUSIONS ET PERSPECTIVES	76
8.	BIBLIOGRAPHIE	79
9.	ANNEXES	85
	ANNEXE 1. MISES EN SITUATIONS FACULTATIVES À DESTINATION DES ACTEURS POUR INTERPRÉTER LES ÉMOTIONS (MONROY ET AL., 2022)	85
	ANNEXE 2. TABLEAU SYNTHÉTIQUE DE DIFFÉRENTES BASES DE DONNÉES ÉMOTIONNELLES ABORDÉES DANS LA LITTÉRATURE AINSI QUE LEUR MÉTHODOLOGIE	86
	ANNEXE 3. MÉTHODES DE DÉCOUPAGE DES AUDIOS ET DE RÉCOLTE DE DONNÉES ACOUSTIQUES DANS LE LOGICIEL PRAAT (VERSION 6.4.07 – MARCH 17, 2024)	87
	ANNEXE 4. NOMENCLATURE POUR RENOMMER LES FICHIERS DES PARTICIPANTS	90
	ANNEXE 5. RÉGLAGES DE LA CARTE-SON EXTERNE DU MICRO (FOCUSRITE ITRACK SOLO LIGHTNING)	90
	ANNEXE 6. TABLEAUX REPRENANT LES MOYENNES DES RÉSULTATS OBTENUS PAR PARAMÈTRE ACOUSTIQUE POUR CHAQUE ACTEUR ET CHAQUE CONDITION ÉMOTIONNELLE.	90
	ANNEXE 7. GRAPHIQUES REPRENANT LES ANALYSES DELTA DES MOYENNES OBTENUES POUR CHAQUE ACTEUR DE L'ÉCHANTILLON, DANS CHAQUE CONDITION ÉMOTIONNELLE PAR RAPPORT À LA CONDITION NEUTRE, ET CE, POUR CHAQUE PARAMÈTRE ACOUSTIQUE ÉTUDIÉ.	93
	ANNEXE 8. EXPLICATIONS DE NOTRE IMPLICATION RESPECTIVE POUR RÉALISER CE MÉMOIRE	97
	RÉSUMÉ	98

1. INTRODUCTION GÉNÉRALE

1.1 Contexte dans lequel s'inscrit le projet

L'impact de l'expression des émotions sur l'évolution des paramètres de la voix et de la parole est un sujet qui intéresse les chercheurs depuis des siècles. En effet, depuis l'Antiquité, les paramètres de la voix et de la parole sont reconnus comme étant très importants dans l'étude de l'expression des émotions. Pour ne citer que lui, Charles Darwin, dans son livre « The Expression of the Emotions in Man and Animals » (1872), soulignait déjà que la voix était un puissant vecteur des émotions, autant chez les animaux que chez les humains (Banse & Scherer, 1996). Depuis lors, il est clairement établi dans la littérature scientifique traitant ce sujet, que nos différents états émotionnels sont associés à des caractéristiques et des manifestations spécifiques des paramètres de la voix et de la parole (Banse & Scherer, 1996 ; Murray & Arnott, 1993).

Ainsi, notre projet de mémoire « Étude des paramètres acoustiques de voix et de parole liés à la production orale d'émotions », s'inscrit dans la continuité des travaux et des études qui ont été réalisés dans ce domaine de recherche.

1.2 Description du projet global dans lequel notre mémoire s'inscrit

Le projet de mémoire « Étude des paramètres acoustiques de voix et de parole liés à la production orale d'émotions » proposé par Mesdames Anne-Lise Leclercq et Angélique Remacle, s'inscrit dans le cadre d'un projet collaboratif mis en place entre le Département de Logopédie et la Haute Ecole de Commerce (HEC) de l'Université de Liège.

Dans le cadre de ce projet, nous avons eu l'opportunité de collaborer avec Madame Elodie Etienne (HEC), qui fait sa thèse de doctorat. En effet, Madame Etienne travaille sur le développement d'une base de données émotionnelle appelée « EVE », associée à une intelligence artificielle. « EVE » permet de reconnaître, grâce à l'analyse de paramètres de la voix et de la parole, les émotions véhiculées dans la prise de parole d'un locuteur face à un public virtuel. Pour ce faire, dix émotions ont été intégrées dans cette base de données : la peur, la joie, le dégoût, la colère, la tristesse, la surprise, le mépris, la confiance en soi, la confusion et la sympathie. Celles-ci ont été choisies par Madame Etienne pour le grand nombre de références scientifiques abordant leur expression au sein de la parole (Gournay & al., 2018) mais aussi pour leur caractère intéressant dans le contexte de prise de parole en public.

Ainsi, la base de données émotionnelle « EVE » et son intelligence artificielle, développées par Madame Elodie Etienne, permettent d'identifier en temps réel, au sein du discours du locuteur,

si certains paramètres de sa voix et de sa parole mettent en évidence des informations émotionnelles particulières. Ensuite, ces dernières sont perçues par un public d'auditeurs virtuels, qui y réagissent automatiquement.

1.3 Description de notre mémoire dans le cadre de ce projet collaboratif

Dans le cadre de ce projet collaboratif et de la thèse de doctorat de Madame Etienne, notre projet de mémoire consistait tout d'abord à réaliser les séances d'enregistrement de 12 acteurs (6 hommes et 6 femmes) produisant les 10 émotions citées précédemment, dans le but d'enrichir la base de données émotionnelle « EVE ».

Ensuite, en ce qui concerne plus spécifiquement notre mémoire, nous avons procédé à l'étude des paramètres acoustiques particuliers de la voix et de la parole (fréquence fondamentale, intensité, LTAS, CPPs, débit, vitesse articulatoire) pour 4 des 12 acteurs sélectionnés (2 hommes et 2 femmes) lorsque ceux-ci interprétaient les 6 émotions « universelles », ou émotions « de base », décrites par Paul Ekman (1993). Celles-ci correspondent à la peur, la joie, le dégoût, la colère, la tristesse et la surprise. Le fait de centrer notre étude sur ces 6 émotions était intéressant, car elles ont fait l'objet de nombreux travaux dans la littérature. Ainsi, nous disposons de diverses données sur leur manifestation dans la voix et la parole et plus précisément sur leurs paramètres, qui faisaient l'objet de notre étude. Cela nous a permis de dresser plusieurs hypothèses confirmatoires sur leur évolution avec la production des 6 émotions d'Ekman, avant d'entamer notre travail d'analyse et d'interprétation de nos résultats.

Néanmoins, pour quelques paramètres, nous ne possédions pas suffisamment de données dans la littérature pour formuler des hypothèses précises sur leur manifestation avec les émotions ciblées. De plus, pour certains d'entre eux, nous avons soulevé quelques contradictions entre les auteurs dans les articles consultés. C'est pourquoi, certaines de nos hypothèses sont exploratoires dans le cadre de ce travail.

Ensuite, nous avons réalisé le traitement des enregistrements de parole de nos 4 acteurs et avons analysé les paramètres acoustiques de leur voix et de leur parole qui nous intéressaient grâce au logiciel PRAAT (version 6.4.07 – March 17, 2024). Les éléments résultant de ces analyses ont été ensuite interprétés grâce à l'utilisation de la méthode des deltas, qui est une technique nous permettant de tirer des conclusions sur l'évolution des paramètres acoustiques de la voix et de la parole avec la production des émotions, en inter et intra-sujet.

Puis, nous avons confronté les données issues de nos analyses et de nos interprétations avec les hypothèses que nous avons formulées au préalable. Enfin, nous avons pu tirer des conclusions

sur celles-ci et ainsi confirmer ou infirmer celles qui sont soulevées dans la littérature et mettre en évidence celles qui devraient être approfondies et analysées davantage dans ce domaine.

Ainsi, nous espérons que notre projet de mémoire permettra de faire évoluer la recherche traitant de la manifestation des paramètres acoustiques de la voix et de la parole avec la production d'émotions, et d'ouvrir de nouvelles perspectives de travail dans ce domaine.

2. INTRODUCTION THÉORIQUE

2.1 Les émotions

2.1.1 Définition

D'après Ekman (1993), les émotions sont des phénomènes survenant à la suite d'évènements sociaux, qu'ils soient réels, ancrés en mémoire ou imaginés. Selon Scherer (1981) cité par Murray et Arnott (1993), les émotions exerceraient une fonction indispensable pour notre organisme. En effet, elles permettraient à notre corps de communiquer avec le monde environnant en réalisant 3 fonctions particulières : mettre en évidence les stimuli que l'organisme préfère et ceux dont il a besoin, préparer physiologiquement et psychologiquement l'organisme pour réaliser une action considérée comme appropriée par celui-ci et enfin, communiquer l'état ainsi que les intentions comportementales de l'organisme à d'autres organismes présents dans l'environnement.

De plus, d'après Gil (2009), une expérience émotionnelle dépendrait de trois composantes essentielles : une composante cognitive (changement d'état mental lié à l'émotion), comportementale (manifestations comportementales et expressives dirigées vers l'extérieur) et physiologique (manifestations physiologiques concomitantes à un évènement émotionnel). Le psychologue Williams James (1842-1910), cité par Gil (2009) illustre l'interaction de ces 3 composantes par l'exemple de la vue d'un ours en forêt. Tout d'abord, à la vue de celui-ci, on ressent ce qui s'apparente à de la peur (cognition), ensuite un comportement de fuite se met en place car les yeux s'ouvrent, le besoin de courir devient important, ... (comportement). Enfin, le rythme respiratoire et le rythme cardiaque s'accélèrent, ... (physiologie).

Les émotions ne doivent pas être confondues avec les humeurs. En effet, les émotions surgissent soudainement en réponse à des stimuli environnants et sont très brèves (quelques secondes à quelques minutes). Ce sont des réponses de l'organisme qui surviennent à la suite de changements dans l'environnement. En revanche, les humeurs sont plus vagues, elles peuvent durer plusieurs heures ou quelques jours. Elles peuvent être considérées comme une réponse à des éléments ou des évènements qui durent dans le temps (Murray & Arnott, 1993).

2.1.2 Les 6 émotions universelles de Paul Ekman

Les différentes études réalisées dans le domaine des émotions ne s'accordent pas sur un nombre d'émotions de base à définir, car il peut être très compliqué de les différencier. En effet, certaines d'entre elles sont parfois le résultat de la combinaison d'autres émotions. Ainsi, certaines théories s'accordent pour qualifier les émotions comme des entités combinées faisant

partie d'un continuum émotionnel, plutôt que des éléments très différents les uns par rapport aux autres (Murray & Arnott, 1993).

Cependant, 6 émotions universelles, ou émotions « de base » ont été mises en évidence par Paul Ekman¹ et ont fait l'objet d'études plus approfondies : la colère, le dégoût, la peur, la joie, la surprise et la tristesse (Murray & Arnott, 1993). Ekman et Cordaro (2011) définissent précisément ces 6 émotions.

La tristesse

La tristesse est une émotion survenant à la suite de la perte d'un objet auquel nous étions attachés ou d'une personne qui comptait énormément pour nous. L'expérience personnelle la plus souvent évocatrice de cette émotion est la mort d'un enfant, d'un parent ou d'un conjoint.

La colère

La colère est une émotion ressentie lorsqu'un obstacle ou un événement interfère avec un objectif personnel qui nous tient à cœur. Cette émotion peut également être ressentie lorsque quelqu'un tente de nous faire du mal physiquement ou psychologiquement. Elle peut provoquer, en guise de réponse, la volonté personnelle de supprimer l'obstacle en question, ou encore de blesser ou faire du mal à la personne concernée.

La peur

La peur est une réponse à une menace perçue dans notre environnement qui peut être physique ou psychologique. Elle déclenche des réactions de fuite ou de stupeur, et peut parfois causer des réactions émotionnelles de colère.

La joie

La joie est une émotion regroupant un ensemble de sentiments agréables appréciés et recherchés.

¹ À titre informatif, Paul Ekman est l'un des chercheurs les plus influents de sa génération dans le domaine de l'étude des émotions et de leur expression (selon « The American Psychological Association »). Il s'est spécifiquement intéressé à leur manifestation dans les expressions faciales et les micro-expressions faciales, c'est-à-dire celles qui sont très brèves et difficilement identifiables (Ramsland, 2012).

Le dégoût

Le dégoût est une sensation de répulsion causée par la vue, l'odeur, ou le goût d'un élément spécifique. Cette émotion peut également être ressentie lorsqu'une personne tient des propos offensants ou réalise des actions que nous jugeons révoltantes.

La surprise

La surprise est une émotion de très brève durée qui est le résultat d'un changement soudain et inattendu dans l'environnement.

Ekman et Cordaro (2011), ont attribué à ces 6 émotions le nom d'émotions « de base », pour souligner leur caractère discret, permettant ainsi de les classer en 6 entités distinctes les unes par rapport aux autres. Celles-ci ont été également définies comme universelles, car leur manifestation dans les expressions faciales serait identique pour chacun de nous (Ekman, 1993), mais aussi car ce sont celles qui seraient les plus courantes dans la culture occidentale (Murray & Arnott, 1993). Les émotions seraient socialement déterminées, car les êtres humains réagissent à des stimuli identiques universellement partagés. Nos rituels et expériences partagés en tant qu'êtres humains seraient alors à la base de nos représentations et expressions communes de nos émotions (Ekman & Cordaro, 2011).

Concrètement, ces 6 émotions universelles correspondraient à des groupes émotionnels ou des « familles » d'émotions, plutôt qu'à des états émotionnels ou psychologiques isolés (Ekman & Cordaro, 2011). Chaque « famille » regrouperait une variété d'expressions émotionnelles partageant des propriétés communes, mais se distinguant les unes des autres par leur intensité, leur contrôle, leur spontanéité et les types d'évènements qui les déclenchent. Par exemple, le groupe émotionnel de la colère reprendrait des états émotionnels allant de l'agacement (colère de basse intensité) à la rage (colère de très haute intensité) (Ekman, 1993). Néanmoins, plusieurs points communs à ces « familles » d'émotions ont été mis en évidence. Par exemple, chacune possède ses propres signaux universels distinctifs, a une physiologie distincte, apparaît rapidement et de manière imprévue, est de courte durée, est présente chez les primates, ... (Ekman & Cordaro, 2011).

Il est important de noter que la notion d'intensité émotionnelle abordée ci-dessus a été prise en considération dans le cadre de la méthodologie de notre mémoire. En effet, chaque émotion interprétée par les participants a été produite avec une faible et une forte intensité lors des séances d'enregistrement. Néanmoins, nous avons interprété les résultats des différents

paramètres acoustiques de la voix et de la parole uniquement lorsque les acteurs exprimaient l'émotion à forte intensité, afin que la variabilité des différents paramètres étudiés soit davantage contrastée lors de l'interprétation de nos données acoustiques.

Enfin, Trinh Van et al. (2022), mettent en évidence que les émotions humaines sont très riches, car elles peuvent prendre des formes très diversifiées. En effet, les émotions ressenties par un individu peuvent être véhiculées par son langage corporel, ses expressions faciales, ses contacts visuels, sa voix, sa parole et bien d'autres.

De plus, les émotions peuvent être déclenchées par différents stimuli. En effet, l'étude de Gil (2009) met en évidence différentes manières d'induire (provoquer) les émotions chez une autre personne : l'utilisation de techniques standardisées et non-standardisées. Les techniques standardisées les plus souvent utilisées sont la présentation de stimuli visuels à caractère émotionnel tels que des images ou des films, ou encore des stimuli auditifs comme l'écoute de musiques. Les techniques non-standardisées, quant à elles, induisent les émotions chez le sujet dans un environnement particulier, en créant au sein de celui-ci une situation ou une interaction expérimentale spécifique. Ces techniques n'utilisent donc pas de stimuli standardisés, car elles varient en fonction des caractéristiques propres à l'environnement. Il est important de noter qu'elles induisent plutôt des états émotionnels en termes de valence, c'est-à-dire sur un continuum positif-négatif, que des émotions discrètes (ex : les 6 émotions d'Ekman). On peut citer parmi celles-ci les expériences conduisant à un « succès » ou un « échec », les tâches de rappel autobiographique, les interactions entre 2 personnes familières à propos de sujets plaisants ou créant des conflits (*dyadic interactions*), ... (Gil, 2009).

Ces émotions, une fois induites, peuvent être mesurées et évaluées selon leurs différentes composantes : cognitive, comportementale et physiologique. L'évaluation cognitive des émotions se réalise principalement via la complétion d'échelles d'auto-évaluation, telles que la DES « Differential Emotions Scale »² (Izard, 1977 ; adaptation française Ouss & al., 1990, cités par Gil, 2009), ou encore l'échelle graphique SAM « Self-Assessment Manikin scale »³ (Bradley & Lang, 1994 cités par Gil, 2009). L'évaluation comportementale consiste à

² Questionnaire d'auto-évaluation d'émotions discrètes (Izard, 1977 ; traduction française Ouss & al., 1990). Il est composé de 30 adjectifs correspondant à 10 états émotionnels isolés. Chaque adjectif est évalué sur une échelle en 5 points (Gil, 2009).

³ Echelle graphique de mesure de dimensions émotionnelles (Bradley & Lang, 1994) sur laquelle l'état d'une figurine est décliné en 9 degrés sur un continuum de valence émotionnelle (négatif à positif). Par exemple, pour la dimension du plaisir, on observe un visage joyeux à l'extrémité de l'échelle, et de l'autre, un visage renfrogné (Gil, 2009).

réaliser des mesures sur les comportements des individus, tels que leurs expressions faciales (contraction des muscles faciaux), les paramètres acoustiques de la voix, ... Enfin, l'évaluation physiologique des émotions peut se faire en mesurant des paramètres comme la fréquence cardiaque, le réflexe psychogalvanique⁴, la fréquence respiratoire, l'électro-encéphalographie, ... (Gil, 2009).

Il est important de garder à l'esprit que l'ensemble de ces variables ont été étudiées en laboratoire (contexte artificiel) afin d'obtenir les données expérimentales les plus robustes sur les variables permettant d'induire les émotions d'autrui et leur évaluation. Néanmoins, il serait plus pertinent d'étudier l'induction des émotions en contexte naturel, afin d'obtenir des données sur les stimuli induisant les émotions dans des situations concrètes et plus proches de la réalité (Gil, 2009).

Ainsi, il est clairement établi que les émotions peuvent prendre des formes diverses et être induites par des stimuli très différents. Néanmoins, dans le cadre de ce mémoire, nous nous sommes intéressées plus particulièrement aux indicateurs acoustiques des émotions au sein du signal de parole des acteurs participant à notre étude. Notre recherche se base donc sur la « forme » acoustique des 6 émotions d'Ekman dans la voix et la parole, lorsque celles-ci sont induites par les acteurs eux-mêmes.

2.1.3 L'expression des émotions au sein de la parole

L'expression des émotions au sein de la parole joue un rôle très important dans l'information transmise entre plusieurs interlocuteurs qui discutent (Trinh Van & al., 2022). En effet, au sein d'une discussion, l'expression émotionnelle fournit des informations sur les intentions du locuteur et influence alors les réactions émotionnelles de son partenaire de communication (Monroy, Cowen & Keltner, 2022).

L'expression émotionnelle communiquée au sein de la parole est principalement modulée par deux composantes : la prosodie⁵ ainsi que le contenu lexical et sémantique de l'énoncé (Lingelbach & al., 2023). De plus, la manifestation des émotions est fortement variable au cours du temps et dépendante du contexte. Ainsi, plusieurs composantes des émotions varient au cours

⁴ Mesure de l'activité électrodermale (AED) des glandes sudoripares de la paume de la main. On mesure la variation de conductance électrique provoquée par les changements de sudation du sujet (Gil, 2009).

⁵ La prosodie fait référence au niveau suprasegmental du langage. Elle concerne l'étude de l'intonation, de l'accentuation, du rythme, du débit et des pauses de la parole (Lacroix, 2016). A titre informatif, dans le traitement émotionnel de la parole, la prosodie est plus importante que son contenu sémantique. On parle alors de dominance prosodique. Cependant, les deux composantes doivent être considérées pour identifier convenablement l'émotion dans le signal de parole (Lingelbach & al., 2023).

du temps selon leur continuum respectif : leur caractère satisfaisant, autrement appelé valence (continuum agréable-désagréable ou positif-négatif), leur éveil ou activation (continuum calme-excitation), et leur caractère contrôlable par le sujet appelé dominance (continuum contrôlable-incontrôlable) (Gil, 2009 ; Lingelbach & al., 2023).

Par ailleurs, la littérature consultée nous a permis de décrire les processus physiologiques sous-tendant la production orale d'émotions et leur variation au sein du signal de parole. Ainsi, la production orale de la parole est principalement modulée par le néocortex⁶. L'excitation (ou activation) émotionnelle dans le signal de parole, quant à elle, dépendrait principalement de l'activité du système limbique⁷. L'activation du système nerveux somatique⁸ et l'activation (sympathique et parasympathique) du système nerveux autonome⁹ permettraient enfin la production des émotions dans le signal acoustique vocal. Néanmoins, étant donné la multiplicité de facteurs déterminant le processus de production vocale, de très légers changements de ces paramètres physiologiques suffisent à apporter des modifications importantes au signal acoustique de la parole d'un individu (Scherer, 1995).

La littérature scientifique traitant de l'expression des émotions au sein de la parole s'est tout d'abord penchée sur l'étude des 6 émotions mises en évidence par Paul Ekman, qui font l'objet de notre mémoire (la colère, le dégoût, la peur, la joie, la surprise et la tristesse). Plus tard, ce sont des émotions plus complexes comme l'embarras, la honte, la fierté, l'amour, le désir, la sympathie, la confusion, la crainte, la satisfaction, ... qui ont été soumises à des analyses plus approfondies (Monroy, Cowen & Keltner, 2022).

En outre, beaucoup d'études se sont penchées sur les processus sous-jacents à l'identification des émotions d'autrui.

⁶ Le néocortex permet la commande et la production des mouvements phonatoires et articulatoires produits intentionnellement (caractéristiques prosodiques) ou non-intentionnellement (Scherer, 1995).

⁷ Le système limbique rassemble les structures corticales et sous-corticales qui sont interconnectées et dédiées à relier les états viscéraux et les émotions à la cognition et au comportement (Mesulam, 2000 cité par Catani & al., 2013).

⁸Le système somatique fait référence aux commandes volontaires que nous pouvons effectuer sur notre corps. Ce système est contrôlé par les circuits hypothalamiques, qui reçoivent les informations sensorielles sur l'état du milieu intérieur et extérieur (Di Scala, 2023).

⁹ Le système nerveux autonome correspond aux commandes involontaires du corps. La production émotionnelle comprend donc deux systèmes parallèles, qui sont anatomiquement et fonctionnellement différents : volontaires (somatique) et involontaires (autonome) (Di Scala, 2023).

2.1.4 L'identification des émotions

En ce qui concerne l'identification des émotions, Cowen et al. (2019) mettent en évidence son caractère indispensable et fondamental pour le développement de l'interaction sociale de chaque individu. En effet, depuis tout-petits, les enfants identifient les émotions sur le visage et dans la voix de leurs parents et se servent de celles-ci pour répondre adéquatement à leur environnement (Cowen & al., 2019).

L'identification des émotions est un processus qui peut être réalisé grâce à la combinaison de multiples paramètres. Par exemple, la combinaison d'enregistrements vocaux avec des expressions faciales ou avec le langage corporel garantit une meilleure reconnaissance des émotions (Trinh Van & al., 2022). En effet, se baser uniquement sur l'analyse du comportement verbal d'un individu est insuffisant pour identifier ses émotions (Ekman, 1993). Il est donc indispensable de prendre également en considération ses expressions faciales, qui seraient particulièrement riches d'informations émotionnelles et permettraient une meilleure reconnaissance des émotions (Russel, 2003), mais aussi les paramètres de sa voix (Ekman, 1993).

Ce processus d'identification prend place systématiquement lorsque nous sommes en interaction avec d'autres personnes (Liu Pan & al., 2021). Ainsi, les scientifiques qui réalisent des recherches dans le domaine de l'identification des émotions se doivent de sélectionner les stimuli émotionnels les plus naturels et complexes possibles afin de garantir la validité de leurs travaux. En effet, étudier les émotions comme des éléments fragmentés et détachés de tout contexte social aurait peu de sens pour en apprendre davantage sur nos capacités à traiter et identifier les émotions (Lingelbach & al., 2023).

2.1.5 L'identification des émotions dans le signal de parole

D'après Van Rijn et Larrouy-Maestri (2023), l'approche perceptive des émotions serait limitée, car elle ne prendrait pas suffisamment en compte les variables inter et intra-individuelles des interprètes. Ainsi, ces auteurs ont réalisé un modèle statistique bayésien¹⁰ afin d'analyser les variables pertinentes à prendre en considération chez les interprètes dans les études s'intéressant à l'identification des émotions. Ainsi, les auteurs soulignent l'intérêt de prendre en compte le

¹⁰ Le modèle statistique bayésien, basé sur le théorème de Thomas Bayes (mathématicien anglais), prend en considération les données observées lors d'une étude et les connaissances préalables à propos des variables qui y sont manipulées, dans le but d'estimer leur distribution. Ainsi, ce modèle considère chaque paramètre étudié comme une variable aléatoire et prend en considération l'incertitude associée à leur estimation (Kruschke 2010).

corpus qui a été utilisé, mais aussi le sexe, la culture et la langue de l'individu qui a produit les émotions dans ces études (Van Rijn et Larrouy-Maestri, 2023).

Lorsque nous percevons une émotion particulière dans le signal vocal de notre interlocuteur, une région cérébrale spécifique serait davantage activée. Plus précisément, l'identification des émotions d'autrui serait sous-tendue par l'activation de nos hémisphères cérébraux. Ainsi, l'activation de l'hémisphère cérébral droit serait plus importante que celle de l'hémisphère gauche pour traiter les expressions faciales et les caractéristiques émotionnelles prosodiques de notre interlocuteur. Les émotions positives (ex : joie) seraient associées à une activation plus importante des aires frontales gauches, alors que les émotions négatives (ex : colère) seraient caractérisées par une activation plus importante des aires frontales droites (Hellige, 2002).

Par ailleurs, le traitement émotionnel de chaque paramètre acoustique du signal vocal de notre interlocuteur implique un délai de traitement et une région cérébrale, spécifiques à ce paramètre. Par exemple, une activation cérébrale droite serait observée lorsque nous percevons des informations émotionnelles dans la hauteur tonale d'un individu (fréquence fondamentale), car la hauteur du son serait traitée dans cet hémisphère cérébral (Mauchand & Zhang, 2023). En outre, s'il est demandé à un individu de déterminer, rien qu'à la voix de son interlocuteur, si celui-ci ressent une émotion ou non, ce processus de traitement prendrait plus de temps (200 ms) en présence d'une émotion dans la voix de celui-ci, qu'en l'absence d'émotion. La mesure de ce délai de traitement permettrait ainsi d'objectiver la présence ou l'absence de contenu émotionnel dans la voix de l'interlocuteur (Mauchand & Zhang, 2023).

Il est important de noter que les émotions associées à un état de menace pour l'organisme, le mettant alors dans un mécanisme de survie, seraient mieux perçues que les autres émotions. C'est le cas de la colère et de la peur, qui seraient les émotions les plus facilement identifiables chez notre interlocuteur (Vos & al., 2023). Des études et des recherches supplémentaires sur le traitement des émotions dans le signal vocal restent nécessaires, afin de comprendre davantage le fonctionnement de ses mécanismes cérébraux et de ses régions cérébrales émotionnelles spécifiques (Mauchand & Zhang, 2023).

L'ensemble de ces données perceptives sont très intéressantes et permettent d'en apprendre davantage sur les mécanismes sous-jacents à l'identification des émotions d'autrui dans sa voix et sa parole. Néanmoins, dans le cadre de ce mémoire, nous nous concentrerons uniquement sur le versant expressif des émotions.

Dans ce mémoire, nous nous sommes également intéressées à la voix et ses paramètres acoustiques. Ceux-ci seront cités et explicités dans la prochaine section de ce travail.

2.2 La production vocale

2.2.1 Définition

D'après Naqvi et Gupta (2023), la voix est le résultat du passage de l'air expiré venant des poumons entre les plis vocaux qui sont en vibration. Une production (phonation) et une mise en forme (articulation) efficaces de la voix nécessitent une synchronisation et un fonctionnement optimal des structures laryngées, diaphragmatiques et articulatoires. Ainsi, selon Cowen et al. (2019), les paramètres acoustiques caractérisant la voix de l'être humain dépendent à la fois de l'activité exercée par les poumons, de la tension des plis vocaux, des mouvements exercés par la cavité buccale (la bouche, les joues et les lèvres), mais aussi du placement résonantiel dans les résonateurs supérieurs (fosses nasales, cavum, pharynx, vestibule laryngé, cavité buccale, espace entre les lèvres).

Les paramètres acoustiques de la voix qui ont été étudiés et analysés dans le cadre de ce mémoire sont définis ci-après.

2.2.2 Les paramètres acoustiques de la voix

La fréquence fondamentale (f_0) - Hz

La fréquence fondamentale de la voix (appelée également fréquence usuelle ou fréquence moyenne) correspond physiologiquement au nombre moyen de cycles vibratoires réalisés par les plis vocaux en une seconde. Elle est perçue comme la hauteur de la voix, c'est-à-dire son caractère grave ou aigu. Son unité de mesure est le Hertz (Hz), mais peut parfois être exprimée en demi-tons (ex : domaine musical) (Banse & Scherer, 1996 ; Cantor-Cutiva, Bottalico & al., 2019). La valeur moyenne de la fréquence fondamentale diminuerait avec l'avancée en âge chez les femmes et augmenterait chez les hommes (Kent & al., 2021). Ses valeurs de référence se situent entre 78 et 166 Hz chez l'homme et entre 167 et 258 Hz chez la femme (De Bodt, Heylen, Mertens, Vanderwegen & Van de Heyning, 2008).

Afin d'étudier ce paramètre dans notre mémoire, nous avons tout d'abord choisi d'analyser sa médiane. Celle-ci nous a permis d'obtenir, pour chaque émotion, une valeur de tendance centrale représentative de la fréquence fondamentale. La médiane correspond à la valeur de la fréquence fondamentale se retrouvant à équidistance de ses deux valeurs extrêmes ($f_{0 \max}$ et $f_{0 \min}$).

Nous nous sommes également intéressés à son étendue. En effet, évaluer l'étendue nous permet d'obtenir des données sur la gamme fréquentielle de la fréquence fondamentale de l'acteur, c'est-à-dire le nombre de Hertz distançant la valeur de fréquence fondamentale maximale et minimale au cours d'un énoncé. Qualitativement, évaluer l'étendue de la fréquence fondamentale nous permet de qualifier les variabilités intonatives de l'acteur au cours de ses productions. L'étendue est calculée comme-suit :

$$f_o \text{ étendue} = f_o \text{ max} - f_o \text{ min}$$

Le niveau de pression sonore - dB

Le niveau de pression sonore, autrement appelé « intensité », correspond à la puissance sonore (ou quantité d'énergie sonore) émise par l'appareil phonatoire (Švec & Granqvist, 2018). Selon son aspect perceptif, elle caractérise le côté « fort » ou « faible » du volume de la voix. De son point de vue physiologique, le niveau de pression sonore correspond à l'amplitude (ou degré d'ouverture) des plis vocaux lors de leur vibration. En effet, l'amplitude de l'onde sonore est dépendante de la quantité de pression d'air sous-glottique produite par l'activité expiratoire des poumons, qui franchit les plis vocaux à chaque cycle vibratoire. Cette pression entre en collision avec la colonne d'air présente dans le tractus vocal. Ainsi, la valeur de l'amplitude obtenue correspond à la différence de pressions existant entre la région sous-glottique et le tractus vocal, mais dépend également de la force d'adduction des plis vocaux (Giovanni, 2021). L'intensité vocale s'exprime en décibels (dB), mais peut parfois être mesurée en pascals (Pa) lorsqu'elle est envisagée sous sa forme aérodynamique (Ghio, 2021 ; Švec & Granqvist, 2018).

Dans ce mémoire, nous avons pris la décision d'étudier l'intensité selon l'évolution de sa médiane et de son étendue avec les émotions. En effet, la médiane nous permet d'obtenir, pour chaque émotion, une valeur de tendance centrale représentative de l'intensité produite chez les acteurs de notre échantillon. L'étendue nous indique, quant à elle, le nombre de décibels distançant les valeurs d'intensité maximale et minimale lorsqu'une émotion est produite et nous renseigne sur l'intonation de l'acteur. L'étendue est calculée comme-suit :

$$\text{Intensité étendue} = \text{Intensité}_{\text{max}} - \text{Intensité}_{\text{min}}$$

Smoothed Cepstral Peak Prominence (CPPs) - dB

Le « Cepstral peak prominence » (CPP) représente à l'heure actuelle la mesure de dysphonie la plus fiable et présentant la meilleure corrélation avec les analyses perceptives de la voix (Halberstam, 2004 ; Hillenbrand & Houde, 1996 cités par Popolo & Johnson, 2020 ; Lowell & al., 2013). Le CPP constitue un outil sensible et spécifique afin de mesurer le degré d'harmonie

au sein d'un échantillon vocal (Heman-Ackah & al., 2014). Il est calculé en mesurant la différence d'amplitude (intensité) entre l'harmonique dominante (qui forme un pic dans le spectrogramme) et la fréquence fondamentale (Antonetti & al., 2020). Ainsi, plus le signal vocal est périodique, plus le degré d'harmonie est important et plus la valeur du CPP augmente. L'avantage de cette mesure est qu'elle peut être réalisée sur n'importe quel échantillon de voix (par exemple sur un discours spontané, un /a/ tenu, ...), qu'il soit périodique ou non, et à n'importe quelle intensité. Plusieurs algorithmes de calcul du CPP existent, dont une version appelée « Smoothed Cepstral Peak Prominence » (CPPs). Cette mesure « lissée » permettrait d'obtenir une valeur d'approximation de la dysphonie plus robuste, moins sujette aux artéfacts, et plus représentative des caractéristiques perceptives de la voix (Antonetti & al., 2020 ; Heman-Ackah & al., 2014). La valeur seuil du CPPs a été définie à 4,0 dB par Heman-Ackah et al. (2014), dans des échantillons de discours. Cela signifie qu'un individu dont la valeur de CPPs est inférieure à 4,0 dB, a une plus grande probabilité d'être dysphonique (Heman-Ackah & al., 2014). Dans le cadre de ce mémoire, les valeurs du CPPs nous permettent d'obtenir des données quant à la stabilité et la régularité de l'onde sonore produite par nos acteurs lorsqu'ils interprètent une émotion.

Long-Term Average Spectrum (LTAS)

Le « Long-Term Average Spectrum » (LTAS) est une mesure acoustique de dysphonie qui, comme le CPPs, peut être mesurée sur n'importe quel échantillon vocal (/a/ tenu, discours spontané, ...) qu'il soit périodique ou apériodique (Antonetti & al., 2020). Elle permet d'obtenir des données sur la variation du spectre sonore de la voix (Nordenberg & Sundberg, 2004). La mesure du LTAS est obtenue en analysant les signaux acoustiques de la source sonore et des éléments qui la filtrent, ainsi que leur interaction.

Il est important de noter que cette mesure comprend les indices acoustiques suivants : le rapport L1-L0 et le pourcentage de non-voisement. Le rapport L1-L0 correspond à la différence d'énergie existant entre le premier formant¹¹ et la fréquence fondamentale. Il fournit des indices sur la fuite d'air à travers le plan glottique et ainsi, sur l'efficacité phonatoire et la présence de souffle dans la voix. Une valeur supérieure de L0 par rapport à L1 met en évidence une hypo-adduction glottique, et lorsque L0 est plus faible que L1, cela démontre une hyper-adduction glottique. Des valeurs de L0 et L1 similaires indiqueraient une production vocale économique,

¹¹ À titre informatif, les formants du signal vocal correspondent à des "pics" spectraux qui déterminent le timbre de chaque son de la parole (Meynadier, 2013).

ne demandant alors pas trop d'énergie (Antonetti & al., 2020). En ce qui concerne le pourcentage de non-voisement, celui-ci mesure le rapport entre les sons voisés et non-voisés produits au cours de la phonation. En d'autres termes, il renseigne sur le pourcentage de phonèmes qui ont été produits sans avoir initié de vibration cordale au cours d'un énoncé. Il nous permet donc également d'obtenir des données sur l'efficacité phonatoire, c'est-à-dire la présence ou l'absence de mise en vibration des plis vocaux pour les sons voisés. Dans le cadre de ce mémoire, l'étude du rapport L1-L0 et du pourcentage de non-voisement nous permet de tirer des conclusions sur le caractère soufflé et non-voisé des différents énoncés produits avec les émotions d'Ekman.

Enfin, après avoir développé les différents paramètres acoustiques de la voix, il convient également d'aborder le concept de fluence de la parole et ses différents paramètres. En effet, ceux-ci font partie intégrante de notre travail et constituent un point central de notre question de recherche. Ainsi, la fluence de la parole et ses paramètres méritent, eux aussi, d'être traités et explicités davantage.

2.3 La fluence de la parole

2.3.1 Définition

La fluence est un concept qui a longtemps été controversé dans la littérature scientifique, car les auteurs qui se sont intéressés à ce sujet ne s'accordaient pas toujours sur sa définition. En effet, initialement, la fluence était décrite comme l'usage régulier, rapide et sans effort de la langue (Crystal, 1987 cité par Rodríguez-Fuentes & Calle-Díaz, 2023). De plus, la définition de cette notion se limitait uniquement à sa modalité dans l'expression orale. On abordait alors le concept de « fluence de la parole » comme une compétence linguistique globale, correspondant à la maîtrise efficace d'une langue, qui pouvait être mesurée quantitativement en calculant le nombre de syllabes et de pauses produites dans la parole par unité de temps. De plus, ce concept a longtemps été confondu avec les notions d'intelligibilité ou de précision de la parole (Chambers, 1997).

Plus tard, les recherches effectuées dans le domaine de la fluence de la parole nous ont permis d'enrichir ce concept et de perfectionner sa définition. Ainsi, la fluence serait actuellement définie comme la capacité à traiter et produire le langage à un moment donné et à vitesse raisonnable (Nation, 2014 cité par Rodríguez-Fuentes & Calle-Díaz, 2023). Cette nouvelle définition est bien plus complète, car elle prend en considération un spectre de compétences ne

se limitant pas uniquement à l'expression orale de la parole. En effet, elle intègre également les capacités liées au langage écrit (lecture et écriture) (Rodríguez-Fuentes & Calle-Díaz, 2023).

2.3.2 Les paramètres de la fluence de la parole

La littérature consultée met en évidence différents paramètres de la fluence de la parole.

Le débit de parole

Le débit de parole est le paramètre de la fluence qui représente le nombre de mots ou de syllabes produits par un locuteur dans un laps de temps donné (par exemple x mots produits/min) ou sur un nombre de mots déterminé (par exemple x syllabes produites /100 mots). Il tient compte des intervalles silencieux (pauses et silences) de l'échantillon de parole (Liu & al., 2022 ; Themistocleous & al., 2020). Cette mesure reflèterait la capacité du locuteur à planifier les mouvements moteurs de sa parole (Liu & al., 2022). La littérature traitant de l'évolution des paramètres de la fluence avec les émotions aborde souvent ce paramètre. Ainsi, il nous paraît pertinent de nous intéresser à ce paramètre afin de confronter nos observations avec ce qui est avancé dans la littérature à son sujet. Dans le cadre de notre étude, nous avons choisi de mesurer ce paramètre en nombre de syllabes produites par seconde. En effet, cette mesure nous paraissait plus précise et cohérente pour le peu de mots constituant les échantillons de parole étudiés (10 phrases de la liste 1 du corpus de Fharvard, Aubanel & al., 2020) et le peu de temps nécessaire pour les produire. La formule du débit de parole est la suivante :

$$\text{débit de parole de l'énoncé} = \frac{\text{nombre de syllabes produites}}{\text{durée totale énoncé (s)}}$$

La vitesse articulatoire

La vitesse articulatoire (ou taux d'articulation) représente une mesure rythmique temporelle de la parole qui calcule le nombre de segments réalisés par seconde, en supprimant les moments de pauses et de silences. Cette exclusion évite de prendre en considération des caractéristiques individuelles propres aux locuteurs comme les pauses silencieuses, les hésitations, ... (Kühnert & Kocjančič Antolík, 2016 ; Themistocleous & al., 2020). Tout comme le débit, l'étude de la vitesse articulatoire dans un contexte de production d'émotions est développée dans la littérature. Sa formule est la suivante :

$$\text{vitesse articulatoire} = \frac{\text{nombre de syllabes produites}}{\text{durée totale énoncé (s)} - \text{pauses (s)}}$$

Les pauses

Les pauses au sein du discours sont de bons indicateurs de la fluence de la parole du locuteur. On distingue les pauses non-remplies, qui sont des périodes de silence dans le discours, et les

pauses remplies, qui représentent des laps de temps durant lesquels des sons (ex : « euh » ou « hum ») ou des expressions (ex : « bien », « tu vois ») sont produits par le locuteur. Ces deux types de pauses indiquent alors que l'individu qui produit le discours est confronté à des difficultés pour formuler son énoncé et planifie ce qu'il va dire par la suite (Liu & al., 2022).

Les pauses peuvent également être étudiées en termes de durée. Certains auteurs signalent que la durée des pauses peut varier selon différents facteurs : le locuteur, le débit de parole, le type de discours produit, la structure prosodique, la longueur de phrase et la syntaxe (Krivokapić, 2007). Cependant, la durée des pauses produites serait surtout dépendante des caractéristiques propres aux locuteurs (Goldman Eisler, 1968 cité par Krivokapić, 2007). De plus, certains auteurs considèrent les pauses comme courtes lorsqu'elles sont inférieures à une durée de 300 milli secondes et longues lorsqu'elles dépassent les 1000 millisecondes (Butcher, 1981 ; Zvonik & Cummins, 2003 cité par Krivokapić, 2007). Il est important de noter que dans le cadre de ce travail, nous avons considéré les silences qui étaient supérieurs ou égaux à une durée de 250 millisecondes comme correspondant à des pauses. Nous avons choisi cette durée de référence, car dans la littérature consultée, une pause silencieuse peut être définie comme telle lorsque sa durée est égale ou supérieure à 250 millisecondes (Grosjean & Deschamps, 1975, cités par Grosman & al., 2018). Enfin, l'étude des pauses produites au cours des énoncés nous a permis de déterminer la vitesse articulatoire, en appliquant sa formule ainsi que la méthodologie décrite en annexe (**cf. Annexe 3**).

Ainsi, les paramètres acoustiques de la voix qui seront approfondis spécifiquement dans le cadre de ce mémoire sur base des enregistrements de parole de nos 4 participants sont les suivants : la fréquence fondamentale (étendue et médiane), l'intensité (étendue et médiane), le LTAS (rapport L1-L0 et pourcentage de non-voisement) et le CPPs. En ce qui concerne l'étude des paramètres acoustiques de la fluence de la parole, nous analyserons le débit (syllabes/seconde) ainsi que la vitesse articulatoire (syllabes/seconde sans les pauses). Ces différents paramètres seront ensuite analysés et étudiés via le logiciel PRAAT (6.4.07) (**cf. Annexe 3**). Enfin, nous confronterons nos observations avec ce qui est avancé dans la littérature à leur sujet.

Enfin, après avoir présenté et développé les multiples paramètres de la voix et de la fluence de la parole qui seront étudiés dans le cadre de ce travail, il convient d'aborder leur manifestation lorsqu'un locuteur interprète des émotions particulières. Ainsi, les prochains paragraphes développent les données issues de la littérature sur ce phénomène, lorsque les 6 émotions universelles de Paul Ekman se présentent.

2.4 L'impact des émotions sur les paramètres de la voix et de la fluence

La voix humaine joue à la fois le rôle de médiateur d'un message à destination d'un interlocuteur, mais transmet également de nombreuses informations sur l'état émotionnel dans lequel le locuteur se trouve (Trinh Van & al., 2022). En effet, une émotion interprétée par un locuteur peut être identifiée dans des enregistrements vocaux aussi brefs que 60 millisecondes. Cela signifie que l'émotion produite par un individu qui prend la parole, est bien présente et identifiable dans chacun de ses énoncés, les plus brefs soient-ils (Banse & Scherer, 1996 ; Murray & Arnott, 1993).

En ce qui concerne plus spécifiquement la fluence de la parole, nous pouvons aborder la notion de prosodie. Pour rappel, la prosodie fait référence, tout comme la fluence, au niveau suprasegmental du langage. Elle concerne l'étude de l'intonation, de l'accentuation, du rythme, du débit et des pauses de la parole (Lacroix, 2016). Ainsi, il est clairement établi que la prosodie de la parole, outre sa dimension linguistique, posséderait également une dimension émotionnelle très importante qu'on ne peut pas sous-estimer, tout comme les expressions faciales et les postures (Lacroix, 2016 ; Philippot, 2007). Cette dynamique émotionnelle prosodique serait présente chez chacun de nous (Robert W. Frick¹², 1985 cité par Bänziger & Scherer, 2005). Plusieurs auteurs ont mis en évidence son rôle essentiel lors de la production d'émotions. Par exemple, Bänziger et Scherer (2005), ont relevé que chaque type d'intonation dans la parole (ex : variation de la fréquence fondamentale) serait associé à une émotion particulière. Ainsi, la prosodie serait l'un des facteurs les plus importants à prendre en considération dans l'expression émotionnelle de la parole (Philippot, 2007 ; Bharadwaj & Acharjee, 2022).

Toutes ces informations suggèrent que les émotions ont un impact sur les différents paramètres acoustiques de la voix et de la parole. Grâce à la littérature consultée, nous pouvons à présent étudier l'évolution de ces paramètres lorsqu'ils sont envisagés plus précisément avec l'interprétation des 6 émotions universelles de Paul Ekman.

La peur

Pour cette émotion, l'ensemble des études réalisées ont mis en évidence les mêmes observations. La peur serait associée à un niveau d'excitation élevé qui se manifesterait dans la voix par une augmentation de sa fréquence fondamentale et de l'étendue de celle-ci (Banse &

¹² Robert W. Frick fut le premier auteur à s'intéresser à la prosodie et ses différents paramètres, ainsi qu'à sa dimension émotionnelle en 1985 (Bänziger & Scherer, 2005).

Scherer, 1996). La peur serait également marquée par une haute intensité (Sobin & Alpert, 1999). Une irrégularité de la voix peut être présente en raison de la perturbation de la respiration (Williams & Stevens, 1972 cités par Murray & Arnott, 1993). Cette dernière information peut nous faire inférer que le CPPs est susceptible de diminuer avec cette émotion. Ces observations seraient également présentes dans les formes plus douces de la peur, telles que l'inquiétude ou l'anxiété (Banse & Scherer, 1996). Dans une forme prolongée de la peur, c'est-à-dire l'angoisse, Johnson et al. (1986) cités par Murray et Arnott (1993), mettent également en évidence une augmentation de la fréquence fondamentale.

A propos des paramètres de la fluence, la vitesse articuloire et le débit de parole (Murray & Arnott, 1993) sont accélérés lorsqu'une personne ressent de la peur.

La colère

En ce qui concerne les paramètres de la voix, la colère se manifeste par une augmentation de l'intensité (Sobin & Alpert, 1999), de la fréquence fondamentale et de son étendue dans la voix (Banse & Scherer, 1996). Davitz (1964), cité par Murray et Arnott (1993) souligne que la colère se manifeste notamment par une sonorité forte et une hauteur plutôt aiguë. Il est important de noter que l'intensité moyenne est corrélée positivement à la colère. Ainsi, plus la colère est forte, plus l'intensité de la voix sera élevée (Bänziger & al., 2015). En ce qui concerne le LTAS, la colère serait associée à des valeurs de rapport L1-L0 plus élevées (Antonetti & al., 2020). Frick (1985), cité par Murray et Arnott (1993), aurait distingué deux émotions différentes au sein de la colère : la menace et la frustration. La menace présenterait une fréquence fondamentale plus faible alors que la frustration aurait une valeur plus élevée pour ce même paramètre.

Pour les paramètres de la fluence, la vitesse articuloire augmente avec la colère en règle générale (Banse & Scherer, 1996). La prosodie est finalement marquée par un rythme rapide et irrégulier (Davitz, 1964 cité par Murray & Arnott, 1993).

La joie

Cette émotion est associée à une excitation élevée. Ainsi, au niveau des paramètres de la voix, les valeurs moyennes de l'intensité (amplitude), de la fréquence fondamentale et de son étendue augmentent (Banse & Scherer, 1996). Davitz (1964), cité par Murray et Arnott (1993) aurait associé la joie à une augmentation de l'intensité de la voix. De plus, les fréquences aigües seraient corrélées positivement avec les émotions positives comme la joie (Frick, 1985). Pour le LTAS, la joie serait liée à des hautes valeurs de rapport L1-L0 (Antonetti & al., 2020). Chez

tous les locuteurs, la valeur de la fréquence fondamentale augmente lorsqu'on sourit. Le sourire peut également augmenter la valeur de l'intensité, mais ce n'est pas le cas pour tous les individus (Tartter, 1980 cité par Murray & Arnott, 1993).

Les paramètres de la fluence de la parole sont aussi modifiés lorsque la joie est interprétée. En effet, la vitesse articulatoire augmente (Banse & Scherer, 1996) tout comme le débit de parole, et la prosodie se caractérise par un rythme régulier (Davitz, 1964 cité par Murray & Arnott, 1993). Le rythme de l'énonciation peut être impacté par la production d'un sourire (Tartter, 1980 cité par Murray & Arnott, 1993).

La tristesse

Au niveau des paramètres de la voix, cette émotion serait associée à une diminution de l'intensité vocale (Sobin & Alpert, 1999), de la fréquence fondamentale et de son étendue et ce, à travers différentes langues et cultures (Banse & Scherer, 1996 ; Vos & al., 2023). Une diminution de l'intensité est également observée (Murray & Arnott, 1993). L'intensité moyenne serait négativement corrélée à la tristesse, ce qui signifie que plus nous sommes tristes, plus l'intensité diminue (Bänziger & al., 2015). Cela remet en question les observations plus anciennes réalisées par Banse et Scherer (1996) sur une forme plus forte de cette émotion : le désespoir. Pour cette dernière, les auteurs mettaient en évidence une augmentation des valeurs moyennes de la fréquence fondamentale et de l'intensité.

Quant à l'impact de la tristesse sur la fluence de la parole, l'énonciation peut être qualifiée de « brouillée », ce qui indique que l'intelligibilité de la parole est altérée (Fonagy & Davitz cités par Murray & Arnott, 1993). La vitesse articulatoire diminue (Banse & Scherer, 1996) et les caractères prosodiques de la tristesse suggèrent que le rythme est irrégulier (Murray & Arnott, 1993).

Le dégoût

Les observations faites sur cette émotion sont assez incohérentes dans la littérature. En effet, concernant les paramètres de la voix, certains auteurs mettent en évidence une augmentation de la fréquence fondamentale avec le dégoût, tandis que d'autres notent une diminution de celle-ci (Banse & Scherer, 1996). Fairbanks et Hoaglin (1941) cités par Murray et Arnott (1993) mettent en évidence une diminution de l'intensité avec cette émotion. Lorsqu'un individu ressent du dégoût, la voix peut être qualifiée de « compressée », car elle est le résultat d'une action musculaire spasmodique des muscles de la périphérie laryngée, entraînant alors la

perturbation de la vibration des plis vocaux (Fonagy, 1981 cité par Murray & Arnott, 1993 ; Trojan, 1952 cité par Murray & Arnott, 1993).

En ce qui concerne les paramètres de la fluence, Fairbanks et Hoaglin (1941), cités par Murray et Arnott (1993), notent que le débit de parole est ralenti à cause d'un temps de phonation prolongé et d'une augmentation de la durée des pauses.

La surprise

Pour cette émotion, seules des informations sur les paramètres de la voix ont été récoltées. Dans la surprise, la fréquence fondamentale serait ascendante dans un premier temps, puis descendante dans un second temps (Fonagy & Magdics, 1963 cités par Murray & Arnott, 1993). Sa valeur moyenne peut augmenter avec la surprise, mais ce n'est pas le cas pour tous les individus. L'étendue de la fréquence fondamentale augmente et le début de la phrase porte une forte accentuation. De plus, la voix est haletante avec cette émotion (Oster & Risberg, 1986 cités par Murray & Arnott, 1993).

En ce qui concerne les paramètres acoustiques LTAS et CPPs, nous disposons de très peu de renseignements sur leur évolution avec les 6 émotions étudiées. Ainsi, la plupart des hypothèses et les observations faites à leur sujet dans le cadre de ce mémoire sont exploratoires.

Il est important de se rappeler que l'ensemble des recherches effectuées dans la littérature et expliquées précédemment ont été entreprises afin d'obtenir des données pour développer et enrichir la base de données « EVE ». En effet, cette base de données émotionnelle, réalisée par Madame Etienne et dans laquelle notre mémoire s'inscrit, s'inspire de nombreuses bases de données qui ont déjà vu le jour dans la recherche traitant des émotions dans le signal de parole. La méthodologie de création de certaines d'entre elles ont inspiré Madame Etienne et nous-mêmes, afin de mettre en place ce projet ambitieux. Quelques-unes d'entre elles seront présentées dans les prochains paragraphes.

2.5 Les bases de données émotionnelles

2.5.1 Méthodologie de création des bases de données émotionnelles

Dans cette section, les différentes méthodologies d'élaboration des bases de données émotionnelles recensées dans la littérature sont décrites. Un tableau récapitulatif reprenant les différentes méthodologies d'élaboration de quelques-unes d'entre elles se trouvent dans les annexes (**cf. Annexe 2**).

Les types de contenus des bases de données émotionnelles

La majorité des bases de données émotionnelles abordées dans la littérature sont multimodales, c'est-à-dire qu'elles récoltent des données émotionnelles provenant de sources différentes. La plupart d'entre elles recueillent des données audio-visuelles afin d'obtenir des informations complémentaires sur l'expression des émotions. Ainsi, elles prennent en considération les enregistrements de voix et de parole des participants, mais aussi leurs expressions faciales et corporelles. Certaines méthodologies de création de bases de données évaluent et distinguent la qualité des données émotionnelles obtenues, en mettant les participants dans différentes conditions d'enregistrement : audio-visuelle, uniquement audio ou uniquement visuelle.

Les types de participants sélectionnés pour interpréter les émotions étudiées

Les participants sélectionnés pour la production d'émotions sont principalement des groupes d'adultes composés à 50% d'hommes et 50% de femmes. Il leur est très souvent demandé d'avoir de l'expérience dans le milieu du théâtre et/ou de l'acting. En effet, les études initiales qui ont été effectuées dans le domaine de la production orale des émotions supposent que les acteurs professionnels sont capables d'interpréter n'importe quel type d'émotion de manière réaliste (Jürgens & al., 2015). Néanmoins, il n'existerait pas de réelle différence d'interprétation « orale » des émotions entre des individus tout-venants et des acteurs professionnels (Jürgens & al., 2015). Ainsi, certaines bases de données comme « Emo-DB » (Burkhardt & al., 2005) ont fait intervenir des participants tout-venants afin d'obtenir des données sur les capacités interprétatives des émotions de l'ensemble de la population.

De plus, nous pouvons remarquer que les différentes bases de données émotionnelles abordées dans la littérature ont été élaborées grâce à la participation d'individus parlant différentes langues, notamment l'anglais, le français, l'allemand, l'espagnol, l'italien, le québécois, ... Celles-ci ont permis de diversifier les banques de données émotionnelles, mais également d'étudier les différences et similarités interculturelles dans la production et la perception des émotions.

Les méthodes d'induction des émotions

En ce qui concerne les techniques d'induction des émotions, différentes méthodes sont utilisées dans les bases de données disponibles. La plupart des émotions sont interprétées via la production de phrases sémantiquement neutres, provenant de différents corpus équilibrés au niveau du nombre de syllabes et de phonèmes (ex : corpus de Fharvard Aubanel & al., 2020).

Il est donc demandé aux acteurs d'interpréter chaque phrase selon chaque émotion décrite. D'autres banques de données ont intégré des données émotionnelles qui ont été induites grâce à l'utilisation de phrases à contenu émotionnel. Toutefois, nous pouvons remarquer que certaines bases de données utilisent d'autres stratégies afin d'induire les émotions. C'est le cas de la base de données "IEMOCAP"¹³, qui s'est penché sur l'étude de la communication entre des dyades d'acteurs afin d'analyser l'interaction des émotions avec leurs gestes, leur parole, ou encore leurs expressions faciales. Une autre base de données "GAUDIE"¹⁴, a été construite sur base de mises en situations menant à un échec ou une réussite (technique non-standardisée¹⁵), afin d'induire des émotions plutôt positives ou négatives (Lingelbach et al., 2023).

Les émotions étudiées

Au niveau des émotions, nous pouvons relever que, pour une grande partie des bases de données abordées dans la littérature, nous retrouvons les 6 émotions « de base » d'Ekman (peur, joie, tristesse, surprise, dégoût, colère). En effet, certaines bases de données étudient ces émotions, car elles sont reconnues comme universelles, alors que d'autres se penchent davantage sur des états émotionnels plus subtils comme la frustration, la dépression, la jalousie, ... Il est important de noter que les émotions sont parfois mises en relation avec la condition "neutre" (condition sans interprétation d'émotion), dans le but de comparer les données obtenues entre les conditions émotionnelles et cette condition contrôle. Enfin, il est régulièrement demandé aux acteurs d'interpréter ces émotions selon différentes intensités émotionnelles (basse et haute).

¹³ Corpus anglais de données émotionnelles créé en 2008 par la SAIL (Speech Analysis and Interpretation Laboratory) de l'Université de Californie du sud (USC). L'IEMOCAP comporte des milliers d'enregistrements de parole réalisés par des acteurs en dyades et selon 9 émotions : le bonheur, la colère, la tristesse, la frustration, la surprise, la peur, l'excitation, la neutralité, autres (Trinh Van & al., 2022).

¹⁴ Base de données reprenant 37 extraits de paroles en allemand provenant de 3 types de contextes afin de représenter des émotions positives, neutres et négatives et se basant sur l'utilisation de l'échelle d'auto-évaluation de l'émotion ressentie (SAM) (Self-Assessment Manikin) (Bradley et Lang, 1994 cités par Lingelbach & al., 2023).

¹⁵ Pour rappel, les techniques non-standardisées permettent d'induire les émotions sans utiliser de stimuli standardisés, car elles varient en fonction des caractéristiques propres à l'environnement. Elles induisent plutôt des états émotionnels en termes de valence (sur un continuum positif-négatif) que des émotions discrètes (Gil, 2009).

Processus de validation de ces bases de données

Afin de valider la conception de ces différentes bases de données émotionnelles, diverses méthodes d'évaluation des émotions ont été privilégiées. C'est le cas des échelles d'auto-évaluation telles que la SAM (« Self Assessment Manikin » Bradley & Lang, 1994 cités par Lingelbach & al., 2023), qui permettent aux participants de juger les émotions qu'ils ont ressenties, ou encore les tâches de reconnaissance des émotions d'autrui par un public tout-venant. Finalement, nous pouvons constater qu'au fur et à mesure des années, les bases de données développées ont diversifié leurs méthodologies, de façon à pouvoir enrichir les données dont nous disposons dans la littérature traitant de la production des émotions.

2.5.2 Les bases de données émotionnelles existantes

Ainsi, les bases de données émotionnelles et leurs méthodes de conception sont très nombreuses et diversifiées dans la littérature scientifique. Néanmoins, certaines d'entre elles ont particulièrement retenu notre attention, car leur méthodologie a inspiré le développement de la base de données « EVE », pour laquelle nous nous sommes investies.

Les bases de données en français

CaFE (A Canadian French Emotional Speech Dataset) (Gournay, Lahaie, Lefebvre, 2018)

Ce corpus a été réalisé dans le but d'obtenir une base de données émotionnelle en français. Pour ce faire, 12 acteurs (6 hommes et 6 femmes, moyenne d'âge= 34 ans) ont été recrutés. Chacun d'eux a été invité à produire 6 phrases différentes, équilibrées au niveau syllabique et à contenu sémantique neutre (c'est-à-dire non-lié à des situations émotionnelles). Ces phrases ont été produites avec les 6 émotions « de base » d'Ekman ainsi que l'état neutre, et selon 2 intensités émotionnelles différentes (faible et intense). Ainsi, chaque acteur a produit 6 phrases x 2 intensités x 7 conditions (6 conditions émotionnelles et la condition neutre). Pour chaque phrase, 2 à 5 enregistrements ont été réalisés selon les demandes des expérimentateurs et des acteurs, afin de conserver l'audio correspondant au mieux à ce que l'acteur souhaitait produire (Gournay & al., 2018).

EMOV-DB (The Database of Emotional Voices) (Adigwe et al., 2018)

Cette base de données a été construite grâce au recrutement de 5 acteurs. L'échantillon se composait de 4 personnes (2 hommes et 2 femmes) ayant l'anglais comme langue maternelle et d'une personne d'origine belge et parlant le français (1 homme). Les 5 participants ont été amenés à lire des phrases provenant de différents corpus équilibrés phonétiquement et selon 5

états émotionnels. Le corpus « CMU-arctic » (Kominek & Black, 2004 cités par Adigwe et al., 2018) a été utilisé pour les natifs anglais et le corpus « SIWIS » (Honnet & al., 2017 cités par Adigwe et al., 2018) a été proposé à la personne parlant le français. Les conditions émotionnelles étudiées dans cette base de données étaient la joie, la colère, le dégoût, et la somnolence. L'état neutre était également mesuré.

Les autres bases de données émotionnelles

De nombreuses bases de données émotionnelles ont été créées et sont disponibles en libre accès via différentes plateformes en ligne. Les bases de données les plus récentes et respectant une méthodologie semblable à celle appliquée dans le cadre de notre travail sont décrites en annexe (**cf. Annexe 2**). Celles-ci nous ont permis de prendre en considération les dernières données dont nous disposons actuellement à propos des méthodologies de conception des bases de données émotionnelles.

Dans les prochaines sections de ce mémoire, nous développerons les objectifs et hypothèses de notre projet. Nous aborderons également la méthodologie précise qui a été mise en place afin de concevoir la base de données « EVE » et de réaliser l'étude des paramètres acoustiques de la voix et de la parole avec les 6 émotions universelles de Paul Ekman. Enfin, nous présenterons les résultats issus de cette étude, et discuterons des biais et des limites que nous avons rencontrés, afin que ceux-ci soient pris en considération si de nouvelles études ou de futurs projets de mémoire s'intéressent également à l'étude de la manifestation des émotions dans la voix et la parole.

3. OBJECTIFS ET HYPOTHÈSES

3.1 Objectifs

La réalisation de ce projet de mémoire vise à atteindre plusieurs objectifs. Tout d'abord, notre étude contribue à la création d'une base de données audio-visuelle des émotions appelée « EVE ». La partie audio de celle-ci est constituée d'un ensemble d'énoncés phonétiquement équilibrés, qui ont été produits avec différentes émotions par des acteurs sélectionnés selon des critères précis. Les énoncés à produire provenaient du corpus de Fharvard (Aubanel & al., 2020), qui est l'adaptation française du corpus d'Harvard (Rothausser & al., 1969). La partie visuelle comporte les enregistrements vidéo des acteurs lorsqu'ils prononçaient les énoncés avec les diverses émotions. Cependant, seules les données audios seront analysées et utilisées dans le cadre ce mémoire. Les données visuelles récoltées enrichiront la base de données pour d'autres études ultérieures éventuelles.

Ensuite, le second objectif de notre projet consiste à étudier l'influence de la production des 6 émotions de Paul Ekman sur les paramètres de la voix et de la parole de 4 acteurs de notre échantillon. En effet, réaliser un travail de recherche sur ce thème est nécessaire, car nous constatons un manque de consensus entre les auteurs qui se sont intéressés à ce sujet dans la littérature consultée (**cf. Tableau 1**). De plus, la majorité des données bibliographiques dont nous disposons sont relativement anciennes (Larrouy-Maestri & al., 2024). Ainsi, la littérature scientifique ne nous permet pas encore de caractériser avec précision l'évolution des paramètres acoustiques de la voix et de la parole avec la production d'émotions, et nous réalisons notre mémoire pour tenter d'atteindre cet objectif.

Enfin, la majorité de la littérature qui s'est penchée sur ce sujet tirait ses conclusions à partir d'enregistrements vocaux, de bases de données émotionnelles et de corpus anglophones. Ainsi, l'élaboration de ce mémoire permet d'enrichir les données dont nous disposons actuellement, en étudiant l'évolution des paramètres acoustiques de voix et de parole lorsqu'ils sont envisagés en français. De ce fait, grâce à l'analyse acoustique et au traitement des enregistrements de la voix des acteurs que nous avons recrutés, les données obtenues nous ont permis d'expliquer comment les paramètres de la voix et de la parole caractérisent l'expression de chaque émotion étudiée.

3.2 Hypothèses

Comme abordé précédemment, certains auteurs mettent précisément en évidence comment les paramètres de la voix et de la parole sont modifiés avec la production d'émotions. Pour ceux-

ci, les hypothèses posées sont confirmatoires. Pour certains paramètres acoustiques cependant, nous disposons de trop peu de données, ou constatons quelques résultats contradictoires. Pour ceux-ci, nos hypothèses sont alors exploratoires. Les données issues de la littérature consultée sont résumées dans le tableau ci-dessous (cf. **Tableau 1**).

Tableau 1. Données issues de la littérature sur l'évolution des paramètres acoustiques avec les émotions de Paul Ekman.

	Fréquence fondamentale		Intensité		LTAS		CPPs	Débit	Vitesse articulatoire
	étendue (Hz)	médiane (Hz)	étendue (dB)	médiane (dB)	non voisement (%)	rapport L1-L0 (dB)	(dB)	(nombre syllabes /sec)	(nombre syllabes /sec) sans pauses
	<i>Banse & Scherer (1996)</i> (1)	<i>Banse & Scherer (1996)</i> (1)	/	<i>Banse & Scherer (1996)</i> (1) <i>Bänziger & al., (2015)</i> (3)	/	<i>Antonetti & al. (2020)</i> (6)	/	<i>Banse & Scherer (1996)</i> (1) <i>Bänziger & al., (2015)</i> (3)	<i>Banse & Scherer (1996)</i> (1) <i>Murray & Arnott (1993)</i> (4)
	<i>Vos et al., (2023)</i> (2)	<i>Frick, (1985)</i> (7) <i>Murray & Arnott (1993)</i> (4) <i>Vos et al., (2023)</i> (2)		<i>Murray & Arnott (1993)</i> (4) <i>Sobin & Alpert (1999)</i> (5)				<i>Murray et Arnott (1993)</i> (4)	
Δ Peur	Δ+ (1)	Δ+ (1)	/	Δ+ (5)	/	/	/	Δ+ (4)	Δ+ (4)
Δ Colère	Δ+ (1)	Δ+ (1)(2)(4)	/	Δ+ (3)(4)(5)	/	Δ+ (6)	/	Δ+ (1)	Δ+ (1)
Δ Joie	Δ+ (1)	Δ+ (1)(2)(7)	/	Δ+ (1)(4)	/	Δ+ (6)	/	Δ+ (1)(3)	Δ+ (1)
Δ Tristesse	Δ- (1) (2)	Δ-(2) ou Δ+ (1)	/	Δ- (3)(4)(5) ou Δ+(1)	/	/	/	/	Δ- (1)
Δ Dégoût	/	Δ- ou Δ+ (1)	/	Δ- (4)	/	/	/	Δ- (4)	/
Δ Surprise	Δ+ (4)	Δ+ (4)	/	/	/	/	/	/	/

Note. / = données absentes dans la littérature consultée. Celles-ci feront l'objet de nos hypothèses exploratoires ; Δ+ = delta positif ; Δ- = delta négatif ; Δ- ou Δ+ = manque de consensus dans la littérature consultée. Celles-ci feront l'objet de nos hypothèses exploratoires ; Numéros en gris clair = liens entre les hypothèses posées et leurs auteurs

Nos hypothèses ont été testées grâce à l'utilisation des deltas. Sur base de l'observation des différences de chaque paramètre présenté dans le tableau 1, nos hypothèses ont été testées en calculant l'écart entre la condition neutre (condition contrôle) et chacune des 6 émotions. Ainsi, nous parlons de delta « positif » ($\Delta+$), lorsque la différence obtenue pour le paramètre acoustique est positive entre chaque émotion ciblée et la condition neutre. Nous parlons de delta « négatif » ($\Delta-$), lorsque la valeur obtenue pour l'émotion ciblée est inférieure à celle de la condition neutre. Cette méthode d'analyse sera développée davantage dans la partie méthodologique de ce mémoire.

3.2.1 Hypothèses sur l'évolution de f_o selon l'émotion produite

Concernant ce paramètre, nous nous sommes intéressées à l'évolution de la valeur de sa médiane (tendance centrale), ainsi que de sa variation (étendue). La littérature consultée à ce sujet nous a permis de dresser plusieurs hypothèses confirmatoires à propos de l'évolution de ce paramètre acoustique, avec la production des 6 émotions universelles de Paul Ekman.

- *Nous nous attendons à ce qu'en moyenne, les valeurs de **médiane et d'étendue de f_o** soient plus élevées lorsque **la peur, la colère, la joie et la surprise** sont produites, par rapport à la condition neutre. Ainsi, pour ces quatre émotions, **le delta devrait être positif ($\Delta+$)** entre ces deux conditions pour ces deux paramètres.*
- *Nous nous attendons à ce qu'en moyenne, la valeur de **l'étendue de f_o** soit moins élevée lorsque **la tristesse** est produite, par rapport à la condition neutre. Ainsi, **le delta devrait être négatif ($\Delta-$)** entre ces deux conditions pour ce paramètre.*

En ce qui concerne l'évolution de f_o (médiane et étendue) avec la production du dégoût, les auteurs ne s'accordent pas sur sa manifestation. Il en est de même pour l'évolution de la valeur médiane de ce paramètre avec la tristesse. Ainsi, nos hypothèses sont exploratoires pour celles-ci.

- *Existe-t-il, en moyenne, une différence des valeurs de **médiane de f_o** lorsque **la tristesse** est produite, par rapport à la condition neutre ?*
- *Existe-t-il, en moyenne, une différence des valeurs de **médiane et d'étendue de f_o** lorsque **le dégoût** est produit, par rapport à la condition neutre ?*

3.2.2 Hypothèses sur l'évolution de l'intensité avec la production d'émotions

En ce qui concerne ce paramètre, nous nous sommes également intéressées à l'évolution de la valeur de sa médiane, ainsi que son étendue. La littérature consultée nous a permis de dresser plusieurs hypothèses confirmatoires à propos de l'évolution de ce paramètre acoustique avec la production des 6 émotions de Paul Ekman.

- *Nous nous attendons à ce qu'en moyenne, la valeur de la **médiane de l'intensité** soit plus élevée lorsque **la peur, la colère et la joie** sont produites, par rapport à la condition neutre. Ainsi, pour ces trois émotions, **le delta devrait être positif ($\Delta+$)** entre ces deux conditions.*
- *Nous nous attendons à ce qu'en moyenne, la valeur de la **médiane de l'intensité** soit plus faible lorsque **le dégoût** est produit, par rapport à la condition neutre. Ainsi, **le delta devrait être négatif ($\Delta-$)** entre ces deux conditions.*

Par rapport à l'évolution de la valeur de l'étendue de l'intensité, la littérature consultée ne nous renseigne pas sur sa manifestation. Néanmoins, les données dont nous disposons pour les autres émotions mettent en évidence que l'étendue de l'intensité évolue généralement dans le même sens que sa médiane. Ainsi, nous pouvons déduire de nouvelles hypothèses confirmatoires pour ce paramètre.

- *Nous nous attendons à ce qu'en moyenne, la valeur de **l'étendue de l'intensité** soit plus élevée lorsque **la peur, la colère et la joie** sont produites, par rapport à la condition neutre. Ainsi, pour ces trois émotions, **le delta devrait être positif ($\Delta+$)** entre ces deux conditions.*
- *Nous nous attendons à ce qu'en moyenne, la valeur de **l'étendue de l'intensité** soit plus faible lorsque **le dégoût** est produit, par rapport à la condition neutre. Ainsi, **le delta devrait être négatif ($\Delta-$)** entre ces deux conditions.*

À propos de la modification de l'étendue et de la médiane de l'intensité avec la production de la tristesse et de la surprise, nous n'avons pas obtenu de données sur leur manifestation. Ainsi, nos hypothèses sont exploratoires pour celles-ci.

- *Existe-t-il, en moyenne, une différence des valeurs de **médiane et d'étendue de l'intensité** lorsque **la tristesse et la surprise** sont produites par rapport à la condition neutre ?*

3.2.3 Hypothèses sur l'évolution du LTAS avec la production d'émotions

En ce qui concerne ce paramètre, nous nous sommes intéressées à l'évolution du rapport L1-L0 ainsi qu'au pourcentage de non-voisement. La littérature consultée nous a permis de dresser des hypothèses confirmatoires à propos de l'évolution de ce paramètre acoustique avec la production de la colère et de la joie.

- *Nous nous attendons à ce qu'en moyenne, la valeur du **rapport L1-L0** soit plus élevée lorsque **la colère et la joie** sont produites, par rapport à la condition neutre. Ainsi, pour ces deux émotions, **le delta devrait être positif ($\Delta+$)** entre ces deux conditions.*

Par rapport à l'évolution de la valeur du rapport L1-L0 et du pourcentage de non-voisement, la littérature consultée ne nous permet pas de dresser des hypothèses confirmatoires. Nos hypothèses sont donc exploratoires pour les autres émotions.

- *Existe-t-il, en moyenne, une différence des valeurs du **rapport L1-L0** et du **pourcentage de non-voisement** lorsque **la peur, la tristesse, le dégoût et la surprise** sont produits, par rapport à la condition neutre ?*
- *Existe-t-il, en moyenne, une différence de valeur du **pourcentage de non-voisement** lorsque **la colère et la joie** sont produites, par rapport à la condition neutre ?*

3.2.4 Hypothèses sur l'évolution du CCPs avec la production d'émotions

La littérature consultée ne nous a pas permis de dresser des hypothèses confirmatoires à propos de l'évolution de ce paramètre acoustique avec la production des 6 émotions « de base » d'Ekman. Pour ce paramètre, nos hypothèses sont exploratoires.

- *Existe-t-il, en moyenne, une différence de valeur du **CCPs** lorsque **la peur, la colère, la joie, la tristesse, le dégoût et la surprise** sont produites, par rapport à la condition neutre ?*

3.2.5 Hypothèses sur l'évolution du débit de la parole avec la production d'émotions

La littérature consultée nous a permis de dresser plusieurs hypothèses confirmatoires à propos de l'évolution de ce paramètre de fluence de la parole avec la production de l'émotion de joie et de peur.

- *Nous nous attendons à ce qu'en moyenne, la valeur du **débit de parole** soit plus élevée lorsque **la peur, la colère et la joie** sont produites, par rapport à la condition neutre. Ainsi, pour ces trois émotions, **le delta devrait être positif ($\Delta+$)** entre ces deux conditions.*
- *Nous nous attendons à ce qu'en moyenne, la valeur du **débit de parole** soit plus faible lorsque **le dégoût** est produit, par rapport à la condition neutre. Ainsi, **le delta devrait être négatif ($\Delta-$)** entre ces deux conditions.*

Pour les autres émotions, nos hypothèses sont exploratoires.

- *Existe-t-il, en moyenne, une différence de la valeur du **débit de parole** lorsque **la tristesse et la surprise** sont produites, par rapport à la condition neutre ?*

3.2.6 Hypothèses sur l'évolution de la vitesse articulatoire avec la production d'émotions

En consultant la littérature au sujet de ce paramètre, nous avons pu dresser plusieurs hypothèses confirmatoires à propos de l'évolution de la vitesse articulatoire avec la production des émotions de Paul Ekman.

- *Nous nous attendons à ce qu'en moyenne, **la vitesse articulatoire** soit plus élevée lorsque **la peur, la colère et la joie** sont produites, par rapport à la condition neutre. Ainsi, pour ces trois émotions, **le delta devrait être positif ($\Delta+$)** entre ces deux conditions.*

- *Nous nous attendons à ce qu'en moyenne, la **vitesse articulatoire** soit plus faible lorsque la **tristesse** est produite, par rapport à la condition neutre. Ainsi, le **delta devrait être négatif** (Δ -) entre ces deux conditions.*

Pour les autres émotions abordées, les observations réalisées nous permettront d'en savoir plus sur son évolution. Nos hypothèses sont donc exploratoires pour le dégoût et la surprise.

- *Existe-t-il, en moyenne, une différence de la valeur de la **vitesse articulatoire** lorsque le **dégoût et la surprise** sont produits, par rapport à la condition neutre ?*

4. MÉTHODOLOGIE

4.1 Participants

Afin de créer les enregistrements de parole constituant la base de données « EVE », 12 acteurs normo-phoniques (6 hommes et 6 femmes) ont été recrutés. Néanmoins, dans le cadre de notre mémoire, nous réalisons notre étude sur un plus petit échantillon composé de 4 acteurs (2 hommes ; moyenne d'âge = 43,5 ans et 2 femmes ; moyenne d'âge = 29,5 ans) (cf. **Tableau 2**). Il était important d'avoir autant d'énoncés prononcés par des voix d'hommes que de femmes, car les paramètres acoustiques de la voix (ex : la fréquence fondamentale) varient en fonction de différents éléments, dont entre autres, le genre de l'individu qui prend la parole (Kramer, 1963 ; Kushner & Bickley, 1995 ; Trent, 1995 cité par Murray & Arnott, 1993).

Pour que l'échantillon de participants soit le plus représentatif et homogène possible, des critères d'inclusion ont été définis pour le recrutement : être majeur, avoir plusieurs années d'expérience dans le domaine de la scène et/ou du théâtre (3 ans au minimum) et être toujours actif dans ce secteur. Les acteurs devaient également posséder une carte « artiste », afin de pouvoir leur facturer l'indemnité de 100 € prévue pour le déplacement effectué jusqu'au studio et pour le temps consacré pour l'enregistrement.

Leur langue maternelle devait être le français et ils devaient disposer d'un moyen de déplacement pour se rendre au studio d'enregistrement. Aucun critère d'exclusion n'a été précisément défini lors de la sélection. Cependant, si le discours de l'acteur n'était pas assez intelligible, ses données n'auraient pas été analysées et interprétées.

Le recrutement des différents acteurs a été réalisé via les réseaux sociaux. Plus précisément, une fiche informative reprenant les différents critères cités précédemment a été publiée sur ces supports.

Les participants sélectionnés pour l'étude ont reçu par e-mail un formulaire de consentement, un document reprenant les informations pratiques du déroulement de la séance d'enregistrement, ainsi qu'une attestation de facturation pour l'indemnisation de 100 €.

La période d'enregistrement s'est déroulée du 15 octobre au 25 novembre 2023. Les rendez-vous étaient fixés et convenus via la complétion d'un agenda en ligne.

Tableau 2. Description de l'échantillon d'acteurs participant au projet « EVE ».

Acteurs (n=12)	Âge (années)	Genre	Années de théâtre/scène
1	24	F	12
2	30	M	10
3	42	M	10
4	54	F	3
5	56	M	35
6	55	F	38
7	35	F	12
8	57	M	15
9	35	M	12
10	41	M	10
11	29	F	12
12	26	F	5

Note. M = homme ; F = femme ; n= nombre ; Données en gras =échantillon des 4 acteurs sélectionnés dans ce mémoire

4.2 Corpus

Le corpus de Fharvard (Aubanel & al., 2020) comporte les phrases qui ont été prononcées par les acteurs lors des séances d'enregistrement. Ce corpus, adapté de la version anglaise « Harvard » (Rothausser & al., 1969 cités par Aubanel & al., 2020), se compose de 700 phrases en français, rassemblées en 70 listes de 10 phrases. Ces regroupements ont été réalisés à l'aide d'un équilibrage automatique des phonèmes. Les phrases sont toutes construites à l'aide de 5 mots-clés, définis pour leur fonction et leur occurrence dans la langue française. La variabilité des structures syntaxiques, la diversité lexicale et l'homogénéité de la difficulté ont été contrôlées. Le corpus de Fharvard centralise également des enregistrements de la voix d'une femme et d'un homme, prononçant chacun les 700 phrases. Ce dispositif a été utilisé à de nombreuses reprises dans la recherche sur l'intelligibilité de la parole et dans la pratique de l'audiologie et fournirait un bon niveau de sensibilité pour l'évaluation de l'audiométrie vocale (Aubanel & al., 2020). Ainsi, la liste 1 du corpus de Fharvard a été sélectionnée pour nos participants (cf. **Tableau 3**).

Tableau 3. Liste des 10 phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020).

1.	La lampe de néon rouge irise ses cheveux.
2.	La nacelle du ballon tire sa souplesse de l'osier.
3.	Devine qui a volé le précieux collier de jade.
4.	C'est facile d'atteindre la branche en sautant du muret.
5.	Ne collez jamais d'affiche sur le mur de la poste.
6.	Découpe de fines lamelles de cette truffe noire.
7.	Une croisière dans un grand bateau, c'est un bonheur parfait.
8.	Une brèche dans le mur laissait entrer de l'air.
9.	Le groupe d'amis s'est quitté devant le kiosque.
10.	La voiture de course partit en trombe sur la piste.

4.3 Matériel

Les enregistrements des acteurs ont été réalisés dans le studio de HEC à Liège. Ce studio dispose d'une bonne isolation acoustique, afin que les bruits environnants ne perturbent pas le signal audio de la voix des acteurs. Néanmoins, il est important de noter que nous n'avons pas effectué de mesures du bruit ambiant de ce local.

L'acteur était placé précisément à 20 centimètres d'un mur du studio grâce à des marquages au sol. Il était préférable de respecter cet agencement afin d'éviter que l'ombre des candidats soit visible sur les enregistrements visuels.

Une fois le positionnement du participant effectué, nous réalisons le placement du micro (MicroMic The original C 544 L vocal – AKG by HARMAN). Ce dernier était installé à 5 centimètres de la commissure droite de la bouche de l'acteur dans le but d'obtenir une distance micro-bouche suffisamment grande pour empêcher le bruit, causé par le souffle, de se répandre dans le signal audio de leur voix. La carte-son externe (Focusrite iTrack Solo Lightning) a été calibrée sur les paramètres « minimum » (cf. **Annexe 5**), afin d'empêcher que le son capté par le micro ne sature lorsque le participant s'exprimait à forte intensité. Ces calibrages permettaient alors d'éviter que le spectrogramme ne soit écrêté et ininterprétable. Le fil reliant le micro au boîtier était suffisamment long pour permettre à l'acteur d'être libre de ses mouvements. Cependant, avant chaque enregistrement, nous lui demandions d'éviter de réaliser des mouvements parasites susceptibles d'empiéter sur la qualité du signal audio (ex : faire du bruit avec ses mains, ses pieds, ses vêtements, ...).

Les dix phrases à produire de la première liste du corpus de Fharvard (Aubanel & al., 2020) devaient être prononcées avec les 10 émotions étudiées dans ce projet et de façon neutre. Pour ce faire, un support visuel (Ipad pro Apple) était disposé devant le participant. Sur l'écran de celui-ci, les phrases à produire défilaient à l'aide d'une télécommande manipulée par l'expérimentateur.

Une caméra (Obsbot tiny 0WB-2004-CE) disposée en face de chaque acteur filmait le haut de leur buste et leur visage. Ce dispositif nous permettait d'enregistrer les expressions faciales des participants. Comme expliqué précédemment, ces données visuelles ne seront pas développées dans ce mémoire. Mais, elles permettront d'obtenir une base de données complète pour chaque émotion qui pourra éventuellement être exploitée ultérieurement pour d'autres études.

L'utilisation d'un ordinateur (MacBook Pro 2.3 GHz Intel Core i5 Dual CoreN) permettait la gestion des différents outils exploités. Un logiciel (Camtasia v22.5.4 [2023 May 02]) nous donnait un feed-back en temps réel de la vidéo et du signal audio de l'acteur durant l'enregistrement.

4.4 Procédure générale

4.4.1 Essais et définition des paramètres avec des participants volontaires

Afin de mettre en place et de nous approprier cette méthodologie, nous avons réalisé 4 sessions d'enregistrement avec des personnes volontaires qui n'étaient pas des acteurs. Ces séances nous ont permis d'adapter les paramètres du matériel utilisé et de nous familiariser une première fois à son mode de fonctionnement.

4.4.2 Elaboration de la base de données avec les participants prenant part à l'étude

Une fois le matériel du studio et des participants correctement calibrés et maîtrisés, les séances d'enregistrement avec les acteurs pour créer la base de données « EVE » ont pu être lancées.

En pratique, les enregistrements se sont déroulés durant une session de 2 heures incluant une pause de 5 à 10 minutes, après avoir interprété les 5 premières émotions. Les acteurs ont été reçus au studio séparément afin d'éviter une quelconque influence par le jeu d'acteur des autres participants.

Durant les tests, les expérimentateurs avaient le rôle de vérifier les paramètres du matériel sur le logiciel, de démarrer les enregistrements et de les sauvegarder dans l'ordinateur. Ils devaient également donner les consignes et faire défiler les différentes phrases à produire à l'aide de la

télécommande. Dès qu'une production contenait des erreurs (ex : omission ou rajout de mots/de phonèmes), l'expérimentateur demandait à l'acteur de répéter l'énoncé.

Les séances d'enregistrement se déroulaient en deux temps. Durant la première phase, il était tout d'abord demandé au participant de se présenter en quelques mots afin de vérifier le calibrage des paramètres de la carte-son externe du micro. La consigne était la suivante : « *Présentez-vous-en quelques mots en donnant votre nom, votre âge et votre profession* ».

Ensuite, l'acteur était invité à réaliser le cri le plus intense qu'il était susceptible de produire durant l'enregistrement, pour vérifier que le signal audio n'était pas écrêté. La consigne était la suivante : « *Nous allons vous demander de réaliser le cri le plus fort que vous êtes susceptible de produire durant l'enregistrement. Il est donc inutile de réaliser un cri très intense et strident si vous ne pensez pas crier aussi fort lorsque vous interprétez les émotions* ».

Enfin, nous avons invité le participant à lire les 10 phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020) de manière neutre, afin d'obtenir des données sur les paramètres de la voix du participant quand il n'exprime pas d'émotions (**cf. Tableau 3**). Ces phrases serviraient alors de modèle de base (condition contrôle) pour réaliser la comparaison des paramètres de la voix et de parole des acteurs en modalité neutre et en modalité émotionnelle. La consigne était la suivante : « *Lisez les 10 phrases suivantes une fois chacune de manière neutre, sans interpréter une émotion* ».

Durant la seconde phase du test, les consignes sur le déroulement de l'enregistrement étaient répétées au participant pour nous assurer de sa bonne compréhension de la tâche. Les 10 phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020) ainsi que la description détaillée des 10 émotions à interpréter (définitions, intensité et exemples de mises en situation) figuraient dans le formulaire d'informations communiqué préalablement par e-mail (**cf. Tableau 3 ; Tableau 4 ; Annexe 1**). L'acteur était donc censé les avoir mémorisées et s'être préparé en s'appropriant les émotions une première fois avant l'enregistrement. Néanmoins, pour éviter les oublis éventuels, l'Ipad affichant les phrases à produire était disposé devant chaque participant.

Pour chaque acteur, les 10 phrases ainsi que les 10 émotions ont été distribuées aléatoirement afin d'empêcher que des effets de fatigue se déclarent pour la même phrase et la même émotion. Tout d'abord, les 10 émotions étaient interprétées l'une après l'autre, pour permettre au participant de rester concentré sur l'expression d'une émotion à la fois. Avant de débiter chaque

émotion, la définition était donnée oralement au participant pour bien se la représenter avant de l'interpréter (cf. **Tableau 4**).

Tableau 4. Définitions des 10 émotions intégrées dans la base de données émotionnelle « EVE » (Ekman, 2005 ; Geer & al., 2000 ; Hareli & al., 2018 ; Hess, 2003 ; Perry, 2011).

Emotion	Définition
Joie	Un état émotionnel positif caractérisé par le bonheur, la satisfaction.
Tristesse	Un état émotionnel négatif marqué par la mélancolie, la douleur émotionnelle.
Peur	Une émotion en réponse à une menace perçue ou à une situation dangereuse, provoquant de l'anxiété.
Dégoût	Une émotion de répulsion envers quelque chose de désagréable.
Colère	Un état émotionnel négatif caractérisé par l'irritation, la frustration.
Surprise	Une émotion momentanée causée par quelque chose d'inattendu.
Confiance en soi	Un sentiment d'assurance et de croyance en ses propres capacités et compétences.
Confusion	Un état mental marqué par un manque de clarté ou de compréhension.
Mépris	Une émotion de dédain ou de supériorité envers quelque chose ou quelqu'un.
Sympathie	Une émotion positive marquée par la compréhension et l'affection envers les émotions ou la situation d'une autre personne.

Ensuite, pour chaque phrase de la liste, l'acteur interprétait l'émotion avec deux intensités émotionnelles différentes : basse intensité et haute intensité (Gournay & al., 2018). Pour que le participant se représente les intensités, des exemples leur étaient proposés oralement (cf. **Tableau 5**). Néanmoins, les acteurs étaient libres de jouer l'intensité basse et haute de chaque émotion comme ils le désiraient. L'important était de percevoir une différence d'interprétation entre les deux intensités de chaque émotion.

Tableau 5. Description des intensités des 10 émotions intégrées dans la base de données émotionnelle « EVE » (Gournay & al., 2018).

Emotion	Intensité
Joie	Basse Intensité : C'est comme ressentir une douce chaleur dans la poitrine. Haute Intensité : C'est comme ressentir une envie incontrôlable de sauter de joie.
Tristesse	Basse Intensité : C'est comme avoir un nœud dans la gorge. Haute Intensité : C'est comme se sentir au bord d'une crise émotionnelle.
Colère	Basse Intensité : C'est comme sentir sa tension artérielle augmenter. Haute Intensité : C'est comme ressentir une colère bouillante et incontrôlable.
Peur	Basse Intensité : C'est comme avoir des papillons dans l'estomac. Haute Intensité : C'est comme se sentir paralysé par la peur.
Dégoût	Basse Intensité : C'est comme sentir quelque chose de désagréable mais supportable. Haute Intensité : C'est comme ressentir une répulsion physique.
Surprise	Basse Intensité : C'est comme être momentanément pris au dépourvu par une nouvelle inattendue. Haute Intensité : C'est comme se sentir complètement stupéfié par une révélation soudaine et étonnante.
Confiance en soi	Basse Intensité : C'est comme se sentir discrètement confiant dans ses capacités. Haute Intensité : C'est comme rayonner une aura d'assurance absolue.
Confusion	Basse Intensité : C'est comme essayer de comprendre un concept complexe. Haute Intensité : C'est comme être complètement perdu et désorienté.
Sympathie	Basse Intensité : C'est comme montrer de la gentillesse et du soutien. Haute Intensité : C'est comme partager pleinement la joie ou la peine de quelqu'un.
Mépris	Basse Intensité : C'est comme ressentir un sentiment subtil de supériorité. Haute Intensité : C'est comme ressentir un mépris fort envers quelqu'un ou quelque chose.

Chaque phrase était prononcée deux fois de suite pour chaque intensité, afin de conserver la seconde production. Cette deuxième production permettait à l'acteur de prendre le temps de s'approprier l'émotion selon la phrase demandée. Ainsi, toutes les phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020) étaient prononcées 4 fois par émotion, ce qui correspond à 410 enregistrements récoltés par acteur (10 phrases x 10 émotions x 2 intensités x 2 répétitions + 10 phrases neutres) et 4920 enregistrements pour l'ensemble des participants.

Enfin, des exemples de mises en situation étaient mis à disposition pour les acteurs s'ils le souhaitaient avant de débiter l'interprétation de chaque émotion (cf. **Annexe 1**). Celles-ci permettaient de les aider s'ils n'avaient pas suffisamment d'inspiration pour interpréter une émotion en particulier. Les exemples n'étaient donc pas donnés systématiquement aux participants, mais uniquement s'ils le demandaient.

4.4.3 Description de la partie de la base de données analysée dans ce mémoire

Pour rappel, de notre échantillon initial composé de 12 participants, nous avons pris la décision de réaliser notre mémoire sur un plus petit groupe de sujets, à savoir, 4 acteurs (2 hommes ; moyenne d'âge = 43,5 ans et 2 femmes ; moyenne d'âge = 29,5 ans) (cf. **Tableau 2**). Ensuite, des 10 émotions interprétées par les acteurs lors des testings, nous avons choisi de centrer notre étude sur les 6 émotions universelles de Paul Ekman (la colère, le dégoût, la peur, la joie, la surprise et la tristesse), et ce, pour le grand nombre de données issues de la littérature dont nous disposons au sujet de leur impact sur les paramètres de la voix et de la parole. Nous avons également souhaité focaliser notre récolte de données et nos analyses sur les énoncés produits à forte intensité, pour que l'évolution des paramètres acoustiques étudiés soit davantage contrastée. Puis, nous avons sélectionné uniquement les énoncés produits lors du second essai, afin que l'acteur ait eu le temps de se mettre en situation et d'interpréter au mieux l'émotion. Finalement, cela signifie que 280 énoncés (4 acteurs x 7 états émotionnels x 10 phrases) ont été analysés afin de tirer des conclusions sur nos questions de recherche.

4.4.4 Traitements des enregistrements

Une fois les testings terminés, nous avons procédé au traitement des enregistrements. Durant cette étape, afin que les expérimentateurs soient dans de bonnes conditions d'écoute, ceux-ci disposaient du même casque (K 271 MK II, Closed-back studio headphones).

Pour réaliser l'analyse acoustique et le découpage des audios, le logiciel PRAAT (version 6.4.07 – March 17, 2024) a été utilisé. La méthodologie appliquée pour l'obtention des données acoustiques est détaillée en annexe (cf. **Annexe 3**). PRAAT (6.4.07) nous a permis d'obtenir, pour chaque état émotionnel, les différentes valeurs correspondant aux paramètres analysés dans ce mémoire. Ces derniers sont détaillés ci-après.

Analyse acoustique de la fréquence fondamentale (Hz)

Pour l'étude de la médiane de la fréquence fondamentale, le logiciel PRAAT (6.4.07) nous fournissait sa valeur. Afin d'étudier la variation de la fréquence fondamentale, PRAAT (6.4.07) nous a fourni les valeurs de f_0 maximale et minimale au cours de chaque audio. Ensuite, pour obtenir la valeur de l'étendue de f_0 , nous avons réalisé la différence entre ces deux valeurs pour chaque échantillon audio.

Analyse acoustique de l'intensité (dB)

Afin d'étudier l'intensité, le logiciel PRAAT (6.4.07), nous a procuré les valeurs de sa médiane, ainsi que son intensité maximale et minimale au cours de chaque énoncé. Ensuite, pour obtenir la valeur de l'entendue, nous avons effectué la différence entre ces deux valeurs (minimum et maximum) pour chaque audio.

Analyse acoustique du LTAS (dB)

En ce qui concerne l'étude du LTAS, le logiciel PRAAT (6.4.07) nous a permis d'obtenir les valeurs du rapport L1-L0 et du pourcentage de non-voisement.

Analyse acoustique du CPPs (dB)

Le logiciel PRAAT (6.4.07) nous a permis d'obtenir la valeur de ce paramètre acoustique pour chaque énoncé.

Analyse du débit de parole (nombre de syllabes/seconde)

Afin de calculer ce paramètre via PRAAT (6.4.07), nous avons calculé le nombre de syllabes composant chaque échantillon vocal, ainsi que la durée totale de l'audio correspondant. Nous avons ensuite divisé le nombre de syllabes par la durée (secondes) de l'enregistrement.

Analyse de la vitesse articulatoire (nombre de syllabes/seconde sans les pauses)

Afin de calculer la vitesse articulatoire, nous avons réalisé la même procédure que celle du débit. Néanmoins, de cette formule, nous avons supprimé les pauses supérieures ou égales à une durée de 250 millisecondes (Grosjean & Deschamps, 1975, cités par Grosman & al., 2018) du temps total de l'audio, avant d'effectuer la division.

Enfin, pour enregistrer chaque fichier audio, une nomenclature a été respectée afin de classer les données de manière cohérente et les interpréter par la suite (**cf. Annexe 4**).

4.4.5 Analyse des paramètres acoustiques de la voix et de la parole des 4 participants

Les données numériques ainsi récoltées ont été rassemblées dans plusieurs tableaux Excel. Les moyennes des données obtenues se trouvent en annexe (**cf. Annexe 6**).

Afin d'analyser et d'interpréter les données, nous avons eu recours aux deltas (symbolisés par Δ ou δ). L'utilisation de cette méthode s'inspire d'un modèle statistique non-paramétrique développé par Spearman et Holzinger en 1924. Celui-ci est souvent utilisé en statistiques pour

estimer la variance de variables dont les distributions sont non-linéaires, même sur de petits échantillons (Bera & Koley, 2023 ; Remacle & al, 2018).

Dans le cadre de notre mémoire, ce procédé d'analyse a été adapté. En effet, nous avons effectué les différences entre les valeurs des paramètres acoustiques étudiés dans une condition émotionnelle et la condition neutre. Cette méthode nous a permis d'estimer la variabilité des paramètres acoustiques de la voix et de la parole en fonction de chaque émotion interprétée par rapport à la condition neutre et ce, pour chacun de nos acteurs. Ainsi, nous avons pu tirer des conclusions sur la variabilité de ces paramètres en intra-individuel (variabilité des données acoustiques en fonction de chaque émotion exprimée par le même acteur), mais aussi en inter-individuel (variabilité de ces données entre les différents acteurs).

La procédure delta qui a été appliquée, est décrite ci-après.

1. En premier lieu, nous avons effectué, pour chaque acteur, le calcul de la moyenne de chaque paramètre acoustique obtenu au sein des phrases, et ce pour chaque état émotionnel en haute intensité.

Exemple¹⁶ : Je souhaite déterminer la moyenne des données obtenues sur l'ensemble des 10 énoncés pour la médiane de la fréquence fondamentale de l'acteur 02, lorsque celui-ci exprimait l'émotion de la joie. Mon calcul est le suivant :

$$\bar{x}_{fo\ médiane\ acteur\ 02_toutes\ les\ phrases_joie_Intensité\ 2_Tentative\ 2}$$

$$= \frac{\sum(fo\ médiane\ acteur02_Phrase1_Intensité2_Tentative2 + fo\ médiane\ acteur02_Phrase2 \dots)}{10}$$

2. Ensuite, nous avons procédé de la même manière pour obtenir la moyenne de chaque paramètre acoustique obtenu au sein des phrases selon l'état neutre (absence d'émotion)

Exemple : Je souhaite effectuer la moyenne des données obtenues sur l'ensemble des 10 énoncés pour la médiane de la fréquence fondamentale de l'acteur 02, lorsque celui-ci n'exprimait aucune émotion. Mon calcul est le suivant :

$$\bar{x}_{fo\ médiane\ acteur\ 02_toutes\ les\ phrases_neutre}$$

$$= \frac{\sum(fo\ médiane\ acteur02_Phrase1_neutre + fo\ médiane\ acteur02_Phrase2 \dots)}{10}$$

¹⁶ Σ = somme
 \bar{x} = moyenne

3. Puis, nous avons réalisé notre opération delta en appliquant une opération de soustraction entre la valeur moyenne résultant des étapes précédentes, pour chaque paramètre acoustique (associé à une émotion) et leur valeur moyenne dans la condition neutre.

Exemple : Je souhaite déterminer le delta des données obtenues pour la médiane de la fréquence fondamentale de l'acteur 02, pour objectiver la variabilité de ce paramètre lorsque celui-ci exprimait l'émotion de la joie, par rapport à la condition neutre. Mon calcul est le suivant :

$$\Delta f_{o \text{ médiane } \text{acteur 02_toutes les phrases_joie_Intensité 2_Tentative 2}}$$

$$= \bar{x} f_{o \text{ médiane } \text{acteur 02_toutes les phrases_joie_Intensité 2_Tentative 2}} - \bar{x} f_{o \text{ médiane } \text{acteur 02_toutes les phrases_neutre}}$$

Ces opérations de delta ont été réalisées pour chaque acteur, paramètre et émotion.

4. Par la suite, afin de tirer des conclusions sur l'évolution du paramètre acoustique étudié selon chaque émotion pour l'ensemble de notre échantillon, nous avons calculé une moyenne des données deltas obtenues pour chaque acteur.

Exemple : Je souhaite effectuer le delta moyen des données obtenues pour tous les acteurs de l'échantillon, pour la médiane de la fréquence fondamentale dans le but d'objectiver la variabilité de ce paramètre lorsque l'émotion de la joie était produite, par rapport à la condition émotionnelle neutre. Mon calcul est le suivant :

$$\bar{x} \Delta f_{o \text{ médiane } \text{acteur 01-02 -07-08_toutes les phrases_joie_Intensité 2_Tentative 2}}$$

$$= \frac{\sum(\Delta f_{o \text{ médiane } \text{acteur 01_toutes les phrases_joie_Intensité2_Tentative2}} + \Delta \text{acteur 02 ...})}{4}$$

Enfin, tous les résultats ont été reportés sur des graphiques (**cf. Annexe 7**) et nous ont permis d'analyser la variabilité des paramètres acoustiques étudiés en intra et inter-acteur. Nous avons tenté d'interpréter ces résultats et de les comparer avec ce qui est avancé dans la littérature à leur sujet. L'ensemble des résultats seront décrits avec précision dans les prochaines sections de ce travail.

5. RÉSULTATS

Les tableaux suivants reprennent les données deltas obtenues pour chaque paramètre acoustique étudié lorsque les 6 émotions d'Ekman étaient produites par les 4 acteurs de notre échantillon.

Ainsi, nous parlerons de « delta » (Δ), pour faire référence à la valeur correspondant à la différence entre les moyennes calculées pour chaque paramètre acoustique, lors de la condition neutre et émotionnelle. Ces valeurs apparaissent dans les colonnes portant le numéro des acteurs (*F01, H02, F07, H08*).

Nous utiliserons les expressions « valeurs moyennes des deltas » ou « delta moyen » pour interpréter les moyennes générales effectuées sur les deltas obtenus par chaque participant et pour chaque émotion étudiée, ainsi que leur évolution positive (Δ^+) ou négative (Δ^-) par rapport à la condition neutre. Celles-ci se présentent dans les colonnes intitulées « $M\Delta$ » au sein de chaque tableau.

L'ensemble de ces résultats seront interprétés en inter et intra-individuel. Les moyennes brutes obtenues pour chaque acteur et chaque condition émotionnelle sont reprises dans les annexes. Il en est de même pour leurs graphiques descriptifs (**cf. Annexe 6 ; Annexe 7**).

5.1 Résultats obtenus sur l'évolution des paramètres de la voix avec les émotions d'Ekman

5.1.1 La fréquence fondamentale

Afin d'étudier l'évolution de f_0 avec les émotions « de base » d'Ekman, nous nous sommes intéressées à sa médiane et son étendue. Les résultats obtenus pour ces 2 paramètres sont repris dans les tableaux ci-après (**cf. Tableau 6 ; Tableau 7**).

Tableau 6. Analyse delta de la médiane moyenne de la fréquence fondamentale lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ médiane de f_0 (Hz)					
	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>	M Δ	
Δ Peur	128,88	174,2	140,47	76,36	129,98	$\Delta +$
Δ Colère	51,53	143,65	103,38	105,31	100,97	$\Delta +$
Δ Joie	96,43	196,31	141,64	113,75	137,03	$\Delta +$
Δ Tristesse	77,46	182,15	48,72	82,30	97,66	$\Delta +$
Δ Dégoût	8,15	113,07	29,65	74,98	56,46	$\Delta +$
Δ Surprise	123,74	148	69,32	80,33	105,35	$\Delta +$

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; $\Delta +$ = delta positif

Synthèse des résultats obtenus pour f_0 : deltas des médianes moyennes

Interprétations inter-individuelles

Dans ce premier tableau (cf. **Tableau 6**), nous pouvons observer que chaque émotion a été produite, en moyenne, avec une évolution positive ($\Delta+$) du delta de la médiane de la fréquence fondamentale, par rapport à la condition neutre. En d'autres termes, cela signifie que tous les acteurs de l'échantillon ont tendance à produire, en moyenne, une valeur de la médiane de la fréquence fondamentale plus importante lorsqu'ils interprètent une émotion, que lorsqu'ils n'en produisent aucune. On peut remarquer que le dégoût est l'émotion qui montre la différence la moins importante par rapport à la condition neutre chez les acteurs 01, 02 et 08. Ainsi, ceux-ci interprètent le dégoût avec une valeur moyenne de f_0 plus faible que celle obtenue pour les autres émotions. Enfin, nous pouvons noter que les valeurs deltas des moyennes obtenues pour ce paramètre chez l'actrice 07 sont supérieures à celles des autres acteurs pour toutes les émotions étudiées.

Interprétations intra-individuelles

En ce qui concerne les données intra-individuelles, nous pouvons tout d'abord relever que pour chaque acteur, le dégoût est l'émotion qui a été produite avec une valeur de delta de médiane de la fréquence fondamentale plus faible, par rapport aux autres émotions.

L'actrice 01 semble avoir produit la peur et la surprise avec des valeurs de deltas supérieures aux autres émotions. Les acteurs 07 et 08 ont, quant à eux, interprété la joie avec des valeurs de deltas supérieures par rapport aux autres émotions. L'acteur 02 a obtenu des valeurs de deltas comparables pour la peur et la joie.

Tableau 7. Analyse delta de l'étendue moyenne de la fréquence fondamentale lors de l'interprétation des 6 émotions d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ étendue de f_0 (Hz) $f_{0 \max} - f_{0 \min}$					
	F01	F07	H02	H08	M Δ	
Δ Peur	6,97	100,63	119,62	107,46	83,67	$\Delta +$
Δ Colère	26,17	143,15	78,13	167,49	103,74	$\Delta +$
Δ Joie	80,63	89,43	148,92	126,73	111,43	$\Delta +$
Δ Tristesse	5,14	208,89	69,85	141,60	106,37	$\Delta +$
Δ Dégoût	23,89	129,93	26,05	100,43	70,08	$\Delta +$
Δ Surprise	105,58	232,07	99,73	192,60	157,50	$\Delta +$

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; $\Delta +$ = delta positif

Synthèse des résultats obtenus pour f_0 : deltas des étendues moyennes

Interprétations inter-individuelles

Dans ce second tableau (cf. **Tableau 7**), nous pouvons observer que, pour l'ensemble des émotions, les valeurs moyennes des deltas obtenus sont positives ($\Delta +$) pour l'étendue de la fréquence fondamentale par rapport à la condition neutre. Cela signifie qu'en moyenne, il y a plus de variations de la valeur de l'étendue de la fréquence fondamentale lorsque les acteurs interprètent une émotion, que lorsqu'ils n'en produisent aucune. Néanmoins, même si cette tendance se marque pour chaque émotion, elle ne se manifeste pas de la même façon pour chaque acteur. En effet, on peut par exemple relever que les deltas des moyennes obtenues par l'actrice 01 sont moins importants pour ce paramètre par rapport aux autres acteurs.

Interprétations intra-individuelles

Au niveau intra-individuel, nous pouvons observer que, par rapport à la condition neutre, l'actrice 01 a produit la tristesse et la peur avec, en moyenne, le moins de variations intonatives. Pour l'actrice 07, l'émotion qui a été sujette à de moindres variations fréquentielles par rapport à la condition sans émotion est la joie. Pour les acteurs 02 et 08, il s'agit du dégoût.

La surprise, quant à elle, est l'émotion qui a été produite, en moyenne, avec la plus grande variabilité de l'étendue de f_0 , par l'actrice 01, l'actrice 07 et l'acteur 08. Ainsi, les valeurs deltas sont supérieures pour cette émotion chez ces 3 acteurs. L'acteur 02, quant à lui, a produit en moyenne le plus de variations intonatives pour interpréter la joie par rapport à l'état neutre.

5.1.2 L'intensité

Nous nous sommes intéressées à la médiane et l'étendue pour l'étude de l'évolution de l'intensité. Les résultats sont décrits ci-après (cf. **Tableau 8** ; **Tableau 9**).

Tableau 8. Analyse delta de la médiane moyenne de l'intensité lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ médiane de l'intensité (dB)					
	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>	M Δ	
Δ Peur	1,78	-0,15	-0,33	0,44	0,44	Δ +
Δ Colère	0,52	-0,72	0,89	0,51	0,30	Δ +
Δ Joie	0,17	-1,28	-0,16	1,71	0,11	Δ +
Δ Tristesse	0,00	-4,01	-3,65	-1,60	-2,31	Δ -
Δ Dégoût	-0,54	-2,08	1,26	-0,59	-0,48	Δ -
Δ Surprise	0,24	-1,38	2,15	1,12	0,54	Δ +

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; Δ + = delta positif ; Δ - = delta négatif

Synthèse des résultats obtenus pour l'intensité : deltas des médianes moyennes

Interprétations inter-individuelles

En ce qui concerne l'étude de la médiane de l'intensité, les données récoltées dans le tableau varient (**cf. Tableau 8**). En effet, pour ce paramètre, les deltas moyens sont positifs ($\Delta+$) entre la condition neutre et la production de la peur, de la colère, de la joie et de la surprise. Cela signifie que la valeur de la tendance centrale de l'intensité est, en moyenne, plus importante lorsque ces émotions sont produites, par rapport à la condition sans émotion. En revanche, pour l'interprétation de la tristesse et du dégoût, on peut relever des deltas moyens négatifs ($\Delta-$). Cela signifie que la valeur de la tendance centrale de l'intensité semble, en moyenne, plus faible lorsque la tristesse et le dégoût sont interprétés, par rapport à la condition non-émotionnelle.

Toutefois, la tristesse semble davantage se distinguer des autres émotions pour la valeur médiane de l'intensité, car la moyenne du delta qui a été obtenue pour ce paramètre est la plus faible ($\Delta_{\text{médiane moyen}} = -2,31$) par rapport à la condition neutre. Ainsi, cette donnée suggère qu'en moyenne, la tristesse est l'émotion qui a été produite avec les valeurs d'intensité les plus faibles, et inférieures à celles de la condition neutre.

Pour l'ensemble de ces données, il est important de prendre du recul, car, pour rappel, il s'agit de moyennes effectuées sur les données obtenues par tous les acteurs. Néanmoins, si on évalue plus précisément les résultats obtenus par chacun d'eux, nous remarquons que les valeurs positives de deltas ($\Delta+$) ne se marquent pas chez tous les participants et pour toutes les émotions.

De plus, il est intéressant de noter que l'acteur 07 se démarque de façon générale des autres participants par des valeurs deltas de médiane de l'intensité particulièrement faibles.

Interprétations intra-individuelles

Au niveau des résultats intra-individuels, nous remarquons que pour les acteurs 07,02 et 08, les deltas obtenus pour la médiane de l'intensité sont les plus faibles avec l'interprétation de la tristesse. Pour l'actrice 01, cela correspond au dégoût.

La peur est l'émotion qui a été produite avec les deltas les plus élevés chez les 2 femmes de l'échantillon (actrices 01 et 07). L'acteur 02 montre un delta supérieur pour la surprise. Pour l'acteur 08, cette tendance se marque pour la joie.

Tableau 9. Analyse delta de l'étendue moyenne de l'intensité lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ étendue de l'intensité (dB)					
	<i>Intensité_{max} - Intensité_{min}</i>					
	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>	<i>M Δ</i>	
Δ Peur	5,14	9,47	2,22	10,67	6,87	$\Delta +$
Δ Colère	6,40	13,75	5,33	12,71	9,55	$\Delta +$
Δ Joie	5,46	15,65	7,17	15,24	10,88	$\Delta +$
Δ Tristesse	6,03	12,72	-2,61	12,03	7,04	$\Delta +$
Δ Dégoût	3,40	11,31	0,70	12,00	6,85	$\Delta +$
Δ Surprise	7,10	15,13	3,40	10,20	8,96	$\Delta +$

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; $\Delta +$ = delta positif

Synthèse des résultats obtenus pour l'intensité : deltas des étendues moyennes

Interprétations inter-individuelles

En analysant ce tableau (cf. **Tableau 9**), nous pouvons observer que les deltas moyens obtenus entre les conditions neutres et émotionnelles pour l'étendue de l'intensité sont positifs pour chaque émotion ($\Delta+$). Cela signifie que les acteurs interprètent, en moyenne, pour toutes les émotions étudiées, des énoncés avec de plus grandes variations de volume sonore, par rapport à la condition neutre. Néanmoins, il est important de noter que l'acteur 02 est celui qui a interprété la tristesse en produisant, en moyenne, le moins de variations de niveau de pression sonore, par rapport à la condition neutre. L'actrice 07, quant à elle, est celle qui a produit en moyenne le plus de variations du niveau pression sonore, par rapport à la condition neutre pour interpréter la colère, la joie, la tristesse et la surprise.

Interprétations intra-individuelles

En ce qui concerne les résultats intra-individuels, nous pouvons relever que les deltas obtenus pour les variations intonatives diffèrent fortement d'un acteur à l'autre.

En effet, l'actrice 01 a interprété le dégoût avec, en moyenne, le moins de variations d'intensité, par rapport à la condition neutre. Pour l'actrice 07, il s'agit de la peur. Les valeurs de delta,

obtenues par l'acteur 08, sont les plus faibles pour la peur et la surprise. Pour l'acteur 02, il s'agit de la tristesse. Celle-ci a été produite avec un delta négatif (Δ^-), suggérant que cet acteur produisait moins de variations d'intensité pour produire la tristesse que lors de la condition neutre.

L'émotion qui a été, en moyenne, sujette aux plus grandes valeurs de deltas pour l'étendue de l'intensité, est la surprise pour l'actrice 01. Pour l'actrice 07, ce sont la joie et la surprise qui ont toutes deux des résultats de deltas comparables. Pour les acteurs 02 et 08, il s'agit de la joie.

5.1.3 Long-Term Average Spectrum (LTAS)

Nous avons choisi d'analyser le rapport L1-L0 ainsi que le pourcentage de non-voisement pour étudier l'évolution du LTAS. Les résultats sont décrits dans les tableaux suivants (**cf. Tableau 10 ; Tableau 11**).

Tableau 10. *Analyse delta du rapport L1-L0 moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).*

	Δ rapport L1-L0 (dB)						
	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>	<i>M Δ</i>		
Δ Peur	5,40	13,22	5,65	9,39	8,42	$\Delta +$	
Δ Colère	8,09	13,14	8,02	13,88	10,78	$\Delta +$	
Δ Joie	6,18	13,28	8,51	13,81	10,44	$\Delta +$	
Δ Tristesse	9,57	18,03	3,07	10,15	10,21	$\Delta +$	
Δ Dégoût	3,92	14,22	4,04	12,25	8,61	$\Delta +$	
Δ Surprise	4,44	9,72	4,09	10,57	7,21	$\Delta +$	

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; $\Delta +$ = delta positif

Synthèse des résultats obtenus pour le LTAS : deltas des rapports L1-L0 (dB) moyens

Interprétations inter-individuelles

En observant ce tableau (**cf. Tableau 10**), nous pouvons conclure que les deltas moyens obtenus pour le rapport L1-L0 du LTAS entre les conditions neutres et émotionnelles sont positifs ($\Delta +$), pour toutes les émotions produites et pour tous les acteurs. Cela met en évidence que tous les

participants ont, en moyenne, interprété les 6 émotions avec une voix plus puissante et moins soufflée que lors de la condition neutre.

Interprétations intra-individuelles

Au niveau intra-individuel, nous pouvons observer que les valeurs de deltas obtenues pour les rapports L1-L0 sont assez similaires chez chaque acteur, mais ne se distribuent pas de manière identique en fonction des émotions produites. En effet, pour l'actrice 01, la valeur delta obtenue est plus faible pour le dégoût. Pour l'actrice 07, il s'agit de la surprise. En ce qui concerne l'acteur 02, c'est la tristesse qui a été produite avec un delta plus faible en comparaison aux autres émotions. L'acteur 08, quant à lui, montre cette tendance pour la peur. Cela signifie que toutes ces émotions ont été interprétées par les acteurs avec le plus de souffle et le moins de puissance, en comparaison aux autres émotions étudiées.

Nous pouvons également observer que les deux femmes de l'échantillon (actrices 01 et 07) ont produit la tristesse avec les plus hautes valeurs de deltas pour le rapport L1-L0. Pour les 2 hommes (acteur 02 et 08), les valeurs de deltas les plus importantes s'observent avec la colère et la joie. Ces émotions étaient alors produites avec plus de puissance et étaient moins soufflées par rapport à la condition neutre, en comparaison aux autres émotions.

Tableau 11. Analyse delta du pourcentage de non-voisement moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ pourcentage de non-voisement (%)					
	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>	M Δ	
Δ Peur	-2,01	-5,12	-10,59	-6,01	-5,93	Δ -
Δ Colère	-4,67	-0,73	-5,91	-5,04	-4,09	Δ -
Δ Joie	-2,82	2,30	-8,97	-6,47	-3,99	Δ -
Δ Tristesse	-2,74	9,93	9,85	2,12	4,79	Δ +
Δ Dégoût	3,13	2,83	-2,42	-1,18	0,59	Δ +
Δ Surprise	4,99	5,66	-4,78	-3,11	0,69	Δ +

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; Δ + = delta positif ; Δ - = delta négatif

Synthèse des résultats obtenus pour le LTAS : *deltas des pourcentages de non-voisement moyens*

Interprétations inter-individuelles

Les données obtenues pour le pourcentage de non-voisement (**cf. Tableau 11**) mettent en évidence des deltas moyens positifs ($\Delta +$) pour les émotions de la tristesse, du dégoût et de la surprise. Cela signifie que ces émotions ont été interprétées, en moyenne, avec davantage de souffle que l'état neutre. Le delta moyen du pourcentage de non-voisement par rapport à la condition neutre est particulièrement important avec l'interprétation de la tristesse chez les acteurs 07,02 et 08. Cela signifie que ceux-ci interprétaient cette émotion avec plus de souffle que l'acteur 01. Pour interpréter la peur, la joie et la colère, la valeur moyenne de delta obtenue est négative ($\Delta -$), ce qui met en évidence que la voix des acteurs de notre échantillon était plus puissante et moins soufflée pour produire ces deux émotions, par rapport à leur voix en condition neutre.

Interprétations intra-individuelles

La peur est l'émotion qui a été interprétée avec les valeurs deltas les plus faibles pour l'actrice 07 et l'acteur 02. L'actrice 01 montre une valeur delta plus faible pour la colère, alors que l'acteur 08 montre cette tendance pour la joie et la peur.

Pour les acteurs 07,02 et 08, la tristesse est l'émotion qui a été produite avec les valeurs deltas les plus importantes par rapport aux autres émotions. Pour l'actrice 01, il s'agit de la surprise.

5.1.4 Smoothed Cepstral Peak Prominence (CPPs)

Les résultats obtenus pour ce paramètre via PRAAT (6.4.07) (**cf. Annexe 3**) sont repris dans le tableau ci-dessous (**cf. Tableau 12**).

Tableau 12. Analyse delta du CPPs moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ CPPs (dB)					
	<i>F01</i>	<i>F 07</i>	<i>H 02</i>	<i>H 08</i>	M Δ	
Δ Peur	-0,58	-1,39	-1,36	-0,13	-0,86	Δ -
Δ Colère	1,14	-0,97	-0,99	-0,37	-0,30	Δ -
Δ Joie	-0,11	-1,67	-1,12	0,07	-0,71	Δ -
Δ Tristesse	0,83	-1,57	-2,42	-0,07	-0,81	Δ -
Δ Dégoût	0,27	-0,58	-0,58	0,02	-0,22	Δ -
Δ Surprise	-0,82	-1,78	-0,88	-0,58	-1,01	Δ -

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; Δ- = delta négatif

Synthèse des résultats obtenus pour le CPPs

Interprétations inter-individuelles

Les résultats obtenus dans ce tableau (**cf. Tableau 12**) soulèvent, que pour toutes les émotions produites, le signal vocal semble en moyenne moins périodique (Δ-) par rapport à la condition neutre. Cela signifie que la voix des acteurs est, en moyenne, moins stable et régulière lorsqu'ils produisent des émotions, par rapport à la condition sans émotion. L'actrice 01 semble cependant montrer plus de stabilité et de régularité par rapport aux 3 autres acteurs de l'échantillon pour la colère, la tristesse et le dégoût.

Il est intéressant de noter que la surprise semble être l'émotion dont la valeur moyenne de CPPs est la plus faible pour tous les acteurs, ce qui suggère que ceux-ci l'ont produite de manière moins stable et régulière, par rapport aux autres émotions. Néanmoins, les valeurs obtenues pour ce paramètre diffèrent très légèrement entre elles. Cette interprétation doit donc être traitée avec beaucoup de prudence.

Interprétations intra-individuelles

Chez les acteurs 01, 07 et 08, la surprise a été interprétée avec un delta négatif (Δ-) plus faible, en regard des autres émotions. Pour l'acteur 02, il s'agit de la tristesse. Cela signifie que ces émotions ont été produites avec moins de stabilité et de régularité dans le signal vocal que dans la condition neutre.

L'actrice 01 montre la valeur delta la plus importante pour interpréter la colère, par rapport aux autres émotions. Pour les acteurs 07 et 02, c'est le dégoût. En ce qui concerne l'acteur 08, les valeurs deltas les plus élevées sont relevées pour la joie et le dégoût. Cela suggère que ces émotions ont été produites avec le plus de stabilité et de régularité dans le signal vocal, par rapport à la condition neutre.

5.2 Résultats obtenus sur l'évolution des paramètres de la fluence avec les émotions

5.2.1 Le débit

Les résultats obtenus pour ce paramètre sont détaillés dans le tableau ci-après (cf. **Tableau 13**).

Tableau 13. Analyse delta du débit moyen lors de l'interprétation des 6 émotions universelles d'Ekman par rapport à la condition neutre pour les 4 acteurs (2 femmes et 2 hommes).

	Δ débit (nombre de syllabes/seconde)					
	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>	M Δ	
Δ Peur	-0,72	-0,36	-0,06	0,53	-0,15	Δ -
Δ Colère	-0,99	-0,88	-0,97	0,39	-0,61	Δ -
Δ Joie	-1,52	-0,64	-1,02	0,32	-0,72	Δ -
Δ Tristesse	-1,75	-1,58	-0,96	-0,24	-1,13	Δ -
Δ Dégoût	-2,33	-1,11	-0,67	-0,19	-1,07	Δ -
Δ Surprise	-1,22	-1,10	-0,28	0,22	-0,59	Δ -

Note. F= femme ; H= homme ; M = Moyenne ; M Δ = Moyenne des deltas ; Δ- = delta négatif

Synthèse des résultats obtenus pour le débit

Interprétations inter-individuelles

Ces données mettent en évidence que, pour toutes les émotions, le débit produit par l'ensemble des acteurs est en moyenne plus lent (Δ-) par rapport à la situation dans laquelle aucune émotion n'est produite (cf. **Tableau 13**). Seul l'acteur 08 montre des deltas positifs (Δ+) pour la peur, la joie et la colère, suggérant qu'il est le seul à avoir interprété ces émotions avec un débit moyen plus rapide que lors de la condition neutre.

Interprétations intra-individuelles

Les 4 acteurs ont interprété la peur avec une valeur delta supérieure aux autres émotions produites. Cela signifie que, pour chaque acteur, la peur a été produite avec un débit de parole plus rapide par rapport aux 5 autres émotions. Néanmoins, celui-ci reste plus lent par rapport à la condition neutre.

En ce qui concerne les acteurs 07 et 08, c'est la tristesse qui a été produite avec un delta plus bas. L'actrice 01, quant à elle, montre un delta inférieur pour le dégoût par rapport aux autres émotions. Finalement, on peut relever une valeur de delta plus faible chez l'acteur 02 pour la colère, la joie et la tristesse.

5.2.2 La vitesse articulatoire

Pour rappel, la vitesse articulatoire se calcule en réalisant le rapport entre le nombre de segments (ou syllabes) produits au cours d'un énoncé et la durée totale de celui-ci, en supprimant les pauses (d'une durée supérieure ou égale à 250 millisecondes) (Grosjean & Deschamps, 1975, cités par Grosman & al., 2018 ; Kühnert & Kocjančič Antolík, 2016).

Néanmoins, pour ce paramètre, il est essentiel de noter que la méthodologie des testings réalisés n'autorisait pas les acteurs à produire de réelles pauses lorsque ceux-ci produisaient les émotions. Les résultats obtenus pour ce paramètre sont donc identiques aux résultats obtenus dans le tableau correspondant au débit de parole. Ainsi, nous ne présenterons pas de tableau de résultats pour ce paramètre de la fluence et ses données seront interprétées avec énormément de prudence.

Dans la prochaine section, nous nous intéresserons aux variations des paramètres de voix et de fluence propres à chaque émotion étudiée. Ces dernières données résument qualitativement les valeurs obtenues pour les deltas, reprises dans les tableaux précédents.

5.3 Identification des paramètres de voix et de fluence spécifiques à chaque émotion

La peur

Nous pouvons observer qu'en moyenne, la peur a été interprétée avec une fréquence fondamentale médiane plus élevée et avec plus de variations de celle-ci, en comparaison à l'état neutre pour chaque participant. Au niveau de l'intensité, on peut relever qu'en moyenne, les acteurs ont produit une intensité médiane plus élevée ainsi que plus des variations intonatives en regard du neutre. Il faut noter que seuls 2 acteurs (01 et 08) ont montré cette tendance concernant la médiane de l'intensité. La peur a également été produite avec une voix puissante

et peu soufflée chez tous les acteurs, en comparaison à la condition neutre. Finalement, cette émotion a été interprétée avec plus d'irrégularités dans la voix, en comparaison à la condition contrôle.

En ce qui concerne la fluence, la peur a été produite avec un débit en moyenne plus lent, par rapport à la condition sans émotion.

La colère

En moyenne, la colère a été interprétée avec une valeur médiane de f_0 plus haute et plus de variations fréquentielles, par rapport aux conditions neutres chez chaque acteur de notre échantillon. Cette émotion a également été produite avec, en moyenne, une valeur d'intensité médiane supérieure et plus de variations d'intensité, en comparaison à la condition contrôle, chez la plupart des acteurs. Plus précisément, la tendance centrale de l'intensité a en effet été plus élevée chez 3 des 4 acteurs (01, 02 et 08). Selon les données obtenues pour le rapport L1-L01 et le pourcentage de non-voisement, nous pouvons remarquer que la colère a été produite avec une voix puissante et peu soufflée chez chaque participant, en comparaison à la condition sans émotion. La colère a finalement été interprétée avec plus d'irrégularités dans la voix en comparaison à la condition contrôle.

Au niveau de la fluence, la colère a été produite avec un débit en moyenne plus lent par rapport à la condition sans émotion.

La joie

Nous pouvons observer que la joie a été produite avec une fréquence fondamentale médiane plus élevée dans cette condition par rapport à la condition neutre. De plus, il y a eu plus de variations fréquentielles en comparaison à la condition contrôle. Du point de vue de l'intensité, on peut relever qu'en moyenne, les acteurs ont produit une intensité médiane plus élevée ainsi que plus de variations intonatives en regard de l'état neutre. Il faut noter que seuls les acteurs 01 et 08 montrent une tendance centrale de l'intensité supérieure au neutre. En ce qui concerne les données obtenues pour le rapport L1-L0 et le pourcentage de non-voisement, par rapport à la condition neutre, nous pouvons remarquer que la joie est marquée par une voix puissante et peu soufflée chez la plupart des participants. L'actrice 07 se démarque en effet des autres acteurs par un pourcentage de non-voisement plus élevé que dans la condition sans émotion. La joie a finalement été interprétée avec plus d'irrégularités dans la voix que dans la condition neutre.

En ce qui concerne le débit, nous pouvons observer que la joie a été produite avec un débit moyen plus lent, en comparaison à l'état neutre.

La tristesse

La tristesse a été interprétée, en moyenne, avec une valeur médiane de la fréquence fondamentale plus haute et avec davantage de variations de fréquence par rapport aux conditions neutres, chez chaque acteur de notre échantillon. Pour le paramètre d'intensité sonore, la tristesse a été produite avec, en moyenne, une valeur d'intensité médiane inférieure, mais plus de variations d'intensité, en comparaison au neutre. Par rapport à la condition sans émotion, nous pouvons observer que la tristesse a été produite en moyenne avec une voix davantage irrégulière et plus soufflée.

Finalement, la tristesse a été produite avec un débit moyen plus lent en comparaison à la condition neutre.

Le dégoût

Nous pouvons remarquer que le dégoût a été interprété avec, en moyenne, une valeur médiane de la fréquence fondamentale plus haute et plus de variations fréquentielles par rapport aux conditions neutres, chez chaque acteur de notre échantillon. En moyenne, le dégoût a été interprété avec une valeur d'intensité médiane inférieure, mais plus de variations d'intensité en comparaison au neutre. Par rapport à la condition sans émotion, nous pouvons observer que le dégoût a été produit en moyenne avec une voix plus soufflée et irrégulière.

Au niveau de la fluence, nous pouvons observer que le dégoût été produit avec un débit moyen plus lent en comparaison au neutre.

La surprise

Quant à la surprise, nous pouvons observer qu'elle a été produite, en moyenne, avec une valeur médiane de la fréquence fondamentale plus haute et plus de variations, par rapport aux conditions neutres chez chaque acteur de notre échantillon. Au niveau de l'intensité, nous pouvons remarquer que la surprise a été interprétée avec, en moyenne, une valeur d'intensité médiane supérieure et plus de variations d'intensité, en comparaison au neutre. Cette émotion est également marquée en moyenne par une voix plus forte, peu soufflée et irrégulière, en regard de la condition contrôle.

Enfin, nous pouvons observer que la surprise a été interprétée avec un débit moyen plus lent en comparaison à la condition sans émotion.

Afin de synthétiser l'ensemble des données obtenues lors de notre travail, un tableau récapitulatif comparant nos résultats avec nos hypothèses initialement posées a été construit (**cf. Tableau 14**).

Tableau 14. Comparaison de nos résultats des deltas obtenus pour chaque paramètre acoustique étudié avec nos hypothèses initialement posées.

	f_0				Intensité				LTAS				CPPs		Débit		Vitesse articulaire	
	étendue (Hz)		médiane (Hz)		étendue (Hz)		médiane (dB)		rapport L1-L0 (dB)		non-voisement (%)		(dB)		nombre de syllabes /sec		nombre de syllabes /sec (sans pauses)	
	H	R	H	R	H	R	H	R	H	R	H	R	H	R	H	R	H	R
Peur	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta+$	/	$\Delta-$	/	$\Delta-$	$\Delta+$	$\Delta-$	$\Delta+$	$\Delta-$
Colère	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta-$	/	$\Delta-$	$\Delta+$	$\Delta-$	$\Delta+$	$\Delta-$
Joie	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta-$	/	$\Delta-$	$\Delta+$	$\Delta-$	$\Delta+$	$\Delta-$
Tristesse	$\Delta-$	$\Delta+$	$\Delta-$ ou $\Delta+$	$\Delta+$	/	$\Delta+$	$\Delta-$ ou $\Delta+$	$\Delta-$	/	$\Delta+$	/	$\Delta+$	/	$\Delta-$	/	$\Delta-$	$\Delta-$	$\Delta-$
Dégoût	/	$\Delta+$	$\Delta-$ ou $\Delta+$	$\Delta+$	/	$\Delta+$	$\Delta-$	$\Delta-$	/	$\Delta+$	/	$\Delta+$	/	$\Delta-$	$\Delta-$	$\Delta-$	/	$\Delta-$
Surprise	$\Delta+$	$\Delta+$	$\Delta+$	$\Delta+$	/	$\Delta+$	/	$\Delta+$	/	$\Delta+$	/	$\Delta+$	/	$\Delta-$	/	$\Delta-$	/	$\Delta-$

Note. H= hypothèse ; R= résultat ; / = donnée absente dans la littérature consultée ayant fait l'objet d'une hypothèse exploratoire ; $\Delta+$ = delta positif ; $\Delta-$ = delta négatif ; $\Delta-$ ou $\Delta+$ = manque de consensus dans la littérature ayant donné lieu à une hypothèse exploratoire ; **couleur verte** = résultat confirmant l'hypothèse confirmatoire initialement posée ; **couleur rouge**= résultat infirmant l'hypothèse confirmatoire initialement posée

6. DISCUSSION

6.1 Interprétation de nos résultats selon les hypothèses initialement posées pour les paramètres de la voix

6.1.1 La fréquence fondamentale (f_0) - Hz

Par rapport à ce qui était avancé dans la littérature, nous obtenons des résultats congruents avec toutes nos hypothèses, sauf pour la tristesse. En effet, pour les émotions de la peur, de la colère, de la joie et de la surprise, les deltas obtenus, sont tous positifs ($\Delta +$). Cela signifie, pour la population étudiée, que les valeurs moyennes de l'étendue et de la médiane de la fréquence fondamentale augmentent avec la production de ces émotions.

Au niveau du dégoût, les résultats obtenus suggèrent une croissance de l'étendue et de la médiane de la fréquence fondamentale ($\Delta+$) par rapport à l'état neutre. Ces informations enrichissent les données dont nous disposions initialement au sujet de cette émotion.

Néanmoins, en ce qui concerne l'étendue de la fréquence fondamentale avec la tristesse, nos résultats ne convergent pas avec ce qui était avancé dans la littérature (Banse & Scherer, 1996 ; Vos & al., 2023). En réalité, nous nous attendions à ce que l'étendue de ce paramètre soit inférieure à la condition neutre ($\Delta-$) avec cette émotion, mais tous les acteurs de l'échantillon ont produit la tristesse avec des valeurs d'étendue de f_0 supérieures à l'état neutre ($\Delta+$). À propos de l'évolution de la médiane de f_0 avec la tristesse, le manque de consensus dans la littérature ne nous permettait pas de poser une hypothèse confirmatoire à son sujet (Vos & al., 2023). Ainsi, nous avons relevé que ce paramètre évoluait positivement ($\Delta+$) avec cette émotion, en comparaison à l'état neutre. Nous pensons que ces résultats peuvent s'expliquer par la méthodologie qui a été appliquée, et notamment par le fait que les données qui ont été analysées dans ce mémoire ne concernaient que la condition à forte intensité émotionnelle. Ainsi, il aurait été intéressant d'observer comment l'ensemble des paramètres évoluaient lorsque les acteurs interprétaient les émotions à basse intensité émotionnelle, pour constater si ces résultats différaient ou non.

En outre, dans la littérature consultée, il est clairement décrit que nos états émotionnels affectent la qualité et la coordination de notre respiration et que l'altération de cette dernière peut entraîner des expériences émotionnelles particulières comme de l'anxiété, de la peur, voire des crises de panique (Ashhad & al., 2022). Ainsi, nous pensons qu'il serait pertinent de s'interroger

sur l'impact de la coordination pneumo-phonique¹⁷ sur les paramètres acoustiques lorsque les émotions sont produites. En effet, certaines émotions, telles que la surprise sont, dans la littérature, caractérisées par une voix haletante. Si la voix est considérée comme telle, nous pouvons facilement imaginer que certains paramètres acoustiques, comme l'étendue de la fréquence fondamentale, pourraient varier avec l'interprétation des émotions (Fonagy & Magdics, 1963 cités par Murray & Arnott, 1993 ; Oster & Risberg, 1986 cités par Murray & Arnott, 1993). Par exemple, qualitativement, nous avons observé lors des enregistrements, que les acteurs produisaient la peur, la joie et la surprise (à forte intensité émotionnelle) avec une voix haletante. C'est pourquoi, évaluer la coordination pneumo-phonique des acteurs lors de la production des émotions aurait été intéressant pour mieux comprendre les résultats obtenus.

De plus, nous pensons qu'il serait judicieux de se questionner sur l'influence du genre de l'individu sur l'interprétation des émotions. En effet, nous notons que les hommes de l'échantillon montrent, en moyenne, des valeurs de deltas plus importantes que celles obtenues par les femmes pour interpréter la peur et la joie, par rapport à la condition sans émotion. Ainsi, les 2 hommes interprètent ces émotions avec plus de variations intonatives que les 2 femmes par rapport à la condition neutre. Quant à l'influence du genre sur la médiane de la fréquence fondamentale, nous observerons que chez les femmes, les valeurs deltas sont en moyenne plus élevées que chez les hommes pour interpréter la surprise, en comparaison à la condition contrôle. Néanmoins, comme cela sera développé ultérieurement, il est important d'interpréter ces données avec beaucoup de prudence compte tenu de la petite taille de notre échantillon d'acteurs.

6.1.2 L'intensité - dB

En ce qui concerne la valeur de la médiane de l'intensité, tous nos résultats coïncident avec les hypothèses confirmatoires posées. En effet, la peur, la colère et la joie ont toutes été produites avec une intensité en moyenne supérieure, par rapport à la condition neutre ($\Delta+$). Le dégoût a également confirmé nos hypothèses, car la valeur moyenne de la médiane obtenue avec cette émotion était inférieure à sa valeur en l'absence d'émotion ($\Delta-$) (Banse & Scherer, 1996 ; Bänziger & al., 2015 ; Murray & Arnott, 1993 ; Sobin & Alpert, 1999). Pour la tristesse et la surprise, les résultats obtenus mettent respectivement en évidence une diminution ($\Delta-$) et une

¹⁷ Coordination entre la respiration et la phonation. Pour que la phonation ait lieu, l'adduction des plis vocaux doit être coordonnée avec le flux d'air expiratoire venant des poumons. On parle alors de coordination pneumo-phonique (Park & al., 2024).

augmentation ($\Delta+$) de la médiane de l'intensité, ce qui nous permet d'enrichir les données dont nous disposons à leur sujet dans la littérature.

Ensuite, la confrontation des données récoltées à propos de l'étendue de l'intensité avec nos hypothèses préalablement posées nous a permis de soulever que ce paramètre augmente ($\Delta+$) avec toutes les émotions étudiées. Ces résultats confirment les hypothèses que nous avons déduites pour la peur, la colère et la joie, et infirment ce qui était attendu pour le dégoût ($\Delta-$). La tristesse et la surprise, quant à elles, faisaient l'objet d'hypothèses exploratoires. Les données obtenues à leur sujet enrichissent ainsi les informations dont nous disposons sur l'évolution de l'étendue de l'intensité avec la production d'émotions.

Nous pouvons à nouveau nous questionner sur l'impact de la coordination pneumo-phonique sur l'intensité, car la littérature consultée suppose que la voix haletante impacte la stabilité de l'intensité vocale (Fonagy & Magdics, 1963 cités par Murray & Arnott, 1993 ; Oster & Risberg, 1986 cités par Murray & Arnott, 1993). Une nouvelle fois, nous avons observé que les acteurs produisaient une voix haletante pour interpréter la peur, la joie et la surprise à haute intensité émotionnelle. Obtenir des données sur la coordination pneumo-phonique permettrait d'appuyer davantage les résultats obtenus.

Si nous nous interrogeons à propos de l'influence du genre de l'acteur sur la médiane de l'intensité, les moyennes des deltas obtenus pour la colère, la joie, le dégoût et la surprise suggèrent que les femmes avaient tendance à interpréter ces émotions avec, en moyenne, moins d'intensité que les hommes, par rapport à la condition neutre. Enfin, les éléments récoltés ne nous permettent pas de tirer de conclusions quant à l'influence du genre de l'acteur sur la modification de l'étendue moyenne de l'intensité, car les résultats obtenus étaient comparables entre les hommes et les femmes pour ce paramètre.

6.1.3 Le LTAS - dB

À propos du LTAS, les résultats récoltés pour le rapport L1-L0 coïncident avec les hypothèses initialement posées pour la joie et la colère (Antonetti & al., 2020). En effet, les deltas moyens obtenus entre les conditions neutres et émotionnelles sont positifs ($\Delta+$) pour ces deux émotions. Les autres états émotionnels, qui faisaient l'objet d'hypothèses exploratoires, suivent également cette tendance. Cela signifie que l'ensemble des émotions étudiées ont été produites avec plus de puissance et moins de souffle, en comparaison à la condition neutre. Néanmoins, ces résultats doivent être interprétés avec précaution, car les données obtenues sont contradictoires avec certains résultats observés pour le pourcentage de non-voisement du LTAS.

Effectivement, les éléments obtenus pour le pourcentage de non-voisement du LTAS mettent en évidence des deltas moyens positifs ($\Delta +$) pour les émotions de la tristesse, du dégoût et de la surprise. Cela signifie que ces émotions ont été interprétées, en moyenne, avec davantage de souffle que l'état neutre. Il est essentiel de noter que cet indicage de souffle est particulièrement important pour la tristesse. Ces derniers résultats entrent en contradiction avec ce qui a été relevé pour le rapport L1-L0 avec ces émotions.

Néanmoins, pour les émotions de la peur, de la colère et de la joie, la valeur moyenne de delta obtenue pour le pourcentage de non-voisement est, quant à elle, négative ($\Delta -$), ce qui signifie que la voix des acteurs de notre échantillon était plus puissante et moins soufflée par rapport à la condition neutre pour ces trois émotions. Ces derniers résultats corroborent avec les données du rapport L1-L0 obtenues pour ces mêmes émotions.

Dans ce cas-ci, nous pensons également qu'il serait intéressant d'évaluer la coordination pneumo-phonique. En effet, nous pouvons imaginer que la perturbation de la respiration lors de la production d'une émotion pourrait créer plus de souffle dans la voix et donc impacter le rapport L1-L0 et le pourcentage de non-voisement du LTAS.

De plus, les résultats obtenus pour ce paramètre acoustique nous rappellent qu'il serait très intéressant de se renseigner davantage sur l'action exercée par les muscles de la périphérie laryngée et leur impact sur la qualité des paramètres acoustiques de la voix et la parole, lorsque des émotions sont interprétées. Effectivement, certaines d'entre elles sont susceptibles de produire un serrage glottique, créant alors une voix plus serrée et moins soufflée. C'est le cas par exemple du dégoût, qui crée une voix que nous pouvons qualifier de « compressée » (Fonagy, 1981 & Trojan, 1952 cités par Murray & Arnott, 1993). Cette action musculaire peut nous interroger également sur les contractions musculaires caractérisant les expressions faciales propres à chaque émotion.

Si nous nous interrogeons sur l'influence du genre de l'acteur sur le rapport L1-L0, nous remarquons que les hommes et les femmes montrent des moyennes de deltas comparables entre les conditions émotionnelles et la condition neutre. Nous ne pouvons donc pas tirer de conclusions à ce niveau. Toutefois, à propos du pourcentage de non-voisement, nous constatons que les hommes ont, en moyenne, obtenu un delta moins important que celui obtenu par les femmes pour le dégoût et la surprise, et ce, par rapport à la condition neutre. Par conséquent, les femmes ont produit ces émotions avec plus de souffle.

6.1.4 Le CPPs - dB

Au niveau du CPPs, nous ne possédions pas de données nous permettant de dresser des hypothèses confirmatoires. Quant aux résultats, nous observons de façon générale des deltas négatifs (Δ -) pour l'ensemble des émotions produites. Cela signifie que les 6 émotions d'Ekman ont été interprétées, en moyenne, avec moins de régularité, par rapport à l'état neutre. Ces éléments nous permettent d'enrichir les informations dont nous disposons à propos de ce paramètre.

En outre, Williams et Stevens (1972) cités par Murray et Arnott (1993) mettent en évidence qu'une irrégularité de la voix peut être présente avec la peur, en raison de la perturbation de la respiration avec cette émotion. Cette dernière information est très intéressante, car elle suggère que la coordination pneumo-phonique est très certainement altérée avec cette émotion et donc, que le CPPs peut certainement être impacté. Cela nous encourage, une nouvelle fois, à évaluer la coordination pneumo-phonique pour obtenir de nouvelles données sur l'importance de la respiration avec l'interprétation des émotions.

En ce qui concerne l'influence du genre de l'acteur sur le CPPs, les résultats obtenus ne nous permettent pas de tirer de conclusions sur la manifestation de ce paramètre avec les émotions d'Ekman.

6.2 Interprétation de nos résultats selon les hypothèses initialement posées pour les paramètres de la fluence

6.2.1 Le débit de parole – syllabes/seconde

Les données recueillies dans ce mémoire nous permettent de conclure que, pour toutes les émotions, le débit moyen produit est plus lent (Δ -) par rapport à la situation dans laquelle aucune émotion n'est produite. Cela rentre en contradiction avec les données que nous détenions au sujet de la fluence de la parole lorsque la peur, la colère et la joie sont produites. Néanmoins, pour le dégoût, les résultats confirment l'hypothèse initialement posée, car nous remarquons bien, en moyenne, une diminution du débit avec cette émotion (Banse & Scherer, 1996 ; Bänziger & al., 2015 ; Murray et Arnott, 1993). Pour la tristesse et la surprise, nous n'avions pas dressé d'hypothèses confirmatoires à leur sujet, mais les résultats récoltés nous ont permis de développer davantage les informations que nous connaissions à propos de l'impact de ces émotions sur le débit de parole.

Néanmoins, nous pensons que les résultats obtenus peuvent s'expliquer en partie par les conditions de passation des testings. Ce phénomène sera développé plus en détail dans la prochaine section de ce travail, mais nous pouvons expliquer ce dernier par le fait que les expérimentateurs soulignaient l'importance de prendre le temps d'articuler chaque phonème, pour chaque phrase de la liste 1 du corpus de Fharvard (Aubanel & al., 2020). De ce fait, les acteurs n'avaient pas la possibilité de produire les phrases avec une aisance et un débit spontané, ce qui a certainement biaisé les résultats obtenus.

De plus, les résultats peuvent être justifiés par le débit produit par chaque acteur lors de la condition neutre. En effet, dans cette situation, les acteurs étaient invités à produire les phrases sans interpréter les émotions et donc, sans réelle contrainte. Ensuite, lorsqu'il leur était demandé de produire les phrases avec les 6 émotions d'Ekman, ceux-ci prenaient le temps de s'approprier l'émotion et de l'intégrer aux phrases à produire, ce qui peut créer une plus grande charge cognitive. Ainsi, nous pensons qu'un acteur ayant produit une phrase avec un débit rapide dans la condition neutre sera susceptible d'obtenir des valeurs de débit qui s'avèreront inférieures (Δ -) lors de la production des émotions.

En outre, nous pensons également que les résultats récoltés pour le débit de parole peuvent s'expliquer par la connaissance des phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020). Par exemple, il est important de noter que l'actrice 01 est revenue une seconde fois pour les enregistrements, car son signal vocal a été écrêté lors de la première séance. Ainsi, elle a été familiarisée davantage aux consignes de passation et aux phrases à produire par rapport aux autres acteurs de l'échantillon. De ce fait, nous supposons que cela lui a permis de connaître davantage les phrases et d'ainsi augmenter son débit de parole par aisance. En revanche, ceux qui ne connaissaient pas les phrases avant les testings regardaient davantage l'écran avant l'enregistrement et produisaient un débit plus lent par manque de connaissance du texte.

Enfin, au vu des résultats obtenus pour le débit, il ne nous est pas possible de tirer de conclusions à propos de l'influence du genre de l'acteur lors de la production des différentes émotions étudiées dans ce travail.

6.2.2 La vitesse articulatoire – syllabes/seconde (sans pauses)

Pour ce paramètre, il est essentiel de préciser qu'aucun temps de pause supérieur à 250 ms n'a été détecté par le logiciel PRAAT (6.4.07) (Grosjean & Deschamps, 1975, cités par Grosman & al., 2018 ; Kühnert & Kocjančič Antolík, 2016). Effectivement, à nouveau, nous pensons que la méthodologie appliquée lors des enregistrements ne permettait pas aux différents acteurs

d'être libres dans leur interprétation des émotions et de produire de réelles pauses s'ils le désiraient. Ce sont les raisons pour lesquelles, en l'absence de pauses, les résultats obtenus pour la vitesse articulatoire sont égaux à ceux relevés pour le débit. Ainsi, les données exposées ci-après doivent être traitées avec beaucoup de prudence. Ce biais méthodologique sera plus amplement analysé dans la prochaine section de ce travail.

Pour la peur, la colère et la joie, nos résultats ne correspondent pas aux hypothèses initialement posées, car nous observons, en moyenne, un ralentissement de la vitesse articulatoire par rapport à la condition neutre (Δ -). Toutefois, pour la tristesse, nos résultats obtenus sont cohérents avec nos hypothèses confirmatoires. Nous observons bien une diminution de la vitesse articulatoire (Δ -) avec cette émotion (Banse & Scherer, 1996 ; Murray & Arnott, 1993). En ce qui concerne le dégoût et la surprise, nos hypothèses étaient exploratoires. Ainsi, les résultats obtenus pour ces 2 émotions nous ont permis d'acquérir de nouvelles données quant à leur manifestation à travers la vitesse articulatoire.

Malheureusement, tout comme le débit, nous ne pouvons pas nous prononcer à propos de l'influence du genre sur ce paramètre.

En conclusion, cette première partie de discussion nous a permis de développer les résultats obtenus pour chaque paramètre étudié et de les confronter avec les hypothèses qui ont été initialement posées. Cette confrontation nous a donné l'occasion de confirmer la majorité de nos hypothèses, d'interpréter les données qui ont été recueillies et de nous questionner sur les variables qui ont pu influencer ces dernières. De ce fait, nous avons tout d'abord pu constater qu'il était difficile, pour la majorité des paramètres, de tirer des conclusions à propos de l'influence du genre des participants de notre échantillon sur ceux-ci. Ensuite, nous avons mis en évidence l'importance de prendre en considération la coordination pneumo-phonique et de se renseigner sur le lien entre les contractions musculaires de la périphérie laryngée et les expressions faciales lors de la production orale d'émotions. Dans la prochaine section de ce mémoire, nous aborderons les biais présents dans notre méthodologie afin de réaliser un recul critique sur ceux-ci.

6.3 Limites méthodologiques

6.3.1 Caractéristiques des acteurs

Les compétences théâtrales

Lors du recrutement des participants, nous avons pris en compte leur âge, leur genre et le nombre d'années de pratique dans le domaine de la scène et du théâtre. Pour rappel, il était impératif que les participants aient au minimum 3 ans d'expérience dans ce domaine. Cependant, étant donné que nous ne possédions pas de connaissances concernant le milieu du théâtre et de l'acting, il était parfois difficile pour nous de réaliser des comparaisons ou de tirer des conclusions inter-individuelles sur les compétences théâtrales des acteurs et sur les résultats obtenus par ceux-ci. Nous pensons que mettre en place un questionnaire sur la qualité de la formation des acteurs ainsi que sur leur parcours artistique aurait été intéressant. De plus, nous pensons que s'entourer de personnes spécialisées dans le domaine théâtral aurait été bénéfique pour réaliser le recrutement. Effectivement, cela nous aurait permis d'obtenir plus de données précises quant à leurs capacités à interpréter des émotions, et en finalité, de prendre du recul sur les résultats obtenus en fonction des compétences de chacun d'eux.

Le sentiment d'auto-efficacité

Il est à noter qu'en plus des compétences propres à chaque acteur, la production orale d'émotions peut être impactée par le sentiment d'auto-efficacité¹⁸. En effet, une personne qui a une faible croyance en ses propres capacités peut ressentir moins de confiance en elle pour réaliser la tâche demandée. Ainsi, elle peut développer des sentiments négatifs et du stress à l'égard de cette tâche, qui vont l'empêcher de réaliser l'exercice confortablement (Bandura, 1977). Ainsi, nous pensons que le sentiment d'auto-efficacité doit être pris en compte dans ce travail, car les conséquences de celui-ci sont susceptibles d'avoir impacté les données acoustiques du signal de parole de l'acteur.

Le niveau de stress

Durant les testings, les expérimentateurs s'assuraient du confort des acteurs pour réaliser la tâche demandée. De ce fait, lorsque l'un d'entre eux nous paraissait stressé, nous lui demandions s'il se sentait anxieux ou non. La majorité d'entre eux nous confiaient qu'ils n'étaient pas stressés, mais que l'exercice demandé leur paraissait difficile. Ainsi, nous pouvions nous faire une idée subjective du niveau de stress ressenti par l'acteur, mais ne

¹⁸ Le sentiment d'auto-efficacité est défini comme la croyance d'un individu en sa capacité de réaliser une tâche (Bandura, 1999).

possédions pas de questionnaire normé nous permettant de l'évaluer objectivement. L'évaluation de ce stress aurait été intéressante afin de comprendre davantage la manifestation des paramètres acoustiques étudiés lors de l'analyse des données. Effectivement, l'effet du stress sur la voix et la parole a été étudié dans de nombreux travaux. Ceux-ci ont mis en évidence, par exemple, que le stress faisait augmenter la valeur de f_0 chez la plupart des individus (Giddens & al., 2013). Selon Larrouy-Maestri (2015), le stress peut également impacter l'intensité et le débit de parole en voix chantée et parlée. L'effet du stress a également été étudié lorsqu'il prend part à certains contextes émotionnels. Ainsi, les émotions les plus associées à un haut niveau de stress seraient la peur et la tristesse. Pour celles-ci, les auteurs constatent que le stress ressenti entraîne une diminution des variations intonatives de l'intensité (Paulmann & al., 2016). Finalement, sur base de ces informations, il est naturel de penser qu'un haut de niveau de stress est susceptible d'empiéter sur les caractéristiques acoustiques de la voix et de la parole de l'acteur, mais aussi sur la qualité de sa performance interprétative.

C'est pourquoi, en sachant que le stress et le sentiment d'auto-efficacité peuvent impacter les paramètres acoustiques de la voix, il aurait été important et pertinent d'inclure systématiquement un questionnaire ou une échelle en début de testing afin d'évaluer ces deux variables. Néanmoins, la proposition de ce support ne doit pas remettre en doute les compétences de l'acteur ou créer chez lui un stress supplémentaire avant de débiter la tâche.

6.3.2 Caractéristiques de la méthodologie appliquée

Méthodologie d'enregistrement des acteurs

La mise en place de la méthodologie des séances d'enregistrement a créé, selon nous, quelques biais susceptibles d'avoir eu un effet sur les résultats obtenus. Ceux-ci seront développés ci-après.

Tout d'abord, pour rappel, les acteurs avaient reçu la consigne de se positionner à 20 cm du mur du studio afin d'éviter que leur ombre ne soit visible sur les enregistrements visuels. De plus, ils portaient un micro-casque, relié par un fil, à la carte-son externe et devaient éviter tout mouvement parasite avec celui-ci, mais aussi avec leurs mains, leurs pieds, leurs vêtements, ... Puis, lorsque les acteurs étaient enregistrés, nous ne leur donnions pas la possibilité de réaliser tous les mouvements et gestes qu'ils souhaitaient. Nous avons donné ces instructions, car nous souhaitions éviter d'ajouter un bruit de fond supplémentaire aux enregistrements audio susceptible d'impacter leur qualité acoustique. Ensuite, il était nécessaire que l'acteur soit bien cadré au niveau du buste et du visage et positionné face à la caméra pour enregistrer ses

expressions faciales. Il est probable que ces consignes de placement aient entravé la qualité de l'interprétation des émotions par les acteurs. En effet, ceux-ci étaient contraints de rester relativement statiques et n'avaient pas la possibilité d'exprimer pleinement l'état émotionnel comme ils l'auraient fait dans une situation plus naturelle (ex : sur scène).

Cela nous interroge sur la qualité des résultats obtenus, car nous ne pouvons garantir que les enregistrements vocaux des acteurs reflètent réellement leurs capacités interprétatives des émotions. De ce fait, nous pensons que ces contraintes méthodologiques ont pu biaiser les valeurs des paramètres acoustiques analysés. Cette réflexion suggère qu'il serait intéressant de récolter des données sur les expressions faciales, la voix, mais aussi les gestes des acteurs, dans un cadre méthodologique leur permettant d'interpréter les émotions de manière plus naturelle. En effet, dans la littérature consultée, les bases de données émotionnelles combinent souvent ces 3 modalités, car elles sont considérées comme complémentaires dans la production des émotions (Trinh Van & al., 2022). Cela permettrait d'obtenir des données audio-visuelles plus objectives et de meilleure qualité.

Ensuite, le micro-casque était positionné précisément à 5 cm de la commissure droite de la bouche de chaque acteur, afin d'obtenir une distance micro-bouche suffisante pour éviter que du souffle empiète sur le signal sonore. Bien que nous vérifiions ce placement avant le commencement des testings et après la pause, et que nous demandions à l'acteur de ne pas y toucher, nous ne pouvions garantir que le micro ne bouge légèrement lors de l'interprétation des émotions. Ainsi, il est probable qu'il se soit rapproché de la bouche du participant au fur et à mesure du déroulement du testing et que du souffle ait été capté progressivement par celui-ci. Cela nous questionne sur la stabilité des données mesurées. Néanmoins, la distribution aléatoire des phrases et des émotions à produire pour chaque acteur nous a permis d'éviter que ce bruit éventuel se présente systématiquement pour les mêmes items et les mêmes émotions.

De plus, les acteurs étaient invités à produire chacune des émotions à faible et forte intensité émotionnelle. Bien que nous leur propositions des exemples de situations concrètes pour décrire ces différentes intensités et que celles-ci leur avaient été préalablement envoyées dans le formulaire d'informations, nous pensons que certains d'entre eux ont éprouvé des difficultés à les intégrer dans leur jeu d'acteur. En effet, nous avons observé que les acteurs avaient tendance à augmenter l'intensité sonore pour contraster ces deux conditions émotionnelles. Or, ce n'est pas ce qui était attendu. Lorsque cela nous paraissait flagrant durant les séances

d'enregistrement, nous leur rappelions brièvement la consigne. Néanmoins, puisque ceux-ci étaient libres de produire les émotions comme ils le souhaitaient, nous n'avons pas interrompu leurs productions. Nous pensons que cette difficulté de compréhension a impacté les résultats obtenus dans ces deux conditions émotionnelles. Il aurait été intéressant de recevoir l'ensemble des acteurs pour un atelier quelques semaines avant la période de testings, au cours duquel la nuance entre les différentes intensités émotionnelles leur aurait été expliquée. Durant les semaines séparant cet atelier de leur séance de testing, les acteurs auraient alors pris le temps de s'approprier ces différentes situations émotionnelles. Leur jeu d'acteur aurait sans doute été plus représentatif de ce qui était réellement attendu lors de leur séance d'enregistrement.

Enfin, il était demandé à l'acteur d'avoir mémorisé les différentes phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020) afin qu'il garde le regard dirigé vers la caméra lorsqu'il produisait l'émotion. Malgré le fait d'avoir laissé les phrases à disposition des participants, il leur était demandé de ne pas détourner le regard pour les relire lors des enregistrements. Nous nous sommes alors questionnées à propos de l'impact de cette contrainte méthodologique sur la charge cognitive des acteurs lorsque ceux-ci produisaient les émotions. En effet, la charge cognitive peut impacter les paramètres acoustiques de la voix et la parole comme l'intensité ou la fréquence fondamentale (Huttunen & al., 2011). Cela nous interroge sur la méthodologie à appliquer afin de limiter ce biais. Par exemple, comme exposé dans certaines bases de données émotionnelles, il serait intéressant de laisser les acteurs produire un discours émotionnel spontané sur un thème orienté (ex : *dyades interactions*) (Gil, 2009), ou encore d'utiliser des caméras qui suivent les mouvements de la tête pour permettre éventuellement aux acteurs de relire les phrases sans perdre de données visuelles.

Méthodologie de traitement des données

Tout d'abord, en ce qui concerne le traitement des audios, nous avons procédé à la sélection, parmi le spectre acoustique de l'enregistrement, des différentes phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020) lorsqu'elles étaient produites à forte intensité et lors du second essai. Pour rappel, nous avons traité ces items en particulier, car nous souhaitons obtenir des données acoustiques plus contrastées lorsque les acteurs avaient eu le temps de s'approprier l'émotion à produire. Cette sélection se réalisait manuellement via le logiciel PRAAT (6.4.07), ce qui nécessitait d'être très précis lors du placement du curseur au début et à la fin de chaque phrase à analyser. En effet, il était important que chaque enregistrement audio ne contienne pas de bruit, pour éviter que les données analysées par le logiciel ne soient

impactées par celui-ci. Malgré cette méthodologie stricte et minutieuse, il est possible que quelques découpages réalisés comportent des périodes de bruit, dues à un placement moins précis du curseur en début ou en fin d'audio. De plus, certaines périodes de bruit ont été conservées volontairement lorsque nous estimions qu'elles faisaient partie intégrante de l'interprétation de l'acteur (ex : inspirations lors des pleurs avec la tristesse, coups de glotte avec le dégoût). Ce bruit doit donc être pris en considération dans l'interprétation des données obtenues, car il a tantôt été conservé, tantôt été supprimé. Pour plus de précision, nous pensons qu'il aurait été préférable de supprimer le niveau de bruit ambiant des enregistrements avant de procéder à leur traitement, afin que cette conservation et cette suppression de bruit ne soient pas réalisées arbitrairement. Enfin, il aurait été intéressant d'utiliser un logiciel acoustique permettant de réaliser automatiquement et avec précision le placement des curseurs en début et fin d'audio, pour que celui-ci ne dépende pas de l'appréciation des expérimentateurs.

Ensuite, il est important de rappeler que, bien que la méthodologie appliquée ne laissât pas aux acteurs la possibilité de produire de réelles pauses dans leur discours, la présence du bruit ambiant dans les audios à traiter a sans aucun doute impacté également la détection de ces pauses par le logiciel PRAAT (6.4.07). Nous nous questionnons alors sur la présence éventuelle de pauses dans le discours des acteurs, qui n'auraient pas été détectées par le logiciel. Cela confirme, à nouveau, qu'il aurait été nécessaire de retirer le niveau de bruit ambiant des audios avant de les analyser pour que le logiciel détecte des périodes de silence d'une durée égale ou supérieure à 250 ms (Grosjean & Deschamps, 1975, cités par Grosman & al., 2018) et que les données acoustiques, telles que la vitesse articulatoire, puissent être calculées plus précisément.

Méthodologie d'interprétation des résultats obtenus

Pour rappel dans ce mémoire, nous avons fait le choix d'analyser les données acoustiques de 4 acteurs issus de notre échantillon initial composé de 12 acteurs. Or, ce petit échantillon ne nous permet pas de réaliser des statistiques descriptives et inférentielles sur les données obtenues et de tirer des conclusions valables pour l'ensemble de la population à partir de celles-ci. Ainsi, au vu de cette faible puissance statistique, il a été nécessaire d'interpréter les données avec beaucoup de prudence. De plus, ce petit échantillon ne nous permettait pas de tirer des conclusions sur les caractéristiques interprétatives des émotions entre les hommes et les femmes. Nous sommes convaincues que réaliser l'analyse des données de l'échantillon initial de 12 acteurs aurait été plus pertinent pour tirer des conclusions plus robustes sur la manifestation des paramètres acoustiques de la voix et de la parole avec les émotions d'Ekman,

et de mettre en évidence pour celles-ci les différences interprétatives inter-individuelles entre chaque participant.

Ensuite, à propos de l'analyse des données acoustiques obtenues, il est primordial de souligner que les interprétations étaient basées sur l'analyse de deltas moyens qui, eux-mêmes, étaient basés sur des moyennes effectuées pour chaque acteur, par émotion. Ainsi, nos analyses ne se basaient pas sur des valeurs précises, mais sur des moyennes. Cela représente une limite méthodologique importante, car chaque acteur a ses caractéristiques interprétatives et acoustiques propres, qui ne peuvent être mises en lumière dans des moyennes. Ainsi, puisque nous confrontions ces moyennes avec nos hypothèses initialement posées, il était difficile de réaliser une analyse intra-sujet précise.

Dans la prochaine et dernière section de ce mémoire, nous exposerons brièvement l'ensemble des résultats obtenus et tenterons de mettre en évidence les variables à prendre en considération et les perspectives qui nous paraissent intéressantes pour les travaux ultérieurs qui s'intéresseront à la production des émotions et leur impact sur la voix et la parole.

7. CONCLUSIONS ET PERSPECTIVES

En conclusion, ce mémoire nous a permis de confirmer plusieurs des hypothèses initialement posées sur les paramètres acoustiques de la voix et de la parole liés à la production orale d'émotions. Néanmoins, nous avons également observé des divergences entre nos résultats et ce qui était avancé dans la littérature. En outre, ce travail nous a permis d'enrichir les données dont nous disposons à ce sujet et d'ouvrir de nouvelles perspectives de travail dans le domaine de l'expression des émotions au sein du signal de parole.

Notre étude nous a permis de soulever qu'en moyenne, la peur, la colère et la joie se caractérisent par des valeurs de médiane et d'étendue plus élevées ($\Delta+$) pour la fréquence fondamentale et l'intensité, par rapport à la condition contrôle. Le rapport L1-L0 du LTAS augmente ($\Delta+$) alors que son pourcentage de non-voisement diminue ($\Delta-$) en comparaison à la condition neutre. Le CPPs, quant à lui, est plus faible avec l'interprétation de ces 3 émotions, tout comme le débit et la vitesse articulatoire ($\Delta-$). En d'autres termes, lorsque les acteurs de notre échantillon interprétaient la peur, la colère et la joie, leur voix était plutôt aigue, forte, et caractérisée par des variations intonatives de fréquence et d'intensité. Elle était puissante, peu soufflée et moins régulière par rapport à la condition neutre. Enfin, son débit et sa vitesse articulatoire étaient plus lents. Ces observations confirment et enrichissent la plupart de nos hypothèses posées avec ces émotions, sauf pour les données du débit et de la vitesse articulatoire, où nous nous attendions à observer une accélération de ces paramètres par rapport à la condition contrôle.

Ensuite, notre mémoire suggère qu'en moyenne, la tristesse et le dégoût se caractérisent par des valeurs de médiane et d'étendue de la fréquence fondamentale plus importantes ($\Delta+$) par rapport à la condition contrôle. La médiane de l'intensité diminue ($\Delta-$) alors que son étendue augmente ($\Delta+$) par rapport à la condition neutre, avec ces deux émotions. En ce qui concerne le LTAS, le rapport L1-L0 et le pourcentage de non-voisement sont plus élevés que la condition contrôle ($\Delta+$) avec la tristesse et le dégoût. Le CPPs, le débit et la vitesse articulatoire, quant à eux, se manifestent par des valeurs inférieures à celles obtenues lors de la condition contrôle ($\Delta-$). Cela signifie que les 4 acteurs de notre étude ont interprété, en moyenne, la tristesse et le dégoût avec une voix que nous pourrions qualifier d'aiguë et faible, produite avec davantage de variations intonatives de fréquence et d'intensité. Elle a été produite, en moyenne, avec plus d'irrégularités et plus de souffle par rapport à la condition contrôle. Enfin, son débit et sa vitesse articulatoire étaient ralentis avec ces émotions. À nouveau, nos observations sont cohérentes avec la plupart

des données issues de la littérature, sauf pour l'étendue et la médiane de la fréquence fondamentale avec la tristesse, où nous pensions obtenir des valeurs inférieures à la condition neutre (Δ^-) pour ces paramètres.

Finalement, pour la surprise, nous possédions peu de données sur la manifestation des paramètres acoustiques avec cette émotion. Néanmoins, nous pouvons conclure, sur base de nos observations, que la médiane et l'étendue de la fréquence fondamentale et de l'intensité augmentent (Δ^+) avec la surprise en comparaison à la situation contrôle. Pour le LTAS, le rapport L1-L0 et le pourcentage de non-voisement sont plus élevés également (Δ^+). À propos du CPPs, du débit et de la vitesse articulatoire, les données obtenues suggèrent que ces paramètres diminuent par rapport à la condition neutre (Δ^-). Concrètement, les acteurs ont produit, avec la surprise, une voix plus aiguë, forte et ce avec des variations intonatives de fréquence et d'intensité. Leur voix était peu soufflée et irrégulière. Leur débit et leur vitesse articulatoire étaient, quant à eux, plus lents par rapport à la condition contrôle.

Bien que l'ensemble de ces résultats soient particulièrement intéressants et confirment la majorité de nos hypothèses posées, il est primordial d'interpréter ces données avec beaucoup de prudence. En effet, de nombreux biais méthodologiques sont présents dans cette étude et sont susceptibles d'avoir impacté nos résultats. Premièrement, les modalités de passation des séances d'enregistrement étaient assez contraignantes (charge cognitive importante, difficulté de compréhension des consignes, positionnement strict, ...). Celles-ci ont, sans doute, empiété sur l'interprétation des émotions par les acteurs. Deuxièmement, plusieurs variables externes sont susceptibles d'avoir impacté nos données et devraient être prises en considération dans les études ultérieures. C'est le cas de la qualité de la formation artistique et des compétences propres aux acteurs, du sentiment d'auto-efficacité et du stress ressenti. Troisièmement, la présence du bruit de fond lors du traitement des enregistrements a très certainement influencé la manifestation des différents paramètres acoustiques étudiés, tout comme la méthode de sélection des périodes de bruit et de placement des curseurs en début et fin d'audio. Enfin, la petite taille de notre échantillon ne nous permettait pas de tirer de conclusions robustes sur les données obtenues. Pour finir, les analyses ont été effectuées sur des moyennes, ce qui « lissait » les caractéristiques propres à chaque acteur.

En outre, pour contourner ces difficultés méthodologiques et développer la recherche dans ce domaine, nous pensons que les futurs projets s'intéressant à la production des émotions dans le signal de parole devraient prendre en considération l'impact des états émotionnels sur la coordination pneumo-phonique et sur la contraction de la musculature laryngée, étant donné les

changements acoustiques qui peuvent être créés par celles-ci dans la voix et la parole. Il serait également intéressant d'étudier davantage les liens entre les contractions musculaires liées aux expressions faciales et leur impact sur les paramètres acoustiques de la voix et de la parole. Ensuite, si les futurs chercheurs collaborent avec des acteurs, il serait judicieux de créer un environnement de testing naturel, peu contraignant et laissant libre cours à leur jeu d'acteur, afin d'obtenir des données émotionnelles reflétant au mieux leurs compétences interprétatives des émotions étudiées. Par ailleurs, il serait intéressant d'évaluer via des questionnaires, des échelles numériques, des échelles verbales ou d'autres supports, les variables susceptibles d'avoir empiété sur l'interprétation de ces émotions (qualité de la formation artistique et compétences, sentiment d'auto-efficacité, stress, ...). Pour le traitement des données acoustiques, il sera important de veiller à ce que le bruit de fond éventuel soit supprimé des enregistrements et que la sélection des audios et des périodes de bruit soit réalisée de manière contrôlée et automatique. Enfin, il sera indispensable de réaliser ces études sur un échantillon plus important, pour pouvoir tirer des conclusions robustes et plus fines sur les données obtenues, mais également pour faire des liens entre les résultats et les caractéristiques propres aux participants (genre, nombre d'années d'expérience dans le domaine du théâtre, ...).

En conclusion, notre mémoire a permis de contribuer à la création de la base de données émotionnelle « EVE » et de pallier le manque de consensus et le peu de données disponibles dans la littérature traitant de l'influence de la production des émotions sur les paramètres de la voix et de la parole. Notre étude a également ouvert de nouvelles perspectives de travail dans ce domaine de recherche. Nous espérons que les données présentées dans ce mémoire pourront être exploitées ou que la méthodologie que nous avons appliquée pourra inspirer le développement de travaux ultérieurs dans ce domaine de travail.

8. BIBLIOGRAPHIE

- Adigwe, A., Tits, N., Haddad, K.E., Ostadabbas, S., & Dutoit, T. (2018). The Emotional Voices Database: Towards Controlling the Emotion Dimension in Voice Generation Systems. *ArXiv, abs/1806.09514*.
- Antonetti, A. E. d. S., Siqueira, L. T. D., Gobbo, M. P. d. A., Brasolotto, A. G., & Silverio, K. C. A. (2020). Relationship of Cepstral Peak Prominence-Smoothed and Long-Term Average Spectrum with Auditory-Perceptual Analysis. *Applied Sciences, 10*(23), 8598. <https://doi.org/10.3390/app10238598>
- Ashhad, S., Kam, K., Negro, C. A. D., & Feldman, J. L. (2022). Breathing Rhythm and Pattern and Their Influence on Emotion. *Annual Review of Neuroscience, 45*(1). <https://doi.org/10.1146/annurev-neuro-090121-014424>
- Aubanel, V., Bayard, C., Strauß, A., & Schwartz, J. L. (2020). The Fharvard corpus: A phonemically-balanced French sentence resource for audiology and intelligibility research. *Speech Communication, 124*, 68–74. <https://doi.org/10.1016/j.specom.2020.07.004>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-Efficacy : The Exercise of Control. *Journal of Cognitive Psychotherapy, 13*(2), 158–166. <https://doi.org/10.1891/0889-8391.13.2.158>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Bänziger, T., & Scherer, K.R. (2005). The role of intonation in emotional expressions. *Speech Communication, 46*, 252–267.
- Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path Models of Vocal Emotion Communication. *PLOS ONE, 10*(9), e0136675. <https://doi.org/10.1371/journal.pone.0136675>
- Benarous, R. & Giltay, Z. (2023). *Etude des disfluences et du débit de parole chez l'enfant tout-venant de 4 et 8 ans* [Mémoire de Master, Université de Liège]. Matheo. <http://hdl.handle.net/2268.2/17109>
- Bera, A. K., & Koley, M. (2023). A History of the Delta Method and Some New Results. *Sankhya B*. <https://doi.org/10.1007/s13571-023-00305-9>
- Bharadwaj, S., & Acharjee, P.B. (2022). Exploring human voice prosodic features and the interaction between the excitation signal and vocal tract for Assamese speech. *International Journal of Speech Technology, 26*, 77–93.

- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F., & Weiss, B. (2005). A database of German emotional speech. *Interspeech*.
- Butcher, A. (1981). Aspects of the speech pause: Phonetic correlates and communication functions. *Arbeitsbericht Kiel*, (15), 1-233.
- Cantor-Cutiva, Bottalico et al., (2019). Toward a Consensus Description of Vocal Effort, Vocal Load, Vocal Loading, and Vocal Fatigue. *Journal of Speech, Language, and Hearing Research*, (63)2, 209-532.
- Cao, H., Cooper, D.G., Keutmann, M.K., Gur, R.C., Nenkova, A., Verma, R. (2014). Crema-d: Crowd-sourced emotional multimodal actors dataset. *IEEE transactions on affective computing*, 5(4), 377–390.
- Catani et al., (2013). A revised limbic system model for memory, emotion and behaviour. *Neuroscience & Biobehavioral Reviews*, 37 (8), 1724-1737
- Chambers, F. (1997). What Do We Mean by Fluency? *System*, 25, 535-544. [https://doi.org/10.1016/S0346-251X\(97\)00046-8](https://doi.org/10.1016/S0346-251X(97)00046-8)
- Cowen et al., (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nat Hum Behav*, 3(4), 369–382. DOI 10.1038/s41562-019-0533-6.
- De Bodt M., Heylen L., Mertens F., Vanderwegen F. & Van de Heyning P. (2008) *Stemstoornissen, handboek voor de klinische praktijk*. Antwerpen : Garant
- Di Scala, E. (2023). Proposition de définition d'un facteur d'autoguérison. *Hegel*, 4, 335-346. <https://doi.org/10.3917/heg.134.0335>
- Ekman, P. (1993). Facial Expression and Emotion. *American Psychologist, Journal of Personality and Social Psychology*, 48, 384-392. <http://dx.doi.org/10.1037/0003-066X.48.4.384>
- Ekman, P. (2005). Basic Emotions. Dans *Handbook of Cognition and Emotion* (p. 45–60). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470013494.ch3>
- Ekman, P., & Cordaro, D. (2011). What is Meant by Calling Emotions Basic. *Emotion Review*, 3(4), 364–370. <https://doi.org/10.1177/1754073911410740>
- Frick, R.W. (1985). Communicating Emotion: The Role of Prosodic Features. *Psychological Bulletin*, 97(3), 412-429.
- Geer, J. H., Estupinan, L. A., & Manguno-Mire, G. M. (2000). Empathy, social skills, and other relevant cognitive processes in rapists and child molesters. *Aggression and Violent Behavior*, 5(1), 99–126. [https://doi.org/10.1016/s1359-1789\(98\)00011-1](https://doi.org/10.1016/s1359-1789(98)00011-1)
- Ghio, A. (2021). L'amplitude du signal sonore. In *La voix : anatomie, physiologie et explorations* (pp. 107-116). Louvain-la-Neuve, Belgique : De Boeck Supérieur.

- Giddens, C. L., Barron, K. W., Byrd-Craven, J., Clark, K. F., & Winter, A. S. (2013). Vocal indices of stress: a review. *Journal of voice: official journal of the Voice Foundation*, 27(3), 390.e21–390.e3.9E29. <https://doi.org/10.1016/j.jvoice.2012.12.010>
- Gil, S. (2009). Comment étudier les émotions en laboratoire. *Revue électronique de Psychologie Sociale*, 4, 15-24.
- Giovanni, A. (2021). 10. Réglage du volume sonore (intensité) de la voix. Dans A. Giovanni, *La voix : anatomie, physiologie et explorations* (pp. 33-35). De Boeck Supérieur.
- Gournay, P., Lahaie, O., & Lefebvre, R. (2018). A canadian French Emotional Speech Dataset. *ResearchGate*, 399-402. DOI: 10.1145/3204949.320812
- Grosman, I., Simon, A. & Degand, L. (2018). Variation de la durée des pauses silencieuses : impact de la syntaxe, du style de parole et des disfluences. *Langages*, 211, 13-40. <https://doi.org/10.3917/lang.211.0013>
- Halberstam, B. (2004). Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL Journal of Otorhinolaryngology and Related Specialities*, 66(2), 70–73.
- Hareli, S., Halhal, M., & Hess, U. (2018). Dyadic Dynamics: The Impact of Emotional Responses to Facial Expressions on the Perception of Power. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01993>
- Hellige, J. B. (2002). Spécialisation hémisphérique : données récentes. *Revue de neuropsychologie*, 12(1), 7-49.
- Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., Rubin, A., Eller, R., Chandran, S., Abaza, M., Lyons, K., Divi, V., Lott, J., Johnson, J., & Hillenbrand, J. (2014). Quantifying the Cepstral Peak Prominence, a Measure of Dysphonia. *Journal of Voice*, 28(6), 783–788. <https://doi.org/10.1016/j.jvoice.2014.05.005>
- Hess, U. (2003). Now you see it, now you don't--the confusing case of confusion as an emotion: Commentary on Rozin and Cohen (2003). *Emotion*, 3(1), 76–80. <https://doi.org/10.1037/1528-3542.3.1.76>
- Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., & Leino, T. (2011). Effect of cognitive load on speech prosody in aviation : Evidence from military simulator flights. *Applied ergonomics*, 42(2), 348–357. <https://doi.org/10.1016/j.apergo.2010.08.005>
- Jürgens, R., et al., (2015). Effect of Acting Experience on Emotion Expression and Recognition in Voice: Non-Actors Provide Better Stimuli than Expected. *J Nonverbal Behav* 39, 195–214. <https://doi.org/10.1007/s10919-015-0209-5>
- Kent et al., (2021). Acoustic parameters of voice in typically developing children ages 4–19 years, *International Journal of Pediatric Otorhinolaryngology* 142.

- Kramer, E. (1963). Judgments of personal characteristics and emotions from nonverbal properties of speech. *Psych. Bull.*, 60,408–420.
- Krivokapić, J. (2007). Prosodic planning : Effects of phrasal length and complexity on pause duration, *Journal of phonetics*, 35 (2), 162-179.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in cognitive sciences*, 14(7), 293–300. <https://doi.org/10.1016/j.tics.2010.05.001>
- Kühnert, B. & Kocjančič Antolík, T. (2016). *Stratégies d'adaptation de la vitesse d'articulation lors de conversations spontanées entre locuteurs natifs et non-natifs* [Conférence]. JEP-TALN-RECITAL, Laboratoire de Phonétique et Phonologie (UMR 7018, CNRS -Sorbonne Nouvelle) 19 rue des Bernadins, 75005 Paris, France. <https://aclanthology.org/2016.jeptalnrecital-jep.74>
- Kushner, R. E., & Bickley, C. A. (1995). Analysis and perception of voice similarities among family members. *The Journal of the Acoustical Society of America*, 98(5), 2936. <https://doi.org/10.1121/1.414098>
- Lacroix, A. (2016). La prosodie dans tous ses états. *Bulletin de Psychologie*, 542 (2), 83-85. DOI 10.3917/bupsy.542.0083.
- Larrouy-Maestri, P. (2015). Influence du stress sur la voix parlée et chantée. *Langage et pratiques*, 56, 33-42.
- Larrouy-Maestri, P., Poeppel, D., & Pell, M. D. (2024). The Sound of Emotional Prosody: Nearly 3 Decades of Research and Future Directions. *Perspectives on Psychological Science*. <https://doi.org/10.1177/17456916231217722>
- Lassalle, A., et al., (2019). The EU-Emotion Voice Database. *Behavior Research Methods*, 51, 493–506. <https://doi.org/10.3758/s13428-018-1048-1>.
- Lingelbach et al., (2023). GAUDIE: Development, validation, and exploration of a naturalistic German AUDitory Emotional database. *Behavior Research Methods*, <https://doi.org/10.3758/s13428-023-02135-z>
- Liu, C.-T., Chu, S.-W., & Chen, Y.-S. (2022). On the Relationship between Speech Intelligibility and Fluency Indicators among English-Speaking Individuals with Parkinson's Diseases. *Behavioural Neurology*, 2022, 1–8. <https://doi.org/10.1155/2022/1224680>
- Liu, P., et al., (2021). Unattended Emotional Prosody Affects Visual Processing of Facial Expressions in Mandarin-Speaking Chinese: A Comparison With English-Speaking Canadians. *Journal of cross-cultural psychology*, 52(3), 275-294.
- Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). *PLoS ONE*, 13 (5), pe019639. Zenodo. <https://doi.org/10.5281/zenodo.1188976>

- Lowell, S. Y., Colton, R. H., Kelley, R. T., & Mizia, S. A. (2013). Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *Journal of voice: official journal of the Voice Foundation*, 27(4), 393–400. <https://doi.org/10.1016/j.jvoice.2013.02.005>
- Mauchand, M., & Zhang, S. (2023). Disentangling emotional signals in the brain: an ALE meta-analysis of vocal affect perception. *Cognitive, Affective, & Behavioral Neuroscience*, 23, 17–29
- Meynadier, Y. (2013). Éléments de phonétique acoustique. Dans N. Nguyen & M. Adda-Decker (dirs.), *Méthodes et outils pour l'analyse phonétique des grands corpus oraux* (pp. 25–83). Hermès.
- Monroy, M., Cowen, A. S., & Keltner, D. (2022). Intersectionality in emotion signaling and recognition: The influence of gender, ethnicity, and social class. *Emotion*, 22(8), 1980–1988. <https://doi.org/10.1037/emo0001082>
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2), 1097–1108. <https://doi.org/10.1121/1.405558>
- Naqvi, Y., & Gupta, V. (2023, 28 avril). Functional Voice Disorders. StatPearls [Internet]. Treasure Island (FL). NCBI Bookshelf. National Center for Biotechnology Information. PMID: 33085329. <https://www.ncbi.nlm.nih.gov/books/NBK563182/>
- Nordenberg, M. & Sundberg, J. (2004). Effect on LTAS of vocal loudness variation. *Logopedics Phoniatrics Vocology*, 29(4), 183-191, DOI : 10.1080/14015430410004689
- Park, J., Choi, S., Takatoh, J., Zhao, S., Harrahill, A., Han, B.-X., & Wang, F. (2024). Brainstem control of vocalization and its coordination with respiration. *Science*, 383(6687). <https://doi.org/10.1126/science.adi8081>
- Paulmann, S., Furnes, D., Bøkenes, A. M., & Cozzolino, P. J. (2016). How Psychological Stress Affects Emotional Prosody. *PloS one*, 11(11), e0165022. <https://doi.org/10.1371/journal.pone.0165022>
- Perry, P. (2011). Concept Analysis: Confidence/Self-confidence. *Nursing Forum*, 46(4), 218–230. <https://doi.org/10.1111/j.1744-6198.2011.00230.x>
- Philippot, P. (2007). Qu'est-ce qu'une émotion ? Dans P. Philippot, *Emotion et psychothérapie*, (pp.11-64). Wavre : Mardaga.
- Popolo, P. S., & Johnson, A. M. (2020). Relating Cepstral Peak Prominence to Cyclical Parameters of Vocal Fold Vibration from High-Speed Videoendoscopy Using Machine Learning: A Pilot Study. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2020.01.026>
- Ramsland, K. (2012). The man of 1,000 faces: Paul Ekman and the science of facial analysis. *Forensic Examiner*, 22(1), 64-70.

- Remacle et al., (2018). Vocal Change Patterns During a Teaching Day: Inter- and Intra-subject Variability. *Journal of Voice*, 32 (1).
- Rodríguez-Fuentes, R. A., & Calle-Díaz, L. (2023). Fluency development beyond speech rate: A study on the effects of read aloud protocols on EFL learners. *Ampersand*, 11, 100133. <https://doi.org/10.1016/j.amper.2023.100133>
- Rothauser, E.H. (1969). IEEE Recommended Practice for Speech Quality Measurements, *IEEE Transactions on Audio and Electroacoustics*, 17(3), pp. 225-246. doi: 10.1109/TAU.1969.1162058.
- Russell et al., (2003). Facial and Vocal Expressions of Emotion. *Annual review of psychology*, 54 (1), 329-349.
- Scherer, K. R. (1995). Expression of Emotion in Voice and Music. *Journal of Voice*, 9(3), 235–248. [https://doi.org/10.1016/S0892-1997\(05\)80231-0](https://doi.org/10.1016/S0892-1997(05)80231-0)
- Sobin, C., & Alpert, M. (1999). Emotion in Speech : The Acoustic Attributes of Fear, Anger, Sadness, and Joy. *Journal of Psycholinguistic Research*, 28(4).
- Spearman, C., & Holzinger, K. (1924). The Sampling Error in the Theory of Two Factors. *British journal of Psychology*, 15, 17–19.
- Švec, J. G., & Granqvist, S. (2018). Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech. *Journal of Speech, Language, and Hearing Research*, 61(3), 441–461. https://doi.org/10.1044/2017_jslhr-s-17-0095
- Themistocleous, C., Eckerström, M., & Kokkinakis, D. (2020). Voice quality and speech fluency distinguish individuals with mild cognitive impairment from healthy controls. *PLoS ONE*, 15(7), Article e0236009. <https://doi.org/10.1371/journal.pone.0236009>
- Trinh Van et al., (2022). Emotional Speech Recognition Using Deep Neural Networks. *Sensors*, 22, 1414. <https://doi.org/10.3390/s22041414>.
- Van Rijn, P., & Larrouy-Maestri, P. (2023). Modelling individual and cross-cultural variation in the mapping of emotions to speech prosody. *Nature Human Behaviour* ,7, 386–396. <https://doi.org/10.1038/s41562-022-01505-5>
- Vos, S., et al., (2023). The Sound of Emotion: Pinpointing Emotional Voice Processing Via Frequency Tagging EEG. *Brain Sci.*, 13, 162.

9. ANNEXES

Annexe 1. Mises en situations facultatives à destination des acteurs pour interpréter les émotions (Monroy et al., 2022)

Joie	Vous recevez un cadeau d'un ami et vous vous sentez heureux parce que c'est quelque chose que vous vouliez vraiment.
Tristesse	Votre cousin vient de mourir et vous êtes très triste. Vous vous retirez dans un coin tranquille pour pleurer en privé, partageant vos souvenirs et votre chagrin avec les membres de votre famille.
Peur	Vous vous retrouvez soudain face à un animal dangereux et vous avez très peur. Vous reculez lentement tout en gardant un œil sur l'animal, et vous cherchez un endroit sûr pour vous abriter.
Dégoût	Vous venez de manger des aliments avariés et vous vous sentez très dégoûté. Vous recrachez immédiatement la nourriture, vous vous lavez la bouche à plusieurs reprises, et vous jetez le reste dans la poubelle avec une expression de dégoût.
Colère	Vous avez été insulté et vous êtes très en colère. Vous répondez avec fermeté à l'insulte, et vous exprimez votre mécontentement de manière assertive.
Surprise	Vous voyez une lumière brillante au milieu de la nuit et vous êtes très surpris. Vous scrutez l'origine de la lumière avec curiosité.
Confiance en soi	Vous êtes sur le point de donner une présentation importante devant un public. Vous parlez avec assurance, et maintenez un contact visuel constant avec votre auditoire, montrant ainsi votre confiance en vos compétences.
Confusion	Vous avez du mal à comprendre quelque chose et vous vous sentez très confus à ce sujet. Vous posez des questions pour éclaircir la situation, et essayez de trouver des réponses pour dissiper votre confusion.
Mépris	Vous voyez une personne riche jeter des vêtements coûteux à la poubelle et vous vous sentez méprisé par un tel acte de gaspillage. Vous secouez la tête avec désapprobation et dédaignez l'acte comme irresponsable et insensible.
Sympathie	Vous voyez une personne blessée et vous éprouvez de la sympathie pour elle. Vous vous approchez avec précaution, exprimez votre préoccupation, et offrez de l'aide ou du réconfort à la personne blessée.

Annexe 2. Tableau synthétique de différentes bases de données émotionnelles abordées dans la littérature ainsi que leur méthodologie

Titre de la base de données	Langue	Méthodologie	Caractéristiques des interprètes	Emotions
Emo-DB (Burkhardt & al., 2005)	Allemand	<i>Modalité</i> : Audio. Les locuteurs ont été enregistrés lorsqu'ils ont produit différents types de phrases selon les émotions ciblées (phrases absurdes ne voulant rien dire et phrases pouvant être utilisées dans la vie quotidienne).	40 locuteurs tout-venants (20 femmes et 20 hommes)	Colère, dégoût, peur, joie, tristesse , l'ennui, le neutre
CREMA-D (Cao & al., 2014)	Anglais	<i>Modalité</i> : Multimodal (audio-visuel, visuel, audio). Les acteurs ont été enregistrés lorsqu'ils ont produit 12 phrases selon les émotions choisies. Seule 1 des 12 phrases a été produite selon 3 intensités (faible, moyenne, forte)	91 acteurs professionnels (48 hommes et 43 femmes)	Colère, dégoût, peur, joie, tristesse , neutre
RAVDESS (Livingstone, & Russo, 2018)	Anglais	<i>Modalité</i> : Multimodal (audio-visuel, visuel, audio). Les acteurs ont été invités à interpréter les états émotionnels selon 60 énoncés parlés et 44 énoncés chantés et ce, selon 2 intensités (faible et forte).	24 acteurs professionnels (12 hommes et 12 femmes)	Colère, dégoût, peur, joie, tristesse, surprise , neutre, calme
Multimodal Database of Emotional Speech, Video and Gestures (Sapinski & al., 2018)	Anglais	<i>Modalité</i> : Multimodal (audio-visuel, visuel, audio). Les acteurs ont été enregistrés lorsqu'ils interprétaient les états émotionnels selon les 6 émotions de base.	16 acteurs professionnels (8 hommes et 8 femmes)	Colère, dégoût, peur, joie, tristesse, surprise
The EU-Emotion Voice Database (Lassalle & al., 2018)	Anglais, suédois et hébreux	<i>Modalité</i> : Audio. Les acteurs ont été enregistrés lorsqu'ils produisaient des phrases (à contenu sémantiquement neutre ou à contenu émotionnel) selon les émotions demandées. <i>Le EU-Emotion Voice Database provient du "EU-Emotion Stimulus Set" qui contient également des stimuli d'expressions émotionnelles à travers les expressions faciales, les gestes, ...</i>	54 acteurs au total (3 groupes de 18 personnes, dont 9 femmes)	Colère, dégoût, peur, joie, tristesse, surprise , honte, ennui, déception, frustration, excitation, blessé, intérêt, jalousie, plaisanterie, sympathie, fierté, sournoiserie, hostilité, inquiétude

Note. Les émotions reprises en gras dans la colonne de droite de ce tableau font partie des 6 émotions universelles de Paul Ekman, qui font l'objet de notre étude.

Annexe 3. Méthodes de découpage des audios et de récolte de données acoustiques dans le logiciel PRAAT (version 6.4.07 – March 17, 2024)

Découpage des audios

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser en se plaçant au tout début du commencement de l’onde sonore. Prendre note du temps de début de l’onde sonore.
3. Se placer à la fin de l’onde sonore.
4. Cliquer sur “Time”, ensuite select.
5. Encoder le temps de début de l’onde sonore dans la section : “Start of selection”.
6. Réécouter l’audio et enregistrer (file > Save as wav file).
7. Enregistrer l’audio selon la nomenclature établie (numéro acteur_Numéro phrase_Numéro émotion_I2_T2).

Analyse acoustique des audios

Analyse de la médiane de la fréquence fondamentale (Hz)

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser.
3. Cliquer sur “Pulses” et ensuite “Voice Report”.
4. Dans la section “Pitch”, nous pouvons ensuite obtenir la médiane (*median pitch*).

Analyse de l’étendue de la fréquence fondamentale (Hz)

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser.
3. Cliquer sur “Pulses” et ensuite sur “Voice Report”.
4. Dans la section “Pitch”, nous avons les deux valeurs suivantes : “Minimum Pitch” (f_o_{min}) et “Maximum Pitch” (f_o_{max}) nous permettant de calculer l’étendue ($f_o_{étendue}$) par la suite
5. L’étendue est finalement calculée comme suit : $f_o_{étendue} = f_o_{max} - f_o_{min}$

Analyse de la médiane de l'intensité (dB)

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser.
3. Cliquer sur “Intensity” et ensuite “Intensity settings” afin de configurer “Averaging method” en cochant “median”. Appuyer ensuite sur “OK”.
4. Retourner sur “Intensity” et ensuite “Get intensity” (F8).

Analyse de l'étendue de l'intensité (dB)

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser.
3. Cliquer sur “Intensity” et ensuite sur “Get minimum intensity” (*Intensité_{min}*) ainsi que sur “Get maximum intensity” (*Intensité_{max}*).
4. Il est ensuite possible de calculer l’étendue (*Intensité_{étendue}*) comme suit : $Intensité_{étendue} = Intensité_{max} - Intensité_{min}$

Analyse du CPPs (dB) (voir <https://www.jamescurtisphd.me/tutorials/voice/cpp>)

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “Analyse to periodicity”.
2. Sélectionner “To Power Cepstrogram” et confirmer “OK”.
3. Sélectionner “Query “ et ensuite “Get CPPs”.
4. Modifier “Time averaging window (s) : 0,01
5. Modifier “Quefreny averaging window (s) : 0,001
6. Sélectionner OK

Analyse du rapport L1_L0 du LTAS (dB)

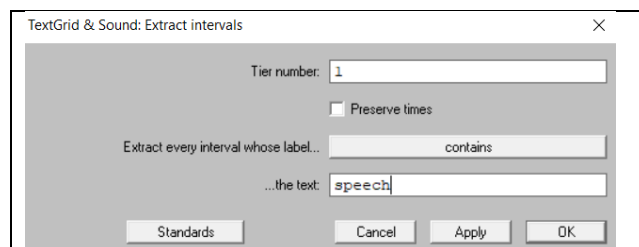
1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner « Analyse spectrum » puis « to LTAS (pitch corrected) ».
2. Modifier la valeur de « Maximum frequency » en 8000 (Hz) et sélectionner « ok ».
3. Sélectionner le nouveau fichier apparu dans la fenêtre objet et appuyer sur « Query » et sélectionner « Get slope ». Une nouvelle fenêtre apparait.
4. Compléter « Low Band » avec les valeurs 50,0 et 300,0. « High band » avec les valeurs 300,0 et 800,0 ensuite sélectionner « dB » puis « ok ». La valeur ainsi obtenue correspond à la valeur absolue du rapport L0-L1.

Analyse du pourcentage de non-voisement (%)

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser.
3. Cliquer sur “Pulses” et ensuite sur “Voice Report”.
4. Dans la catégorie “Voicing”, nous pouvons obtenir la valeur “*Fraction of locally unvoiced frames*”. Cela correspond donc au pourcentage de non-voisement de l’audio sélectionné.

Analyse de la vitesse articulatoire (Nombre de syllabes/seconde) (Benarous, R. & Giltay, Z., 2023).

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “Annotate” et ensuite “To TextGrid (Speech activity)”.
2. Compléter “Min. Nonspeech interval (s)” par 0.25.
3. Sélectionner les deux fichiers présents dans la fenêtre objet et ensuite cliquer sur “Extract”.
4. Ensuite, cliquer sur “Extract intervals where...” et modifier “Extract every interval whose label” en sélectionnant “Contains”.
5. Compléter “the text” par “speech”.
6. Sélectionner les nouveaux fichiers dans la fenêtre objet, nommés “Sound_XX_XX_speech_X” et analyser le son total de l’audio.



Note. Screenshot des étapes 4. et 5. permettant l’analyse de la vitesse articulatoire via le logiciel PRAAT (6.4.07).

Analyse du débit (Nombre de syllabes/temps total) :

1. Une fois le fichier audio sélectionné dans la fenêtre objet, sélectionner “View and edit”.
2. Sélectionner la partie de l’audio à analyser.
3. On obtient ensuite le temps total de l’audio (en seconde).
4. Compter le nombre de syllabes produites par enregistrement.
5. Le débit peut finalement être calculé en divisant le nombre de syllabes par le temps total de l’audio.

Annexe 4. Nomenclature pour renommer les fichiers des participants

AA_PP_EE_I2_T2

1. AA: Numéro de l'acteur (*01, 02, 07 ou 08*)
2. PP: Numéro de la phrase provenant de la liste 1 du corpus de Fhavard (Aubanel & al., 2020) (*de 01 à 10*)
3. EE: Numéro de l'émotion produite
 - 01: Peur
 - 02: Colère
 - 03: Joie
 - 04: Tristesse
 - 05: Dégoût
 - 06: Surprise
4. I2: Intensité 2 (*pour préciser qu'il s'agit de l'émotion produite à forte intensité*)
5. T2: Tentative 2 (*pour préciser qu'il s'agit de l'émotion produite lors du second essai*)

Annexe 5. Réglages de la carte-son externe du micro (Focusrite iTrack Solo Lightning)



Annexe 6. Tableaux reprenant les moyennes des résultats obtenus par paramètre acoustique pour chaque acteur et chaque condition émotionnelle.

- a) *Moyennes brutes des médianes obtenues pour la fréquence fondamentale (Hz) chez chaque participant et pour chaque émotion*

	F01	F07	H02	H08
Neutre	205,98	191,15	108,83	113,22
Peur	334,86	365,37	249,30	189,57
Colère	257,51	334,80	212,21	218,53
Joie	302,41	387,46	250,47	226,97
Tristesse	283,44	373,30	157,56	195,51
Dégoût	214,13	304,22	138,48	188,20
Surprise	329,72	339,15	178,15	193,55

- b) *Moyennes brutes des étendues obtenues pour la fréquence fondamentale (Hz) chez chaque participant et pour chaque émotion*

	F01	F07	H02	H08
Neutre	157,15	114,65	75,99	83,69
Peur	164,11	215,28	195,61	191,15
Colère	183,32	257,80	154,11	251,18
Joie	237,77	204,08	224,90	210,42
Tristesse	162,29	323,53	145,84	225,30
Dégoût	181,04	244,58	102,04	184,12
Surprise	262,73	346,72	175,72	276,29

- c) *Moyennes brutes des médianes obtenues pour l'intensité (dB) chez chaque participant et pour chaque émotion*

	F01	F07	H02	H08
Neutre	71,75	70,22	70,64	68,07
Peur	73,54	70,07	70,32	68,51
Colère	72,27	69,50	71,53	68,59
Joie	71,93	68,94	70,48	69,79
Tristesse	71,75	66,21	67,00	66,47
Dégoût	71,22	68,14	71,90	67,49
Surprise	71,99	68,84	72,80	69,20

- d) *Moyennes brutes des étendues obtenues pour l'intensité (dB) chez chaque participant et pour chaque émotion*

	F01	F07	H02	H08
Neutre	30,86	34,56	34,34	36,31
Peur	36,00	44,03	36,56	46,98
Colère	37,27	48,32	39,67	49,02
Joie	36,32	50,21	41,52	51,55
Tristesse	36,89	47,28	43,41	48,34
Dégoût	34,26	45,87	35,04	48,30
Surprise	37,96	49,69	37,74	46,51

e) *Moyennes brutes des rapports L1_L0 (dB) obtenus chez chaque participant et pour chaque émotion*

	F01	F07	H02	H08
Neutre	-6,44	-3,41	-5,79	-13,37
Peur	-1,04	9,82	-0,14	-3,98
Colère	1,65	9,73	2,23	0,51
Joie	-0,27	9,88	2,71	0,44
Tristesse	3,13	14,62	-2,72	-3,22
Dégoût	-2,52	10,82	-1,75	-1,12
Surprise	-2,00	6,32	-1,70	-2,80

f) *Moyennes brutes des pourcentages de non-voisement (%) obtenus chez chaque participant et pour chaque émotion*

	F01	F07	H02	H08
Neutre	28,85	34,31	34,06	44,44
Peur	26,84	29,20	23,47	38,44
Colère	24,18	33,58	28,15	39,40
Joie	26,03	36,62	25,09	37,98
Tristesse	26,11	44,25	43,91	46,57
Dégoût	31,98	37,14	31,64	43,27
Surprise	33,84	39,97	29,28	41,33

g) *Moyennes brutes des valeurs obtenues pour le CPPs (dB) chez chaque participant et pour chaque émotion*

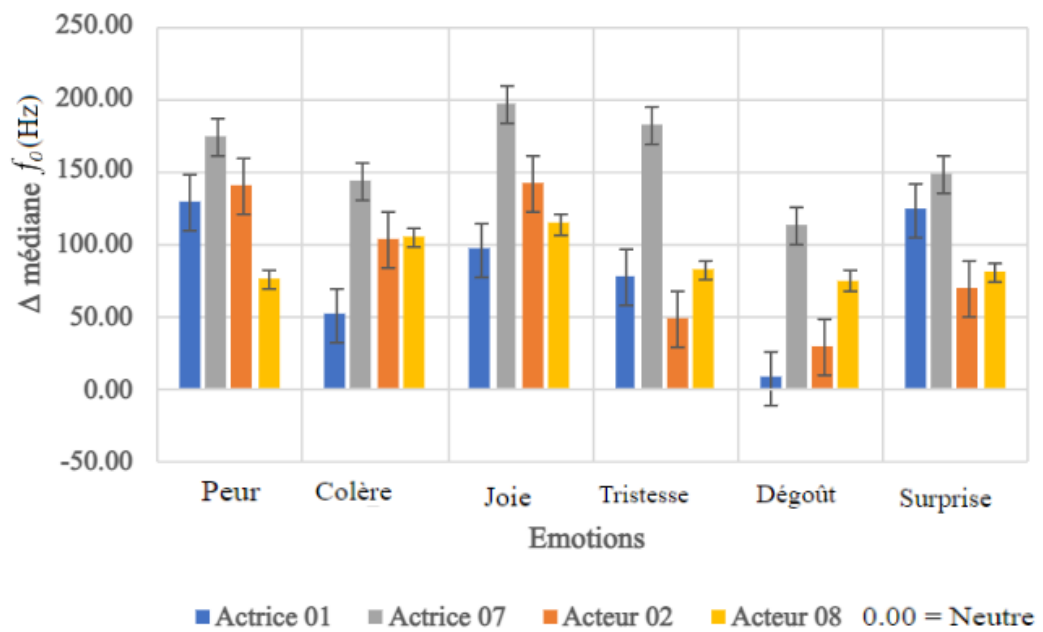
	F01	F07	H02	H08
Neutre	5,59	6,82	8,25	6,19
Peur	5,01	5,44	6,89	6,06
Colère	6,73	5,85	7,26	5,82
Joie	5,48	5,16	7,13	6,26
Tristesse	6,42	5,25	5,83	6,12
Dégoût	5,86	6,24	7,67	6,20
Surprise	4,77	5,05	7,37	5,61

h) Moyennes brutes des valeurs du débit (nombre de syllabes/seconde) obtenues chez chaque participant et pour chaque émotion

	<i>F01</i>	<i>F07</i>	<i>H02</i>	<i>H08</i>
Neutre	6,79	4,64	5,45	4,15
Peur	6,07	4,29	5,39	4,68
Colère	5,80	3,77	4,48	4,54
Joie	5,27	4,00	4,44	4,47
Tristesse	5,04	3,06	4,49	3,91
Dégoût	4,46	3,54	4,79	3,97
Surprise	5,57	3,55	5,17	4,37

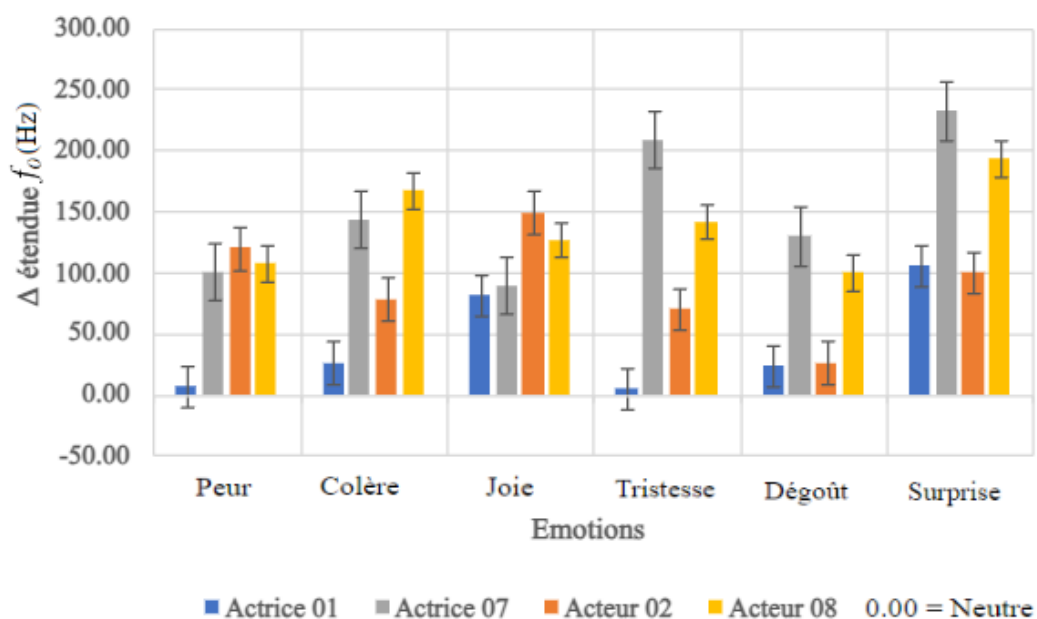
Annexe 7. Graphiques reprenant les analyses delta des moyennes obtenues pour chaque acteur de l'échantillon, dans chaque condition émotionnelle par rapport à la condition neutre, et ce, pour chaque paramètre acoustique étudié.

a) Delta émotions/neutre de la médiane moyenne de f_0 pour chaque émotion



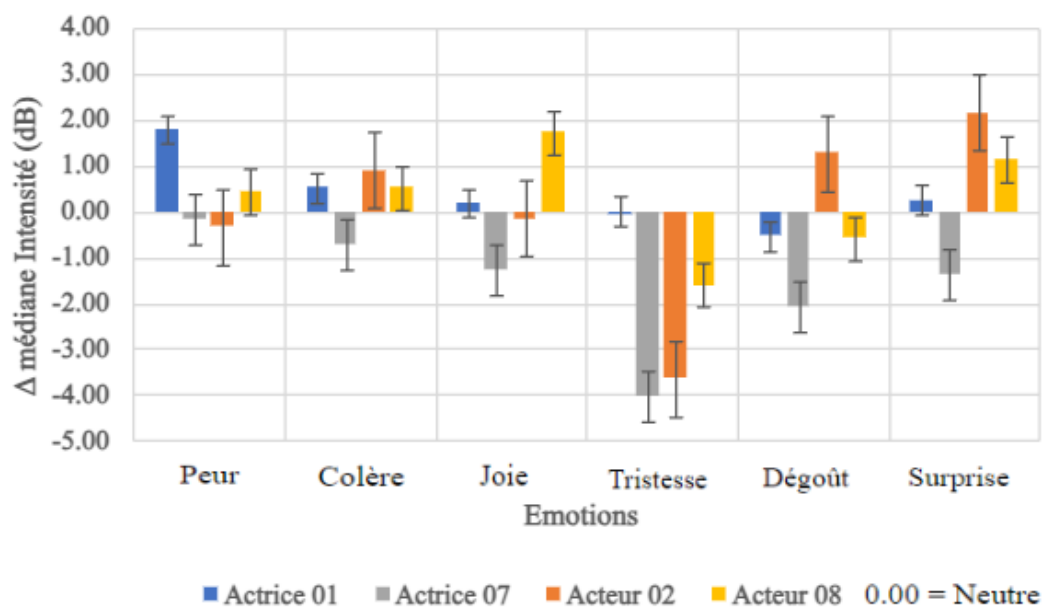
Note. Barre verticale I = erreur standard

b) Delta émotions/neutre de l'étendue moyenne de f_0 pour chaque émotion



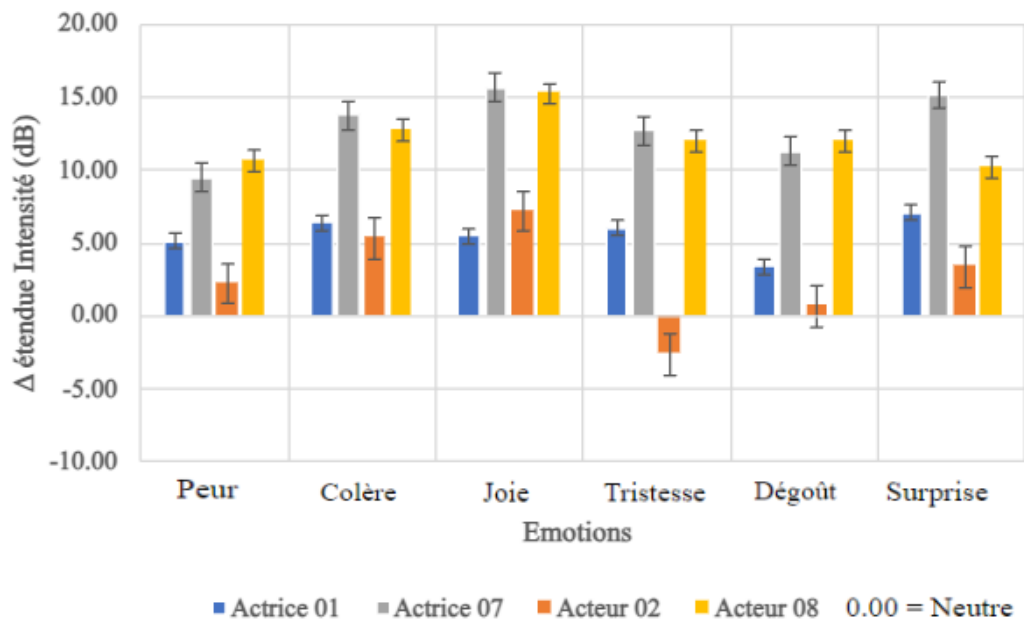
Note. Barre verticale I = erreur standard

c) Delta émotions/neutre de la médiane moyenne de l'intensité pour chaque émotion



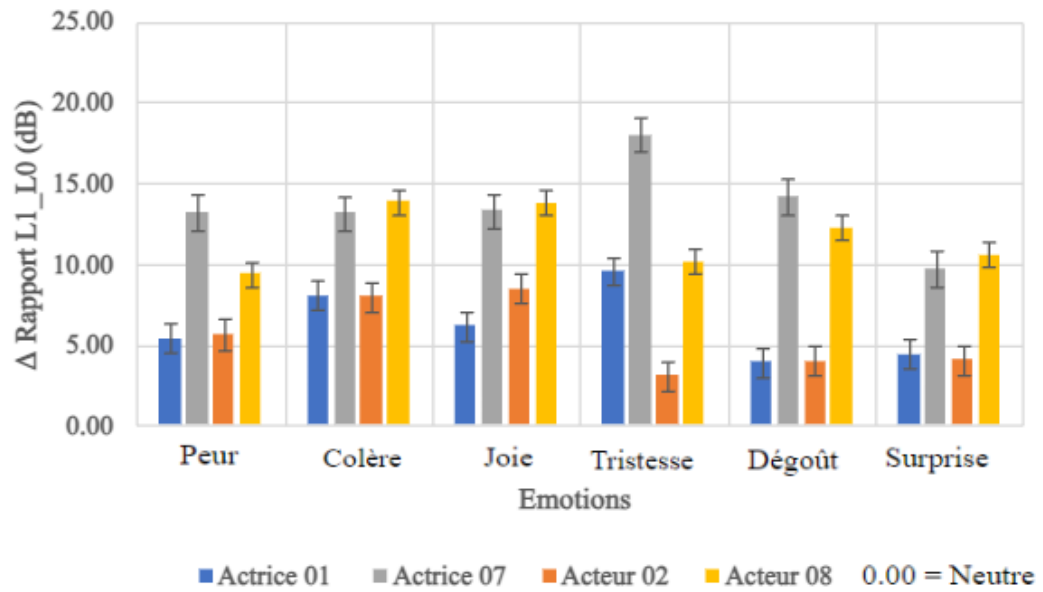
Note. Barre verticale I = erreur standard

d) Delta émotions/neutre d'étendue moyenne de l'intensité pour chaque émotion



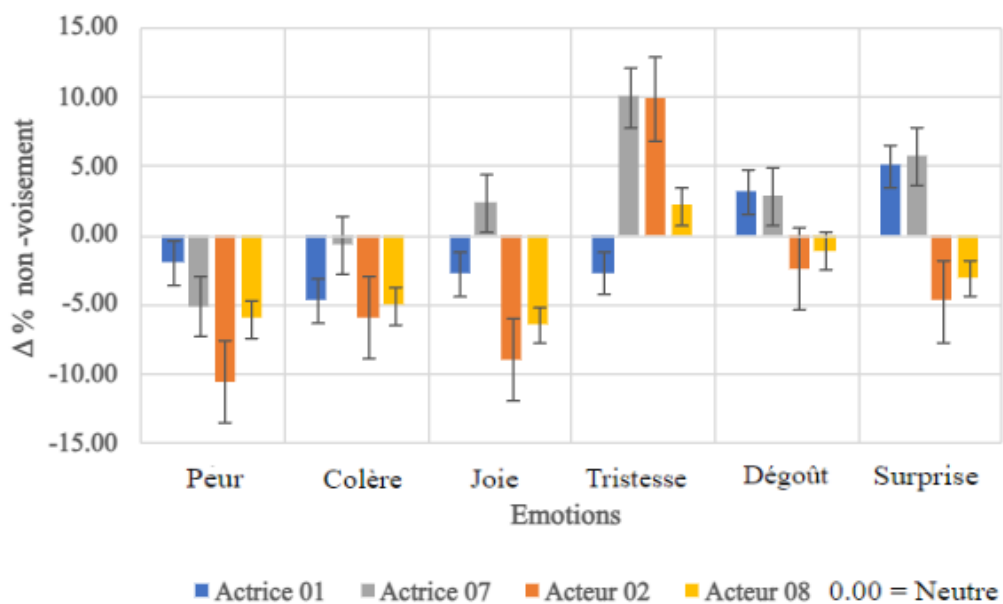
Note. Barre verticale I = erreur standard

e) Delta émotions/neutre du rapport L1-L0 moyen du LTAS pour chaque émotion



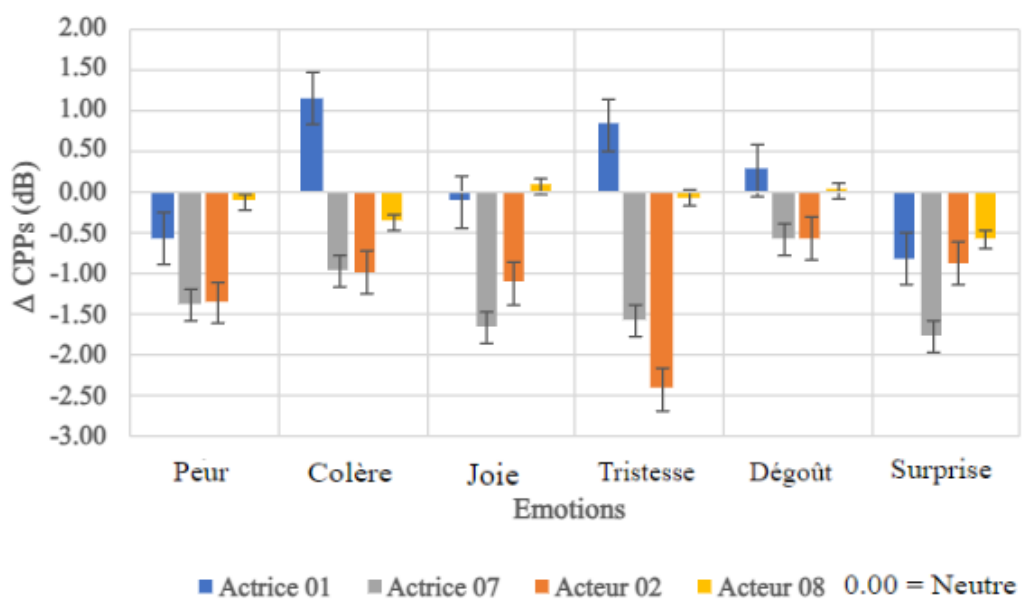
Note. Barre verticale I = erreur standard

f) Delta émotions/neutre du pourcentage de non-voisement moyen du LTAS pour chaque émotion



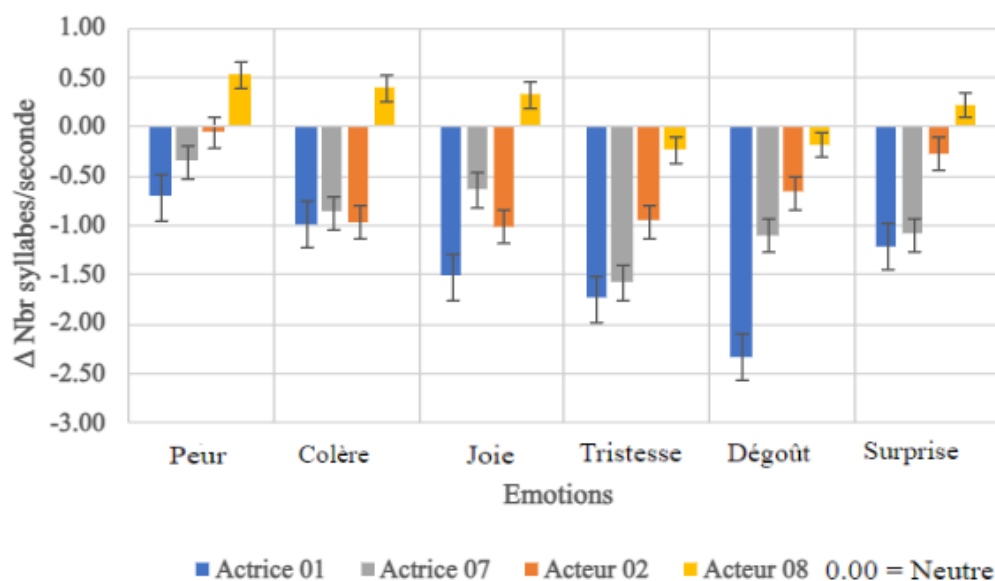
Note. Barre verticale I = erreur standard

g) Delta émotions/neutre du CPPs moyen pour chaque émotion



Note. Barre verticale I = erreur standard

h) *Delta émotions/neutre du débit moyen pour chaque émotion*



Note. Barre verticale I = erreur standard

Annexe 8. Explications de notre implication respective pour réaliser ce mémoire

Il est important de noter que lors de la rédaction de ce mémoire, nous nous sommes autant investies l'une et l'autre pour contribuer à chaque section présentée dans celui-ci (50% Ella et 50% Marie). En effet, notre organisation de travail durant toute cette année académique s'est déroulée comme suit : nous nous fixions des appels téléphoniques régulièrement (environ 1 fois par mois) pour débriefer du travail accompli, se poser des questions et planifier ce qu'il nous restait à produire durant les semaines suivantes. Dès que l'une d'entre nous consultait des ressources bibliographiques intéressantes ou progressait dans ce travail, l'autre consultait à son tour la littérature, commentait, développait ou corrigeait le travail qui avait été proposé par le binôme, afin de s'approprier également le travail accompli. Ainsi, l'ensemble de ce projet est le fruit d'une collaboration rigoureuse et complémentaire entre nous.

RÉSUMÉ

L'impact de l'expression des émotions sur l'évolution des paramètres de la voix et de la parole est un sujet qui intéresse les chercheurs depuis des siècles. Dès lors, il est clairement établi dans la littérature scientifique traitant ce sujet que nos différents états émotionnels sont associés à des manifestations spécifiques des paramètres de la voix et de la parole. Ainsi, notre mémoire s'inscrit dans la continuité des travaux qui ont été réalisés dans ce domaine de recherche.

Notre mémoire contribue tout d'abord à la création d'une base de données audio-visuelle des émotions appelée « EVE », de laquelle nous réalisons ensuite l'étude des données audio. En effet, notre projet consiste à étudier, sur base d'enregistrements, quelle est l'influence de la production des 6 émotions universelles de Paul Ekman (peur, colère, joie, tristesse, dégoût, surprise) sur les paramètres de la voix et de la parole d'un échantillon de 4 acteurs francophones (2 femmes et 2 hommes) âgés de 24 à 57 ans. Les paramètres acoustiques étudiés dans ce travail sont les suivants : f_0 (médiane et étendue), intensité (médiane et étendue), LTAS (rapport L1-L0 et % non-voisement), CPPs, débit, vitesse articulatoire.

Durant les séances d'enregistrement, les acteurs étaient invités à produire oralement les 10 phrases de la liste 1 du corpus de Fharvard (Aubanel & al., 2020). Il leur était demandé d'interpréter, pour chaque phrase, les 6 émotions d'Ekman selon 2 intensités émotionnelles et une condition contrôle. L'acteur avait droit à deux essais, mais seules les phrases produites à haute intensité émotionnelle lors du second essai ont été traitées. Les 280 énoncés ainsi obtenus ont ensuite été analysés via le logiciel PRAAT (version 6.4.07 – March 17, 2024).

Enfin, nos données ont été interprétées via des analyses de deltas (Δ) entre les conditions contrôles et émotionnelles. Ainsi, pour notre échantillon, les résultats obtenus suggèrent que les 6 émotions d'Ekman ont été, en moyenne, caractérisées par une augmentation de la hauteur tonale et de sa variabilité, une voix plus puissante et moins soufflée, davantage de variabilité intonative d'intensité ($\Delta+$), moins de régularité dans le signal acoustique produit, et un ralentissement du débit et de la vitesse articulatoire ($\Delta-$), par rapport à la condition contrôle. Les autres données obtenues permettent de nuancer et d'enrichir les informations dont nous disposons pour chaque émotion étudiée et de les confronter avec nos hypothèses exploratoires et confirmatoires afin d'ouvrir de nouvelles perspectives de recherche dans ce domaine.