

École Nationale Supérieure Louis-Lumière

La perception de l'identité de genre
dans la voix enfantine en langue
française : étude du rôle de la fréquence
fondamentale

Mémoire de Master 2 - Spécialité Son

Autrice :

Lou JULLIEN

Directeur de mémoire interne :

Laurent MILLOT

Directeur de mémoire externe :

Erwan PÉPIOT

Responsable universitaire :

Corsin VOGEL

Rapporteur :

Éric URBAIN

1^{er} juin 2022

Remerciements

Je tiens à remercier tout le personnel de l'école du Chevalier Saint-Juste pour leur accueil. J'adresse mes sincères remerciements aux enfants qui ont prêté leur voix ainsi qu'à leurs parents et aux volontaires qui ont participé au test perceptif, sans qui ce travail n'aurait pas pu voir le jour.

Je suis particulièrement reconnaissant envers Philémon, Garance, Solenn, Valentin, Hugo, Line, Erwan Kerzanet et à l'équipe du service IPT de Paris 8 pour leur aide lors des enregistrements.

Je remercie Laurent Millot et Corsin Vogel pour leur encadrement pédagogique tout au long de l'année ainsi que Florent Fajole pour son aide et sa présence silencieuse.

Merci à Erwan Pépiot pour ses précieux conseils, son enthousiasme et sa bienveillance.

Un grand merci à tou·te·s mes camarades de la promotion Son 2022 pour ces trois précieuses années passées à leurs côtés. Merci aussi à la promotion Son 2023 pour les moments de légèreté partagés entre les séances de travail à Louis Lumière.

Merci à Enora, Jeanne, Maeva et à tou·te·s mes partenaires de bibliothèques et de cafés. Merci à Hugo pour son aide précieuse et sa tendresse.

Merci, enfin, à Valérie pour ses encouragements, sa patience et son soutien constant.

Résumé

L'élaboration, dans l'imaginaire collectif, de représentations de ce que sont les « voix de garçon » et les « voix de fille », par analogie avec les voix adultes, relèvent de problématiques complexes et pluridisciplinaires. Nous avons cherché à étudier dans ce mémoire les tenants et les aboutissants phonétiques et psychoacoustiques de cette question. Nous y abordons les différences intergenres dans les voix enfantines en langue française et posons la question du processus d'identification du genre des enfants par l'écoute de leur voix. Le premier chapitre est consacré à un état des lieux de la recherche sur l'expression et la perception du genre à travers la voix.

Le deuxième chapitre décrit la collecte et l'analyse d'un corpus de voix d'enfants, nommé CoVEP, composé de 10 filles et 8 garçons francilien-ne-s âgé-e-s de 8 à 10 ans. Le corpus est constitué d'une série de 33 pseudo-mots dissyllabiques et de 12 phrases lues avec un mode de phonation normal et avec une voix chuchotée ; ainsi que d'un extrait de discours semi-spontané d'une trentaine de secondes avec un mode de voix normal. Nous présentons l'analyse des F0 des enfants sur l'ensemble du corpus. Les garçons présentent en moyenne des F0 plus basses que celles des filles, cette différence ne s'est avérée significative que pour l'extrait du discours semi-spontané des enfants.

Une partie des enregistrements effectués a été utilisée dans le cadre d'un test perceptif. 44 adultes ont été chargé-e-s d'identifier le genre de 8 des filles et 8 des garçons enregistré-e-s à partir d'une sélection de pseudo-mots dissyllabiques et de phrases voisés et chuchotés. La confrontation de stimuli identiques, voisés et chuchotés nous a permis d'évaluer l'impact de la F0 sur les réponses données par les participant-e-s. Les auditeur-ice-s ont produit un nombre de bonnes réponses supérieur à la chance pour tous les types de stimuli. Les phrases ont permis de mieux identifier le genre que les stimuli courts. Les stimuli voisés ont été mieux identifiés en moyenne que les stimuli chuchotés ce qui indique que la F0 joue un rôle dans la perception du genre des voix pré-pubères. Nos résultats laissent penser qu'il existe bel et bien une différence de F0 en fonction du genre en langue française, dès 8 ans. D'après notre étude bibliographique, ces résultats ne s'expliquent pas par une différence morphologique de l'appareil vocal avant la puberté, l'explication de cette observation est probablement d'ordre culturel. Enfin, il semble que les adultes sont capables de reconnaître le genre des enfants à la seule écoute d'une phrase mais le rôle de la F0 et des paramètres acoustiques utilisés pour identifier le genre restent à déterminer. Mots-clés : Genre, voix d'enfant, perception du genre dans la voix, traitement de la parole, socio-phonétique, psychoacoustique

Abstract

The collective representations of what "boy voices" and "girl voices" are, by analogy with adult voices, raise complex and multidisciplinary issues. In this thesis we have sought to explore the phonetic and psycho-acoustic implications of the matter. We address the gendered differences in French children's voices and seek to understand how one is able to identify the gender of children only by listening to their voices. The first chapter is devoted to an overview of research on the expression and perception of gender through the voice.

The second chapter describes the collection and analysis of a corpus of children's voices, named CoVEP, made up of 10 girls and 8 boys aged 8 to 10 from the Ile-de-France region in France. The corpus consists of a series of 33 disyllabic pseudowords and 12 sentences read with both normal and whispered voices; as well as a 30-second semi-spontaneous speech extract with a normal voice. We present the analysis of the children's F0 over the whole corpus. The boys' F0s were lower on average than those of the girls, but this difference was only significant for the semi-spontaneous speech extract.

Part of the recordings was used in a perceptual test. 44 adults were asked to identify the gender of 8 of the recorded girls and 8 of the recorded boys from a selection of disyllabic pseudowords and sentences both voiced and whispered. The confrontation of identical voiced and whispered stimuli allowed us to evaluate the impact of F0 on the responses given by the participants. The listeners produced a number of correct responses higher than chance for all types of stimuli. Sentences were better identified than short stimuli. Voiced stimuli were better identified on average than whispered stimuli, indicating that F0 must play a role in the perception of the gender in pre-pubescent voices.

Our results suggest that there is indeed a difference in F0 as a function of gender in French as early as 8 years old. According to our literature review, these results cannot be explained by a morphological difference in the vocal apparatus before puberty. The explanation for this observation may be cultural. Finally, it seems that adults are able to recognise the gender of children by listening to a sentence, but the role of F0 and other acoustic parameters used to identify gender remain largely to be determined.

Keywords : Gender, child voice, gender perception, gendered acoustic differences, voice perception, speech processing, French, phonetics, socio-phonetics, psychoacoustics

Table des matières

Résumé	iii
Abstract	iv
Introduction	1
1 Voix et genre, état de la recherche	5
1.1 Définition de la notion de genre	5
1.1.1 Sexe et genre	5
1.1.2 Quelques approches du genre	6
1.2 Le genre dans la voix adulte	7
1.2.1 Quelques critères acoustiques	7
1.2.2 Perception du genre dans la voix	14
1.2.3 Les différences inter-langues des indices de genre	17
1.2.4 Le cas de la voix chuchotée	20
1.3 Physiologie de la production de la voix avant la puberté	24
1.4 La F0 en fonction du genre chez les enfants	26
1.4.1 État de la question	26
1.4.2 Influence de l'intensité	29
1.5 Les formants en fonction du genre chez les enfants	31
1.6 Autres paramètres acoustiques des voix d'enfants	35
1.7 La perception du genre des enfants par leur voix	36

1.7.1	État des lieux	36
1.7.2	Les indices utilisés pour identifier le genre	38
1.7.3	Rôle de l'âge	42
1.7.4	Le degré de certitude	43
1.8	La conscience des rôles de genre et la perception de la voix par rapport au genre chez les enfants	45
2	Enregistrement et analyse acoustique préliminaire d'un corpus de voix d'enfants	49
2.1	Méthodologie	51
2.1.1	Enfants enregistré·e-s	51
2.1.2	Matériau linguistique enregistré	52
2.1.3	Matériel et procédure d'enregistrement	54
2.1.4	Réduction du corpus pour les analyses acoustiques	54
2.1.5	Méthodologie des analyses acoustiques	55
2.2	Analyse des enregistrements	57
2.2.1	Présentation des résultats	57
2.2.2	Discussion	59
3	Test perceptif autour d'une partie du corpus enregistré	63
3.1	Procédure expérimentale	63
3.1.1	Participant·e-s	63
3.1.2	Matériau linguistique sélectionné	63
3.1.3	Déroulement du test	65
3.1.4	Méthodologie de l'analyse des résultats	66
3.2	Analyses des résultats du test perceptif	67
3.2.1	Présentation des résultats	67
3.2.2	Discussion	70
	Conclusion	79
	Annexes	83
	Annexe 1 : Schéma du système phonatoire supérieur	83
	Annexe 2 : Corpus lu	84
	Annexe 3 : Template du questionnaire de profil pour les enfants	87
	Annexe 4 : Détails des scores obtenus par stimulus	89

Table des figures	91
Lexique	97
Bibliographie	99

Introduction

À l'écoute d'une voix, nous sommes capables de déduire un certain nombre de caractéristiques concernant notre interlocuteur·ice. C'est ce qu'on appelle le phénomène d'acquisition d'informations extralinguistiques que Simpson décrit ainsi¹ :

« As soon as we hear a new voice we do much more than just understanding the message it contains. We make judgements about how old somebody is or where they come from. We also try to ascertain whether we are listening to a man or a woman. Often we are only conscious that we are trying to make this judgement when we realise we have made an incorrect assignment or are confronted with an ambiguous voice. » (Simpson 2009, p. 621)

Ce processus correspond à la nécessité de caractériser son interlocuteur·ice afin d'adopter un comportement approprié. Même en n'ayant accès uniquement à la voix de notre interlocuteur·ice (au téléphone par exemple), chacun·e doit être capable de déterminer son genre afin d'adopter les pronoms correspondants, ainsi qu'une tranche d'âge afin de choisir entre le vouvoiement et le tutoiement, mais aussi un niveau de langue approprié et toutes les subtilités de communication induite par le rapport social à l'œuvre. Comme le souligne Simpson, cette tâche est bien souvent automatique.

Le cerveau identifie avec une grande fiabilité le genre en particulier. Nous sommes capable de reconnaître le genre d'une personne grâce à sa voix et ce même à l'écoute d'un

1. « Dès que nous entendons une nouvelle voix, nous faisons bien plus que simplement comprendre le message qu'elle contient. Nous portons des jugements sur l'âge de la personne ou sur ses origines. Nous essayons également de déterminer si nous écoutons un homme ou une femme. Souvent, nous ne sommes conscients que nous essayons de porter ce jugement que lorsque nous réalisons que nous avons fait une attribution incorrecte ou que nous sommes confronté·e·s à une voix ambiguë. » (Traduction de l'autrice)

INTRODUCTION

énoncé de quelques dizaines de millisecondes (S. Whiteside 1998). Pourtant, il est difficile de s'accorder sur ce qui fait une voix de femme ou une voix d'homme et *a fortiori* de déterminer quelles sont les origines des différences entre les deux. Parmi les spécialistes de la question, deux points de vue s'affrontent : l'essentialisme et le constructivisme. Selon le point de vue essentialiste : les caractéristiques vocales de la voix découlent uniquement d'attributs physiques, elles sont innées. D'après ce point de vue, les différences entre une voix d'homme et une voix de femme sont la conséquence des différences physiologiques entre leurs appareils vocaux. Tandis que selon les constructivistes, la façon de poser sa voix correspond également à des normes culturelles et sociales qui s'apprennent, elle est en partie acquise. D'après cette théorie, les différences entre une voix d'homme et une voix de femme sont représentatives de la différence entre les rôles sociaux assignés à chaque genre dans un contexte culturel donné.

Ce questionnement nous pousse à nous intéresser à la voix des enfants. Existe-il des différences inter-genres dans la voix avant la puberté ? Si oui, ces différences sont-elles dues à des différences physiologiques ou à des stratégies articulatoires acquises par mimétisme pour se conformer à une identité de genre assimilée ? Les différences entre les voix des filles et des garçons avant la puberté sont-elles équivalentes à celles existant entre les hommes et les femmes pubèr·e-s ?

Il nous paraît essentiel de mettre ces questionnements en regard de la perception du genre dans la voix des enfants. Dans l'imaginaire collectif, l'enfant est souvent considéré comme un profil ambigu par rapport au genre. En langue française, la dichotomie du genre est moins primordiale en ce qui concerne les enfants (le terme « les enfants » est neutre alors que pour les adultes, les termes usuels sont « les hommes et les femmes »). Dès lors on peut se demander si les voix d'enfants s'affranchissent de la notion de genre. Si des différences acoustiques existent entre les voix des filles et des garçons avant la puberté, sont-elles perçues par les auditeur·ice-s ? Le cerveau discrimine-t-il les voix d'enfants en fonction du genre de la même manière qu'il le fait avec les voix d'adultes ?

Dans ce mémoire, nous commencerons par faire un état des lieux de la recherche sur l'expression et l'identification du genre par la voix avant la puberté, à la lumière des connaissances, plus approfondies, disponibles sur les caractéristiques des voix d'hommes et de femmes adultes. Nous chercherons à dégager les principales différences acoustiques entre les voix de garçons et de filles, leur potentielles origines et comment celles-ci sont interprétées par le cerveau. Dans un second temps, nous présenterons les résultats d'analyses acoustiques d'un corpus de voix d'enfants francilien·ne-s âgé·e-s de 8 à 10 ans, enregistré à l'occasion de ce travail. Nous avons utilisé les enregistrements réalisés dans le cadre

d'une expérience d'identification du genre par la voix. Nous présentons ici l'analyse des résultats de cette expérience, au regard de quelques caractéristiques spectrales élémentaires des voix du corpus. Cette étude constitue une première approche sur la perception du genre dans la voix enfantine en langue française. Il cherche à donner une meilleure compréhension du processus cognitif à l'œuvre dans l'identification du genre à l'écoute de la voix pré-pubère et de l'élaboration des représentations des voix de garçons et de filles dans l'imaginaire collectif. Les résultats préliminaires présentés s'inscrivent dans la lignée de recherches autour de la question de l'expression et de la perception du genre dans la voix et apportent des éléments de réponse à la question de la part d'essentialisme et de constructivisme dans les différences observées entre les voix féminines et masculines.

1.1 Définition de la notion de genre

1.1.1 Sexe et genre

L'Organisation Mondiale de la Santé (OMS) définit le genre comme « les rôles qui, selon la représentation que s'en fait la société, déterminent les comportements, les activités, les attentes et les chances considérés comme adéquats pour tout un chacun dans un contexte socio-culturel donné. Ce terme désigne aussi les relations entre personnes et la répartition du pouvoir dans ces relations. Le genre a un lien avec le sexe biologique (masculin ou féminin), mais il est distinct de lui.»¹ Le sexe d'un individu est défini par un ensemble d'attributs physiques (anatomiques et physiologiques) comme les chromosomes, les hormones, les gonades ou encore les organes génitaux. D'après Fausto-Sterling (2000), le sexe n'est pas un concept binaire (contrairement à ce que laisse entendre la définition de l'OMS), l'autrice introduit la notion de « continuum sexuel ». Laqueur (1990) constate que la compréhension du sexe comme une notion binaire est relativement récente. Il explique que la conception de deux sexes fondamentalement différents et donc incomparables, ainsi que l'absence de considération pour les corps qui se situent « entre-deux » a pris son essor à la fin du *XVIII^e* siècle. Butler (1990), elle, affirme que ce que l'on appelle *sexe* est une

1. <https://www.who.int/fr/news-room/fact-sheets/detail/gender>

construction culturelle au même titre que le genre.

La notion de genre s'est construite en opposition au sexe biologique, comme un « sexe social » (Oakley 1972) ; pourtant la question de savoir si le sexe et le genre sont deux notions distinctes ou si elles se confondent fait débat. Certains, comme Kessler et al. (1978), considèrent que le genre englobe les questions liées au corps et que la distinction entre les deux notions n'a pas lieu d'être. On peut donc dire, suivant en cela l'OMS, que le sexe et le genre sont deux notions qui sont liées et se recoupent sans s'opposer et qui n'entretiennent pas de lien de causalité directe.

Dans cette étude, nous faisons le choix d'utiliser la notion de genre. Le sexe nous semble une notion trop restreinte pour étudier les enjeux de l'identité dans la voix et, comme nous le détaillerons dans la partie **I.3**, un prisme peu pertinent pour aborder la production vocale avant la puberté.

1.1.2 Quelques approches du genre

Depuis l'émergence du terme *genre*, sa définition a évolué et a fait l'objet de plusieurs théories. Citons par exemple Kessler et al. (1978) qui se sont intéressées à la présentation de genre de profils « ambigus » (des profils transsexuels en l'occurrence). Arnold résume ainsi leurs travaux : « Elles étudient quatre aspects de la présentation de soi : la façon de parler, l'apparence physique publique, le corps privé et la biographie. Cette présentation de soi constitue une dimension de ce que Kessler et McKenna appellent le « travail de genre » (gender work) (ibid., 1978, p. 136). Toutefois, elles ne pensent pas que ce travail de genre est réalisé uniquement par la personne qui se présente. Elles soutiennent que la majeure partie du travail de genre est effectuée par la personne à qui la présentation est destinée (ibid., 1978, p. 136). C'est la présentation de soi qui va dans un premier temps déterminer l'attribution de genre, mais une fois que celle-ci est faite, c'est le destinataire qui va être responsable de son maintien. » (Arnold 2015, p. 37-38). Le genre est donc une notion qui ne peut pas s'étudier sans s'attarder sur la manière dont il est perçu par l'autre. Selon Kessler et McKenna, c'est la perception qui est le processus actif de la construction d'une représentation culturelle d'un genre.

Judith Butler, elle, développe la théorie de la performativité. Elle reprend la notion linguistique introduite par Austin (1962) d'un « énoncé performatif » qui fait advenir ce qu'il énonce, en opposition à un énoncé constatif. Selon Butler, le genre n'est pas le constat d'une réalité autrement dit l'expression de genre n'est pas une causalité du corps.

Elle dit¹ :

« *We act and talk and speak and walk in ways that consolidate an impression of being a man or being a woman.* »

Le genre performé est identifiable grâce à ce que Butler appelle la « citationnalité ». Arnold explique cette notion ainsi : « pour qu'un acte performatif de genre soit efficace, il doit être identifiable – il doit faire « écho à des actions antérieures » [Butler, 1997 [2004, p. 92]] qui par leur répétition se sont sédimentées en norme. Ainsi, à travers la citation, tout acte performatif de genre constitue un assujettissement à une norme de genre. » (ibid., p. 32)

Retenons comme définition du genre qu'il est une identité s'accomplissant à travers un ensemble de pratiques sociales et qui peut évoluer au cours de la vie d'un individu. Comme nous avons pu le voir, la façon de parler est régulièrement citée comme un indice employé pour indiquer son genre. Demandons-nous maintenant quels sont les attributs genrés de la voix et comment la façon dont nous posons notre voix renseigne nos interlocuteur·ice·s sur notre genre.

1.2 Le genre dans la voix adulte

1.2.1 Quelques critères acoustiques

La fréquence fondamentale

La valeur de la fréquence fondamentale moyenne de la voix, fréquence que nous nommerons F0 dans la suite, est considérée comme la principale différence entre les voix d'hommes et de femmes. Elle est directement liée à la hauteur perçue de la voix et elle dépend de la taille des plis vocaux² : toutes choses égales par ailleurs, plus les plis vocaux sont longs et épais, plus la F0 sera basse, plus la voix sera perçue comme grave, et inversement. La F0 est donc directement liée à la taille du locuteur mais également à son genre. Pendant la puberté « les cartilages laryngés s'agrandissent et les plis vocaux s'allongent et s'épaississent » (ibid., p. 68-69). Ce phénomène étant plus marqué chez les garçons

1. « Nous agissons, parlons et marchons d'une manière qui consolide l'impression de ce qu'est être un homme ou une femme. » (Traduction de l'autrice). Ces propos sont tirés d'une interview réalisée par Max Miller, 2011, <https://bigthink.com/videos/your-behavior-creates-your-gender/>

2. Les plis vocaux, appelées communément *cordes vocales*, sont deux replis de membrane muqueuse qui se situent au niveau de la cavité intérieure du larynx (voir la figure 23, en annexe de ce document). Leurs mise en vibration est à l'origine des sons voisés qui composent la parole.

que chez les filles, les voix d'hommes sont en moyenne plus graves que celle des femmes après la puberté. Toutefois, les pratiques articulatoires permettent de faire varier la F0, la valeur de celle-ci dépend donc également de la façon dont le·a locuteur·ice module sa voix. Une voix grave, avec une F0 basse, sera associée à une voix d'homme et une voix aiguë, avec une F0 élevée, sera associée à une voix de femme (Arnold 2015). Notons également qu'après la fin de la puberté, la F0 continue d'évoluer avec l'âge tout au long de la vie (Pegoraro-Krook 1988 et Traunmüller et al. 1994).

De nombreuses études se penchent sur l'écart entre les F0 moyennes entre les genres dans différentes langues. Chacune de ces études pointe que la valeur moyenne de la F0 se différencie significativement en fonction du genre. Néanmoins les différences observées varient significativement en fonction de la langue étudiée comme le résume Pépiot :

« Dans une recherche portant sur des locuteurs de l'anglais américain (24 hommes et 24 femmes), Takefuta et al. (1972) trouvent une fréquence fondamentale moyenne de 117 Hz pour les hommes et de 186 Hz pour les femmes, soit un ratio d'environ 1,59. Boë et al. (1975), avec une étude réalisée sur 60 locuteurs francophones, arrivent quant à eux aux moyennes de 118 Hz et 207 Hz, c'est-à-dire un ratio d'environ 1,75. Dans une étude menée sur près de 200 locuteurs suédois, Kitzing (1979) avance les chiffres de 110 Hz pour les hommes et de 193 Hz pour les femmes (correspondant à un ratio d'environ 1,75), des résultats très proches de ceux trouvés plus récemment par Pegoraro-Krook (1988) pour la même langue, avec une étude portant sur plus de 600 locuteurs. [...] Chen, dans une étude conduite en 1974 sur des locuteurs du mandarin standard, arrive à une moyenne de 108 Hz pour les hommes et 184 Hz pour les femmes, des résultats proches de ceux de Hwa Chen (2007) sur des taiwanais locuteurs du mandarin (110 Hz contre 196 Hz), alors que Rose (1991), avec des locuteurs d'un dialecte Wu (quatre hommes et trois femmes), avance les chiffres de 170 Hz pour les hommes contre 187 Hz pour les femmes. » (Pépiot 2013, p. 27)

Il semble donc que l'explication physiologique se couple à une pratique culturelle assimilée puisque la hauteur moyenne des voix d'hommes et de femmes dépend de la langue étudiée. C'est ce qu'explique Simpson¹ :

1. « Lorsque l'on compare les différences moyennes de F0 entre les femmes et les hommes dans différentes langues, il devient vite évident que même la fréquence fondamentale moyenne de la voix est en partie apprise. [...] Comme il serait déraisonnable de rendre compte de différences aussi importantes en termes de différences anatomiques dans les populations étudiées, une partie de la différence doit être attribuée à des comportements appris. » (Traduction de l'autrice)

« *When we compare average female-male F0 differences from a range of languages, it soon becomes clear that even average fundamental frequency of the voice is in part learned. [...] Since it would be unreasonable to account for such large differences in terms of anatomical differences in the populations being investigated, part of the difference must be attributed to learned behaviours.* »
(Simpson 2009, p. 625)

Les formants

Un autre paramètre prépondérant dans la discrimination du genre est la position des formants. On appelle formant, les fréquences de résonance des voyelles. On étudie en général les trois voire quatre premiers formants que l'on note F1, F2, F3 et F4. Ces zones renforcées du spectre sont notamment ce qui permet de différencier les voyelles les unes des autres. On peut expliquer la production de la parole selon un modèle source/filtre. Lors de la production d'une voyelle, le larynx (la source) entre en vibration, c'est ce qui définit la hauteur du son. Les articulateurs (comme les lèvres, la mâchoire inférieure, la langue ou le voile du palais) s'actionnent de sorte à ouvrir, fermer ou déformer des cavités de l'appareil vocal qu'on appelle résonateurs ; ce sont par exemple les cavités buccale, nasale, pharyngale, laryngale ou encore labiale. Ce système de résonateurs agit comme un filtre sur le son produit, il modifie son timbre. En fonction des résonateurs en action et de leur dimension, quelques fréquences sont excitées et ainsi certaines zones spectrales du voisement se trouvent renforcées. Les différentes combinaisons de résonateurs possibles engendrent les différentes voyelles d'une langue. Traditionnellement en phonétique, on les illustre sous la forme de ce qu'on appelle le triangle vocalique, dont chaque sommet représente un paramètre d'articulation : l'ouverture, c'est-à-dire la distance entre la mâchoire inférieure et le palais ; l'antériorité définie par la position de la langue dans la bouche (elle se place dans l'arrière de la bouche pour les voyelles antérieures comme [i] et en avant pour les voyelles postérieures comme [o]) et enfin l'arrondissement des lèvres (maximal pour la voyelle [u :] par exemple). Chacun de ces trois modes d'articulation a une action sur les zones formantiques de la voyelle produite. En phonétique acoustique, on trace donc le triangle formantique qui indique les fréquences centrales des deux premiers formants de chaque voyelle (Calliope 1989). Arnold (2015), s'appuyant sur les données de l'ouvrage de référence *La Parole et son traitement automatique*, trace le triangle formantique des voyelles de la langue française (voir la figure 1).

On constate que la différence entre les voix d'hommes et de femmes est plus marquée pour les voyelles ouvertes (comme [a]) et antérieures (comme [i]), en résulte un triangle

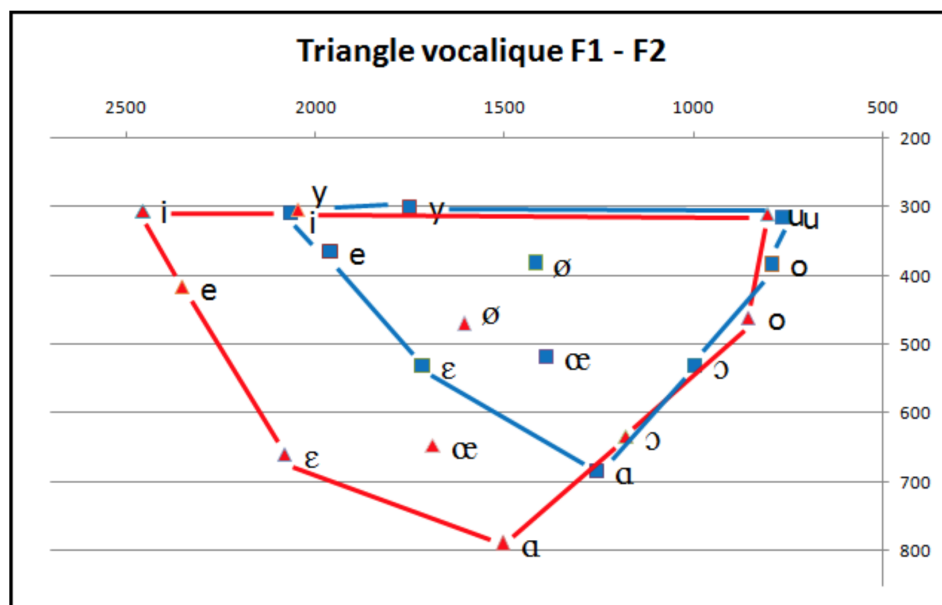


FIGURE 1 – Triangle vocalique représentant les valeurs fréquentielles moyennes (en Hz) de F1 (en ordonnée) et F2 (en abscisse) des voix de femmes (triangles rouges) et d’hommes (carrés bleus) à partir des données de Calliope (1989) (Arnold 2015)

formantique plus étendu chez les femmes. Cette observation est valable également en langue anglaise (Trautmüller 1988). Pépiot souligne que cette disparité reste significative « même en passant d’une échelle linéaire (en Hertz) à une échelle non-linéaire plus proche de la perception humaine (en barks ou demi-tons, par exemple), ce qui ne permet de gommer que très légèrement ces disparités. » (Pépiot 2013, p. 37)

Plusieurs explications ont été avancées pour expliquer ces disparités. En premier lieu, la différence de la taille du conduit vocal, « de manière schématique, plus une cavité est longue, plus les fréquences de résonance seront basses (voir notamment Stevens, 1955, 1999, et Fant, 1970). Ainsi, pour une position articuloire équivalente, les formants des hommes tendent à être globalement situés dans des fréquences plus basses que ceux des locutrices (voir Turner et al., 2009 ; De Boer, 2010). » (ibid., p. 37). Cependant, la taille du conduit vocal ne permet d’expliquer qu’une partie des différences observées, seulement 30% d’après Johnson (2005). Simpson (2000) suggère que ces différences sont le reflet des dimensions de la cavité buccale, la langue des femmes ayant de plus courtes distances à parcourir, la même vitesse d’articulation leur permet d’atteindre des postures articuloires plus extrêmes que les hommes. Un autre point de vue avance que ces différences découleraient d’une stratégie articuloire de compensation liée à la hauteur de la voix.

La F0 des femmes étant en moyenne plus haute, l'enveloppe spectrale de leur voix est moins riche en harmoniques audibles, les locutrices auraient donc tendance à articuler les voyelles ouvertes à la périphérie de la cavité buccale (élargissant ainsi leur triangle vocalique) pour mieux différencier chaque voyelle et ainsi les rendre plus identifiables (Goldstein 1980 ; Ryalls et al. 1982 ; Maurer et al. 1992 ; Diehl et al. 1996 ; Simpson et al. 2007 et Ditcheva et al. 2008).

Cependant, Blandon et al. (1984), Henton (1995) et Johnson (2005) constatent que les différences formantiques entre les hommes et les femmes varient considérablement en fonction de la langue parlée. Cette réalité est illustrée par Johnson dans son article de 2005 (voir la figure 2).

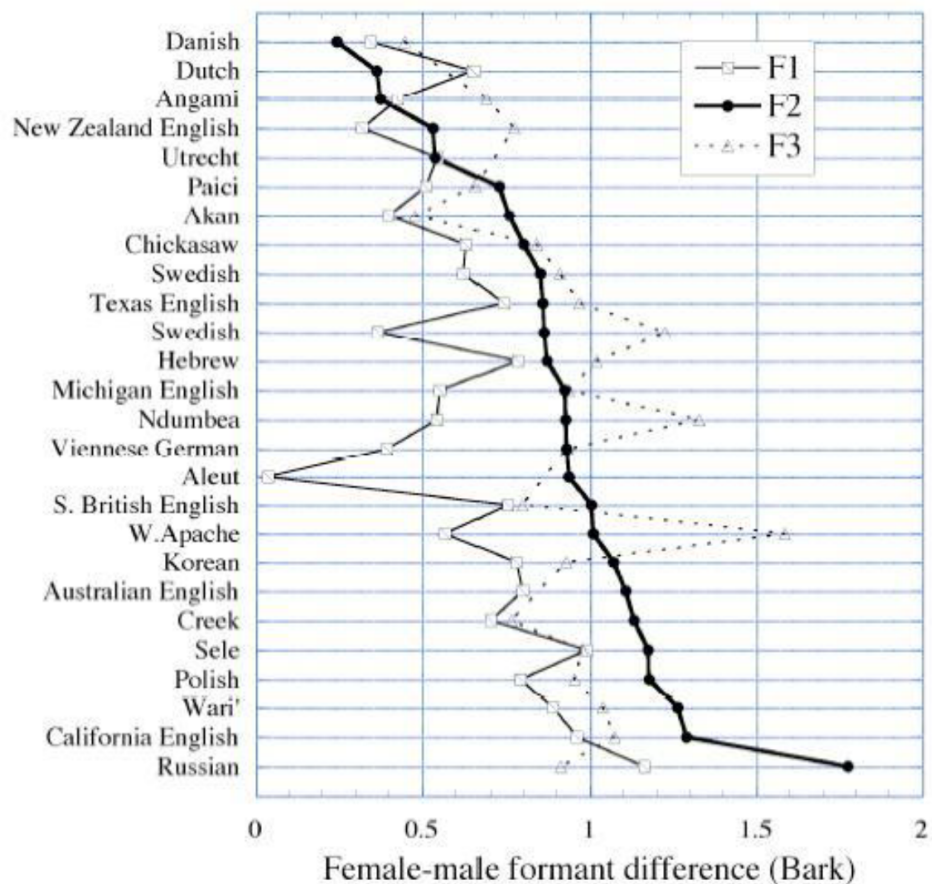


FIGURE 2 – Différences entre les valeurs moyennes de F1, F2 et F3 (en bark) chez les femmes et chez les hommes, dans 26 langues (Johnson 2005)

À la vue de ce graphique, il devient évident que les structures formantiques d'homme et de femme ne sont pas universelles et qu'il n'existe de pas de coefficient multiplicateur

qui permette de passer de l'une à l'autre. Dès lors, il semble peu vraisemblable que cette différence spectrale s'explique uniquement par une dissemblance d'ordre physiologique. Johnson (2006) rappelle que cette différence est utilisée comme un indice pour reconnaître le genre, il en conclut qu'elle est une preuve que les locuteur·ice·s performant leur genre. Cela suggère que les hommes et les femmes ont développé des pratiques articulatoires qui renforcent les dissemblances entre leurs voix afin de se conformer à une certaine représentation des « voix d'hommes et de femmes » et ainsi faciliter l'identification de leur propre genre. Simpson arrive à la même conclusion¹ : « *One factor that is undoubtedly involved in all of the differences we have described is sociophonetic, that is, male and female speakers have learnt to speak in ways appropriate to their gender.* » (Simpson 2009, p. 632)

Autres paramètres

Au-delà de la F0 et des formants vocaliques, il existe bien d'autres paramètres fréquents qui permettent de décrire la voix. L'intonation, c'est-à-dire l'évolution de la F0 au cours d'un énoncé, est souvent citée comme une distinction entre les voix d'homme et de femme. Plusieurs études ont montré que les femmes utilisent une plage de variation de F0 plus importante que les hommes, et ce, même lorsque la F0 est étudiée en demi-ton ou en bark² afin de s'affranchir du caractère logarithmique de la perception humaine de la hauteur. Citons par exemple Johns-Lewis (1986) pour l'allemand, Haan et al. (1999) et Haan (2002) pour le hollandais, Pegoraro-Krook (1988) en suédois et Pépiot (2013) en français. Malgré tout, cette variable semble fortement dépendante du dialecte parlé comme le soulèvent Clopper et al. (2011). Enfin Pegoraro-Krook (1988) remarque que la plage de variation de la F0 évolue avec l'âge : si elle est bel et bien plus étendue chez les femmes que chez les hommes entre 20 et 50 ans, une telle différence n'est plus notable entre 50 et 79 ans. Cette différence entre les voix d'homme et de femmes pourrait donc tendre à disparaître passé un certain âge.

Citons aussi le pic spectral et le centre de gravité spectral des zones de bruit des consonnes fricatives³. En français on observe un centre de gravité spectral ainsi qu'un pic spectral sensiblement plus aigus chez les locutrices que chez les locuteurs (Pépiot et al.

1. « Un facteur qui est sans aucun doute impliqué dans toutes les différences que nous avons décrites est d'ordre sociophonétique, c'est-à-dire que les locuteur·ice·s ont appris à parler d'une manière appropriée à leur genre. » (Traduction de l'auteur)

2. Le bark est une échelle perceptive décrivant la hauteur, proposée par Eberhard Zwicker.

3. Les consonnes fricatives sont des consonnes produites par un rétrécissement du conduit vocal en un point. Elles consistent donc en un bruit de friction, caractéristique du phonème [f], par exemple

2021). On peut aussi imaginer d'autres indicateurs de ces mêmes paramètres comme le \bar{G}_{talker} , qui décrit la position des formants, proposé par Assmann et al. (2008) et repris par Barreda et al. (2021).¹

On peut également étudier des paramètres purement temporels comme le *Voice Onset Time* (VOT), le débit de parole ou encore la durée des voyelles. Le VOT se définit comme le « laps de temps entre le relâchement d'une consonne occlusive ([p], [b], [t], [d]...) et l'entrée en vibration des plis vocaux sur la voyelle qui suit. » Pépiot (2013, p. 50), il dépend de la consonne étudiée. En français, on observe un VOT plus long pour les occlusives² sourdes chez les femmes que chez les hommes et un pré-voisement des occlusives voisées³ plus long chez les femmes que chez les hommes. Les hommes ont un débit de parole plus rapide que celui des femmes en français et la part allouée aux consonnes sur le temps de phonation total est plus importante chez les femmes que chez les hommes. Ces résultats rejoignent une tendance observée plus largement selon laquelle les locutrices ont tendance à adopter des stratégies articulatoires qui favorise l'intelligibilité (Pépiot et al. 2021).

Mentionnons enfin le *mode de voix* qui est largement considéré comme un marqueur sociolinguistique. Aussi appelé *type de phonation*, il s'agit de la manière dont des plis vocaux vibrent lors de la phonation et de la quantité d'air qu'ils laissent ainsi passer. Au cours d'un mode de phonation normal, les plis vocaux oscillent entre deux positions extrêmes, l'une ouverte, dans laquelle le flux d'air qui traverse le larynx est maximal, et l'autre fermée, dans laquelle les plis vocaux se touchent et bloquent le flux d'air. Lors de la production d'une voix modale, les plis vocaux se trouvent en position fermée une fois par cycle de phonation. En adoptant un grand degré d'ouverture, les plis vocaux ne se ferment jamais totalement lors de leur cycle de vibration, ils laissent donc passer un flux d'air important qui caractérise une « voix soufflée » qu'on peut associer à un bruit d'aspiration à l'écoute de la parole. Au contraire, avec un degré d'ouverture plus petit, les plis vocaux, effectuant le même mouvement, se trouvent en position fermée pendant un plus long temps du cycle de phonation. Cette irrégularité dans la vibration des plis vocaux se traduit par une voix aux accents « grinçants » à l'écoute⁴, on qualifie alors la voix de

1. Voir la partie **I.2.2**

2. Les occlusives sont des consonnes produites par la fermeture du conduit vocal suivie d'un relâchement brutal de l'air. Le son produit est un bruit plosif, caractéristique du phonème [p], par exemple.

3. Le voisement désigne tout son produit par la mise en vibration des plis vocaux. Les sons produits par le voisement sont dits « voisés » et les sons produits en l'absence de voisement sont dits « sourds ». La présence du voisement permet de différencier les sons voisés des sons non-voisés, par exemple l'occlusive voisée [b], de l'occlusive sourde [p].

4. Des fichiers audio montrant un exemple de chaque mode de voix sont disponibles sur ce [lien SoundCloud](#).

creaky (Pépiot 2013). Si le mode de voix *creaky* est d'avantage adopté par les locuteurs masculins en anglais, cela ne semble pas être le cas en français (Pépiot et al. 2021). En revanche, les locutrices du français utilisent des voix significativement plus soufflées que les hommes (Pépiot 2013).

La multiplicité des études permet d'établir un consensus solide sur le différentiation de la F0 et de la position des formants entre les voix d'hommes et de femmes. De nombreux autres paramètres et leur expression en relation avec le genre des locuteur·ice·s sont étudiés ; qu'ils soient d'ordre fréquentiel ou temporel, leur prépondérance dans l'expression du genre dans la voix semble bien souvent fonction de la langue parlée.

1.2.2 Perception du genre dans la voix

Sokhi et al. (2005) et Lattner et al. (2005) ont montré que le cerveau traite différemment les voix d'hommes et de femmes, indépendamment du genre de l'auditeur·ice. On peut penser que si le cerveau traite différemment les voix d'homme et de femme, c'est qu'il est capable au préalable de discriminer les voix en fonction du genre. Mullennix (1995) suggère que les auditeur·ice·s utilisent la représentation qu'ils se font des voix d'hommes et de femmes d'après leurs expériences passées afin de déterminer le genre des nouvelles voix auxquelles iels sont confronté·e·s.

De nombreuses études ont cherché à mettre en évidence les paramètres acoustiques utilisés pour identifier le genre d'une voix. Selon le consensus, la fréquence fondamentale et la position des formants vocaliques sont les deux critères principaux qui, associés, permettent d'identifier le genre d'un·e locuteur·ice (Pépiot 2013). Par exemple, S. Whiteside (1998) montrent que des auditeur·ice·s sont capables d'identifier correctement le genre à 98,9% à partir d'extraits de voyelles d'une durée de 50 à 100 ms. Schématiquement, plus la F0 et les fréquences des formants sont élevées, plus les auditeur·ice·s ont tendance à identifier une femme ; inversement, plus elles sont basses, plus iels identifieront un homme. Lorsque l'on présente des voyelles dépourvues de l'un de ces deux paramètres, l'identification du genre est moins souvent correcte mais elle reste largement supérieure à la chance (supérieure à 75%) (Lass et al. 1976). D'autres études ont permis de montrer que les auditeur·ice·s sont susceptibles d'utiliser l'information des zones de bruits des parties apériodiques du spectre¹ pour déterminer le genre (Schwartz 1968, Ingemann 1968, Wu

1. Les parties apériodiques du spectre correspondent aux sons produits en l'absence de voisement. Ces sons s'apparentent à du bruit, leur spectre présentent des zones fréquentielles renforcées par l'effet des résonateurs du conduit vocal, on les appelle *zones de bruit*.

et al. 1991 et S. P. Whiteside 1998). Dans l'étude de Schwartz (1968), les participant·e-s ont été capables d'identifier le genre des locuteur·ice-s à l'écoute de consonnes fricatives sourdes isolées¹. Pépiot (2010) a pu mettre en évidence une légère influence de l'intonation dans l'identification de genre des locuteur·ice-s français·e-s lorsque l'information de la F0 est neutralisée. Il compare la performance de discrimination de deux groupes d'auditeur·ice-s à l'écoute de voix d'homme et de femme avec une F0 moyenne resynthétisée à 169Hz (valeur intermédiaire entre les F0 typiquement masculine et féminine); l'un des groupes écoute ces voix avec le contour intonatif originel conservé, l'autre avec une intonation aplatie. Le pourcentage d'identifications correctes obtenues est de 97,4% pour le premier groupe contre 91,41% pour le second. L'intonation semble donc jouer un rôle dans la discrimination du genre en français; en revanche l'auteur n'a obtenu aucun résultat signifiant pour la même expérience conduite avec des locuteur·ice-s de l'anglais. Enfin, Borsel et al. (2009) ont montré que la qualité de voix joue un rôle dans l'évaluation du degré de féminité de voix de femmes.

L'effectivité et la prépondérance des critères cités précédemment diffèrent en fonction de la langue étudiée, indiquant que le phénomène est au moins en partie culturel (Pépiot 2013). Nous étudierons plus en détail les différences dans le processus d'identification du genre entre l'anglais et le français dans la partie **I.9.1**.

Dans une récente étude très complète sur la perception du genre dans la voix de locuteur·ice-s de 5 à 18 ans, Barreda et Assmann font une remarque cruciale² :

« Overall, our results show that listeners are more influenced by the average spectral characteristics of a talker (e.g., \bar{G}_{talker})^a than they are by independent variation in the absolute formant frequencies of a given token. This leads to an apparent paradox : the best predictor of talker-gender judgments (\bar{G}_{talker}) is not directly present in the signal. » (Barreda et al. 2021, p. 3956)

a. \bar{G}_{talker} est une variable étudiée par Barreda et Assmann, reprise des travaux précédents d'Assmann (ibid 2008) qui correspond à la moyenne des valeurs logarithmiques de F1, F2 et F3 de toutes les voyelles (sur une échelle log-Hz en base 0,13)

Barreda et Assmann suggèrent ici que le cerveau utilise comme indicateur de genre, non pas une valeur spectrale absolue, mais plutôt une moyenne déduite du profil spectral

1. Ce sont des consonnes produites sans vibration des plis vocaux, elles se caractérisent donc par une absence de F0).

2. « Dans l'ensemble, nos résultats montrent que les auditeur·ice-s sont plus influencé·e-s par les caractéristiques spectrales moyennes d'un·e locuteur·ice (par exemple, \bar{G}_{talker}) qu'ils ne le sont par la variation indépendante des fréquences absolues des formants d'un stimulus donné. Ceci conduit à un paradoxe apparent : le meilleur prédicteur des jugements de genre, le \bar{G}_{talker} , n'est pas directement présent dans le signal. » (Traduction de l'autrice)

de plusieurs phonèmes. L'analyse menée par le cerveau serait donc d'avantage temporelle et relative plutôt que paramétrique et discrète ; elle recouperait les paramètres spectraux de plusieurs phonèmes accumulés, analysés relativement les uns par rapport aux autres.

Pépiot, quant à lui, critique la tendance dans la littérature à chercher à hiérarchiser les indices acoustiques liés à la perception du genre. En effet, selon lui, il n'existe pas un indicateur principal du genre mais le cerveau pondérerait l'importance d'un critère d'analyse en fonction des l'états des autres paramètres ; ainsi « lorsqu'une voix présente une fréquence fondamentale ambiguë, le rôle joué par les autres paramètres acoustiques, tels que la position des formants, se trouverait automatiquement renforcé. » (Pépiot 2013, p.72-73). Les états de plusieurs paramètres acoustiques de la voix se combineraient pour constituer un faisceau d'indices concordants qui permettent au cerveau de discriminer le genre. Smith (2016) souligne que les auditeur·ice·s adaptent leur stratégie de discrimination du genre à la nature du stimuli qui leur est présenté. Exposé·e·s à des voyelles isolées, iels fondent leur identification sur la position des formants mais, à l'écoute d'un énoncé plus long, iels utiliseraient davantage la F0¹.

« The suggestion is that in a speaker-sex discrimination task, the listener combines what information is available using early-available (but less reliable) information at the start of the decision process but, as time exposed to the stimulus increases, switches to late-available (but more reliable) information. » (Smith 2016, p. 2-3)

Enfin, Mattingly (1966) suggère que les hommes et les femmes accentuent par leurs pratiques articulatoires des différences acoustiques d'origines physiologiques afin d'être perçu·e·s conformément aux stéréotypes associés au genre performé. C'est aussi ce que soutient Sachs² :

« Adult men and women may modify their articulators, lowering or raising the formant frequencies, to produce voices that aim toward male-female archetype. [...] In other words, men may try to talk as if they are bigger than they actually are, and women may talk as if they are smaller than they actually are. » (Sachs 1975, p. 154)

1. « Nous suggérons que, pour discriminer le genre d'un·e locuteur·ice, l'auditeur·ice combine les informations disponibles en utilisant les premières informations disponibles (mais moins fiables) au début du processus de décision mais, au fur et à mesure que le temps d'exposition au stimulus augmente, iel privilégie des informations disponibles plus tard (mais plus fiables). » (Traduction de l'autrice)

2. « Les hommes et les femmes adultes peuvent modifier la position de leurs articulateurs, en abaissant ou en augmentant les fréquences des formants, pour produire des voix qui visent des archétypes masculin/féminin. [...] En d'autres termes, les hommes essaieraient de parler comme s'ils étaient plus grands qu'ils ne le sont en réalité, et les femmes parleraient comme si elles étaient plus petites qu'elles ne le sont en réalité. » (Traduction de l'autrice)

1.2.3 Les différences inter-langues des indices de genre

Les différences français/anglais

Il est à noter que l'écrasante majorité des études sur le genre dans la voix des enfants concernent la langue anglaise. Attardons-nous donc plus particulièrement sur les différences d'indices de genre entre l'anglais et le français.

Pépiot (2013) réalise une étude conjointe sur des locuteur·ice·s anglophones de l'anglais américain et des locuteur·ice·s francophones du français parisien. Il met en évidence plusieurs différences spectrales subordonnées au genre entre les deux langues. Par exemple, si les voix de femmes sont plus hautes que les voix d'hommes dans les deux langues, les voix d'hommes anglophones sont significativement plus basses que celles de leurs homologues francophones. De même, en français, F2 et F3 ont des valeurs significativement plus élevées chez les femmes que chez les hommes toutes voyelles confondues mais pas F1 ; tandis qu'en anglais, F1, F2 et F3 sont significativement plus élevées chez les femmes que chez les hommes. Il trouve aussi qu'en français, le genre a un effet significatif sur le pic spectral des consonnes, ce qui n'est pas mis en évidence en anglais.

Pépiot remarque également que les indices de genre dans la voix ne sont pas utilisés de la même manière par les auditeur·ice·s en fonction de la langue écoutée. En anglais, la F0 moyenne semble jouer un rôle plus important dans la détermination du genre qu'en français ; à l'inverse, la valeur des formants vocaliques semble influencer davantage les auditeur·ice·s francophones que les anglophones. L'auteur relève que ces résultats corroborent les résultats trouvés dans différentes études qui semblent contradictoires si l'on ne prend pas en compte le facteur de la langue : « Il est tout à fait probable, comme le suggèrent notamment les études de Coleman (1976) et Pausewang Gelfer & Mikos (2005), que le F0 moyen est bien le paramètre acoustique le plus important pour reconnaître le genre chez les auditeurs anglophones américains, mais que les fréquences de résonance et en particulier les valeurs formantiques sont l'indice le plus saillant pour les auditeurs francophones (Arnold, 2008, 2012). » Pépiot (*ibid.*, p. 232). Il dresse un tableau récapitulatif de la prépondérance de certains indices de genre dans la perception de celui-ci en anglais et en français (voir la figure 3).

Les données présentées dans ce tableau sont à tempérer puisque, comme l'indique l'auteur, il n'est pas exhaustif et ne rend pas compte de l'interaction des critères entre eux. Toutefois il donne une bonne idée de la manière dont les perceptions du genre se fondent sur différents paramètres d'une manière qui est propre à chaque langue. Cela permet de valider l'hypothèse selon laquelle le processus d'identification du genre par la

IDENTIFICATION DU GENRE PAR LA VOIX A PARTIR DE MOTS DISSYLLABIQUES		
Paramètre acoustique	Degré d'influence estimé	
	<i>Auditeurs francophones</i>	<i>Auditeurs anglophones américains</i>
<i>F0 moyen</i>	****	*****
<i>Fréquence des formants vocaliques</i>	*****	***
<i>Qualité de voix (H1-H2)</i>	∅	***
<i>Vitesse d'élocution (durée du mot)</i>	*	**
<i>Plage de variation de F0</i>	**	∅
<i>Zones de bruit des consonnes</i>	*	*

FIGURE 3 – Degré d'influence estimé des paramètres acoustiques pouvant jouer un rôle dans l'identification du genre par la voix à partir de mots dissyllabiques, pour les auditeurs francophones et anglophones américains. Le symbole « ∅ » indique l'absence d'influence, les autres indications utilisent une échelle graduée allant de « * » (très faible influence) à « ***** » (très forte influence). (Pépiot 2013)

voix diffère en fonction de la langue maternelle de l'auditeur·ice.

Ainsi, si notre étude bibliographique, fondée très largement sur des études de la langue anglaise, laissera entrevoir certains consensus comme le fait qu'il est possible de discriminer le genre des enfants à l'écoute de leur voix seule et le rôle primordial que jouent les formants et la F0 dans ce processus ; ces attentes sont faussées par la différence d'expression de genre entre le français et l'anglais. Aucun de ces phénomènes n'est pour le moment démontré en français.

Les stratégies articulatoires des multilingues

B. Lee et al. (2017), dans une étude sur des locutrices bilingues, analysent les F0 de 11 femmes bilingues coréen/anglais et 11 femmes bilingues mandarin/anglais sur un corpus de phrases lues, de descriptions d'images et de discours semi-spontané. Elles trouvent que les locutrices des deux groupes produisent une F0 significativement plus élevée respectivement en coréen et en mandarin qu'en anglais. Altenberg et al. (2006) avaient trouvé des résultats similaires dans une étude menée sur des locutrices du russe et de l'anglais. Alors que les locutrices bilingues adoptaient une F0 significativement plus élevée en russe qu'en anglais, il n'y avait pas de différence significative entre les F0 moyennes des locutrices bilingues lorsqu'elles s'expriment en anglais et celles des locutrices monolingues de l'anglais. Altenberg et Ferrand concluent que les productions vocales de personnes bilingues et monolingues sont comparables sur le plan phonétique. Ces deux études ne testant que des femmes, les différences observées pourraient être imputées à des différences phonologiques entre les langues étudiées, toutefois d'autres études ont mis en regard des participant·e·s

bilingues des deux genres. Citons Ordin et al. (2017), qui étudient la F0 de 18 hommes et 14 femmes bilingues écossais/anglais dans des phrases lues. Iels observent une plage de variation de la F0 plus étendue chez les locutrices en écossais qu'en anglais, mais n'observent aucune différence chez les locuteurs entre les deux langues. Cheng (2020) étudie le discours semi-spontané de 10 femmes et 5 hommes bilingues coréen/anglais. Il note que les locuteur·ice·s s'expriment avec une F0 plus élevée en coréen qu'en anglais et qu'iels utilisent une plage de variation de F0 plus étendue en coréen qu'en anglais, le genre ayant un effet significatif sur ces différences inter-langues. Les auteur·ice·s de ces deux études en concluent qu'en passant d'une langue à l'autre, les locuteur·ice·s ne modulent pas leur voix de la même manière en fonction de leur genre.

Pépiot et al. (2021), dans une étude très complète sur des locuteur·ice·s bilingues du français et de l'anglais américain, montrent que les locuteur·ice·s bilingues s'adaptent aux normes de prononciation de genre de la langue dans laquelle iels s'expriment. Ils constatent que les F0 des femmes sont plus élevées que celles des hommes dans les deux langues mais que les femmes ont une F0 significativement plus élevée lorsqu'elles parlent en français (une F0 supérieure de 2,76% en français en moyenne pour des phrases lues). La déviation standard de la F0 diminue chez tout·e·s les locuteur·ice·s lorsqu'iels passent de l'usage de l'anglais au français mais le phénomène est significativement plus marqué chez les locuteurs. L'étude montre également que les locuteur·ice·s présentent des triangles formantiques différents en fonction de la langue dans laquelle iels s'expriment.

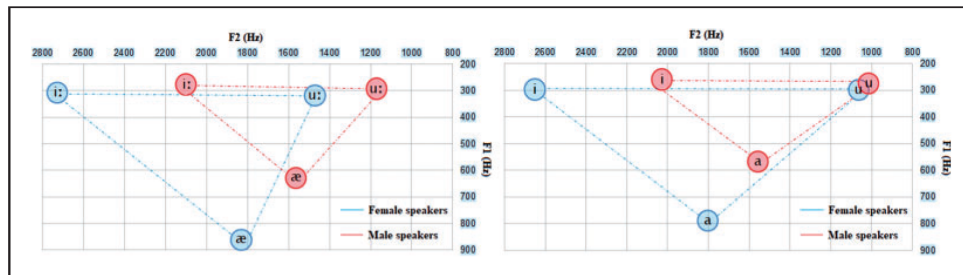


FIGURE 4 – Triangles formantiques des voyelles de l'anglais [i:] [æ] and [u:] (gauche), et des voyelles du français [i], [a], [u] (droite) prononcées par des locutrices (bleu) et des locuteurs (rouge) bilingues (Pépiot et al. 2021)

On remarque en particulier la valeur de F2 du phonème [u] en français, qui n'est pas différent en fonction du genre mais qui est en revanche plus élevée dans le phonème [u:] chez les locutrices en anglais.

L'étude met également en évidence des différences significatives entre les langues en fonction du genre pour le VOT ou dans la qualité de voix. Par exemple, les femmes ont

tendance à utiliser une voix légèrement plus *breathy* en anglais qu'en français tandis que les hommes utilisent une voix plus *creaky* en anglais qu'en français.

On peut conclure de ces études que les personnes bilingues semblent avoir acquis la capacité d'adapter leurs stratégies articulatoires afin de moduler leur voix conformément aux normes de genre spécifiques à la langue dans laquelle iels s'expriment.

En conclusion de cette partie, on peut considérer que la voix est bel et bien un indicateur de genre. Les locuteur·ice·s modulent leur voix, dans les limites des capacités de leur appareil vocal, d'une manière qui renseigne leurs interlocuteur·ice·s sur leur genre et qui, simultanément, nourrit l'idée qu'iels se font de ce qu'est une « voix d'homme » et une « voix de femme ». Par la suite nous examinerons de quelle manière les notions de « voix d'homme » et « voix de femme » se télescopent sur la représentation que nous nous faisons des voix d'enfants et s'il est possible de mettre en évidence d'éventuelles représentations analogues que nous appellerons « voix de garçon » et « voix de fille ».

1.2.4 Le cas de la voix chuchotée

Description acoustique du chuchotement

La voix chuchotée se caractérise par l'absence de voisement. Lors de la production de la voix chuchotée, les plis vocaux n'entrent pas en vibration. Un flux d'air constant traverse la glotte ouverte, comme lors de la respiration. Il existe cependant une constriction de la glotte qui prend généralement une forme de 'Y' afin de réduire l'aire glottique par rapport à une position d'ouverture totale. Cette constriction crée une turbulence lors du passage de l'air qui génère un bruit. Le son produit lors du chuchotement est donc entièrement généré à partir de bruits de turbulence. Il est apériodique, ce qui se traduit par une absence de F0 (Fux 2012).

Bien que le chuchotement soit entièrement composé de bruit, l'articulation permet de renforcer des zones de bruit analogues aux zones formantiques de la parole voisée ce qui permet de différencier les voyelles et de conserver l'intelligibilité de la parole. On appelle improprement « formants » ces zones renforcées du spectre par analogie avec la parole voisée (ibid.). L'étude de la modélisation du système acoustique de l'appareil de production vocal permet de constater qu'une ouverture glottique plus grande entraîne une impédance glottique plus faible. Ainsi la voix chuchotée entraîne un déplacement vers les hautes fréquences des formants, en particulier pour les deux premiers (voir la figure 5, Morris 2003).

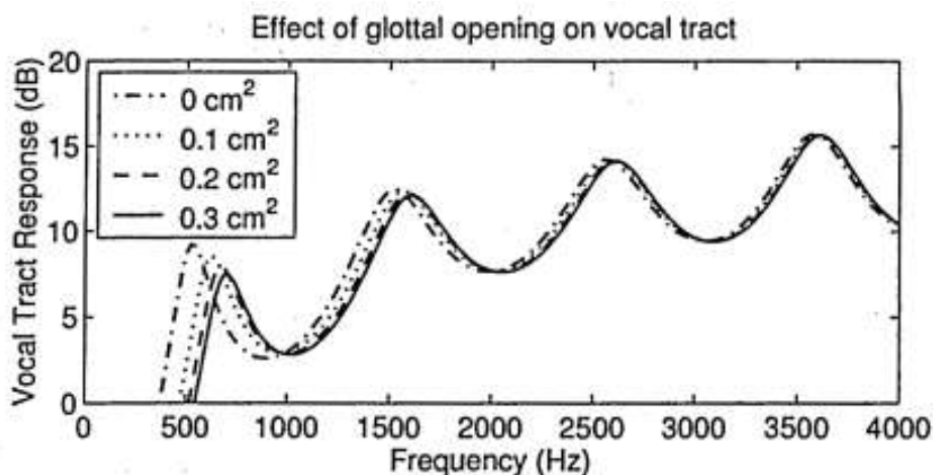


FIGURE 5 – Conséquences de la variation de l’ouverture glottique sur le spectre d’une voyelle, calculées à partir d’une simulation du conduit vocal (Morris 2003)

Cette translation des zones formantiques a pu être observée expérimentalement à plusieurs reprises. Une augmentation importante de la valeur de F1 et plus réduite de la valeur de F2 a été constatée par Jovičić (1998) en serbe ; Matsuda et al. (1999) et Ito et al. (2005) en japonais ; Petrushin et al. (2010) en russe ; ; Sharifzadeh et al. (2012) en anglais britannique et Swerdlin et al. (2010) en anglais australien. L’hypothèse selon laquelle la translation des formants vocaliques est identique chez les hommes et les femmes ne fait pas consensus. Sharifzadeh et al. (2012) montrent une augmentation de la valeur des formants moins marquée chez les femmes, en particulier en ce qui concerne F1 tandis que Jovičić (1998) qualifiait les variations de la position des formants de similaires entre les hommes et les femmes. Ito et al. (2005) présente des tableaux de variation du spectre entre voyelles chuchotée et voisées relativement semblables pour les hommes et les femmes (voir figure 6).

La figure 7 montre la différence de l’allure du spectre d’une voyelle voisée comparée à celle d’une voyelle chuchotée. Jovičić (1998) note un élargissement de bandes des formants lors du chuchotement, là aussi plus important pour F1 que pour F2. La voix chuchotée présente un spectre globalement aplati par rapport à la voix parlée. L’amplitude de tous les formants est diminuée et leur énergie décroît plus lentement le long du spectre. Dans la région basse du spectre des voyelles, l’enveloppe spectrale est décroissante en voix chuchotée contrairement à la parole voisée. Enfin le spectre est généralement plus plat dans la zone comprise entre 200 Hz et 2000 Hz (Fux 2012).

Mentionnons rapidement un effet qui va de soi sur le niveau sonore de la parole. Ito et

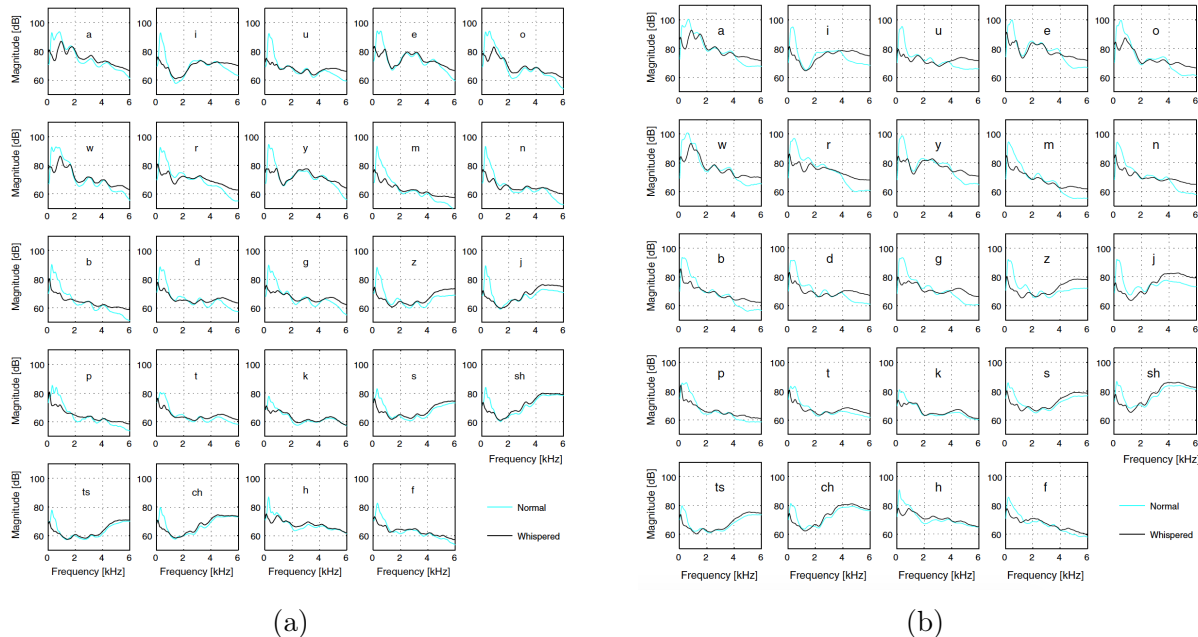


FIGURE 6 – Spectres moyens de 24 voyelles voisées et chuchotées prononcées par des femmes (a) et des hommes (b) (Ito et al. 2005)

al. (2005) mesurent une diminution de 20 à 25 dB de l'intensité des voyelles en l'absence de voisement. Les consonnes non-voisées n'étant pas à peu impactées par cette diminution d'intensité Jovičić et al. (2008), le rapport de niveau entre les voyelles, les consonnes voisées et les consonnes non-voisées se trouve modifié lors du chuchotement. Fux (2012) relèvent également des effets sur la prosodie, le débit de parole étant en général ralenti lors du chuchotement.

Implications sur la perception du genre

L'étude de la perception des voix chuchotée peut être utile pour évaluer l'impact de la F0 dans l'identification du genre en première approximation. On compare la capacité à identifier le genre d'une voix parlée (présence de l'indice F0) et d'une voix chuchotée (absence de F0). Cette méthode a été utilisée par Schwartz et al. (1968), Lass et al. (1976), Bennett et al. (2010), Weinberg et al. (1979b) et Smith (2016) en langue anglaise et Eklund et al. (1997) en suédois. Cette méthode est d'autant plus intéressante, qu'à l'écoute du chuchotement, l'intonation continue à être reconnue, malgré l'absence de F0 (Vercherand 2010). Pour faire la comparaison entre un signal voisé et chuchotée, on normalise en général le niveau des deux signaux. Smith, lors de son étude de 2016, s'est assuré que la

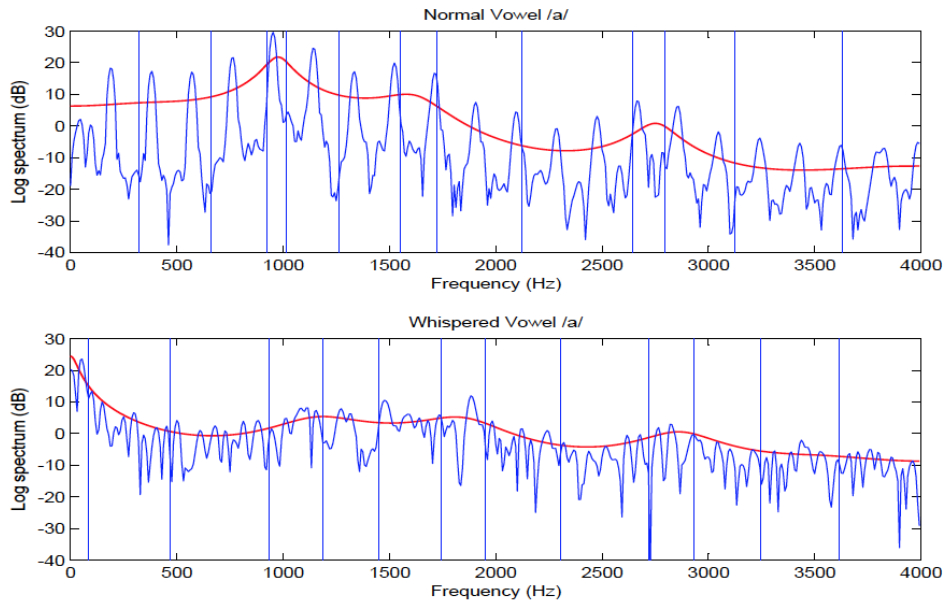


FIGURE 7 – Comparaison des spectres de la voyelle [a] voisée (haut) et chuchotée (bas) d’un même locuteur (Sharifzadeh et al. 2009)

modification du niveau n’a pas d’influence sur la perception du genre. Dans son expérience, les niveaux des voix parlées et chuchotées étaient alignés à 77dB SPL. Dans une expérience annexe, il a fait écouter à deux groupes d’auditeur·ice·s des voyelles isolées chuchotées et voisées. L’un des groupes a écouté les sons augmentés de 6dB, toutes choses égales par ailleurs. Il conclut qu’il n’y a aucun effet notable de l’intensité perçue sur les réponses données : « *changes in perceptual loudness do not underlie differences in speaker-sex discrimination performance between voiced and whispered vowels.* » (Smith 2016, p. 11).¹

Cette méthode, bien qu’imparfaite - puisque le chuchotement impacte d’autres paramètres que la F0 comme nous l’avons vu - a l’avantage de ne pas nécessiter une manipulation numérique de la F0 comme le morphing de voix ou création de voix hybrides. En effet, ces méthodes posent le problème de la « naturalité » des voix obtenues et des effets de la manipulation numérique sur la perception de la voix (problème des voix robotiques, artificielles, qui s’éloignent d’une écoute réaliste, comme le note Arnold (2015) dans une étude où il a eu recours à de telles manipulations).

1. « Les changements de niveau perçu n’entraîne pas de différences dans la performance de discrimination du genre des locuteur·ice·s entre les voyelles voisées et chuchotées. » (Traduction de l’auteurice)

1.3 Physiologie de la production de la voix avant la puberté

L'étude des voix d'enfants est un enjeu important au sein du débat sur la part d'origine physiologique et la part liée aux pratiques sociales dans l'examen des différences entre les voix d'hommes et de femmes. En l'absence de différences physiologiques entre les appareils vocaux des garçons et des filles, le constat de disparités acoustiques entre les voix en fonction du genre avant l'adolescence est souvent considéré comme une preuve de l'existence de pratiques articulatoires apprises spécifiques à un contexte culturel (Sachs 1975).

Kahane (1978), procède à l'étude du larynx de 20 cadavres de garçons et de filles âgé·e·s de 9 à 18 ans. Il n'observe aucune différence claire entre les sexes avant la puberté¹. Rogers et al. (2014) ont étudié la taille des plis vocaux de 87 filles et 118 garçons âgé·e·s de 1 mois à 20 ans. Iels ont observé quatre dimensions différentes : la longueur totale des plis vocaux, la taille de la partie membraneuse des plis vocaux, la taille de la partie cartilagineuse des plis vocaux et le ratio de la partie membraneuse sur la partie cartilagineuse. Aucune des quatre dimensions n'a présenté de différence significative en fonction du genre. Sur le plan morphologique, il n'y a donc pas de raison que les garçons aient des F0 significativement plus basse que les filles.

Fitch et al. (1999), d'après l'étude d'une banque de données d'imagerie médicale, constatent qu'au cours de l'enfance, la longueur du conduit vocal (tracé en rouge dans la figure 8) est proportionnelle au logarithme de la masse corporelle et à la taille de l'enfant, ce qui indique que le conduit vocal observe une croissance stable à mesure que l'enfant grandit. Ces résultats sont cohérents avec ceux obtenus par King (1952) et Goldstein (1980). Fitch et Giedd n'observent aucune différence de la longueur du conduit vocal entre les garçons et les filles avant la puberté. Ils en déduisent que les différences acoustiques observées entre les voix de garçons et de filles sont liées à des pratiques articulatoires.

Vorperian et al. (2009) étudient la croissance du conduit vocal de la naissance à l'âge adulte à partir d'une banque de données d'imagerie médicale (TDM et IRM²) des têtes et cous de 605 individu·e·s. Iels constatent que la croissance du conduit vocal n'est pas uniforme en fonction de la région étudiée et de l'âge. En 2011, les mêmes auteur·ice·s

1. L'auteur précise que les sujets sont considéré·e·s comme prépubèr·e·s s'iels ne présentent aucune trace de pilosité pubienne ni aucun signe de développement de la poitrine.

2. L'acronyme TDM désigne la tomodensitométrie qui est une méthode de scanographie médicale. L'acronyme IRM désigne l'imagerie par résonance magnétique.

étudient l'évolution de 9 dimensions du conduit vocal à partir de la même base de données. Iels se penchent sur la longueur totale du conduit vocal, *VTL*; la section verticale du conduit vocal, *VT-V*, qu'iels décomposent en une partie postérieure, *PCL*, et une partie nasale, *NPhL*; la section horizontale du conduit vocal, *VT-H*, qu'iels décomposent en trois sous-sections : l'épaisseur des lèvres, *LTh*, la cavité antérieure, *ACL*, et la largeur oropharyngée, *OPhW*; enfin la longueur de la cavité buccale, *VT-O* (chacune de ces dimensions est indiquée sur la figure 8). Sans surprise, 6 des 9 dimensions étudiées se sont avérées significativement différentes en fonction du sexe de 15 à 19 ans. De 10 à 14 ans, en revanche, seule l'*OPhW* s'est révélée significativement liée au sexe; pourtant, de 5 à 9 ans, deux critères sont significativement distincts entre les filles et les garçons (*VT-H* et *VT-O*). Vorperian et al. en déduisent que certains dimorphismes sexuels de l'appareil vocal, qui existeraient avant le début de la puberté, pourraient être masqués par la temporalité de la croissance, propre à chaque sexe. Iels choisissent de mener une seconde analyse des mêmes individu·e·s sur une fenêtre glissante d'un mois. En ce qui concerne la tranche d'âge des 8-10 ans, cette analyse indique que seules les dimensions *PLC* et *NPhL* (respectivement en bleu clair et bleu foncé sur la figure 8) sont significativement différentes entre les garçons et les filles, la longueur totale du conduit vocal, elle, n'est pas différenciée entre les deux groupes. Les auteur·ice·s citent les travaux de Fant (1975) qui simplifient le conduit vocal en un modèle à deux tubes (correspondants à la cavité buccale et au pharynx). D'après cette modélisation, la longueur de la cavité orale peut être affiliée à la distance entre la position des formants *F1* et *F3* et la longueur du tube modélisant le pharynx est liée à la différence entre les formants *F1* et *F2*. Si les distinctions physiologiques constatées entre les garçons et les filles entre 8 et 10 ans peuvent donc éventuellement expliquer une potentielle dissemblance entre les positions relatives de *F1* et *F2* dans cette tranche d'âge, elles ne permettent pas de justifier des différences observées entre *F1* et *F3*.

En conclusion, il n'existe pas de différence significative de la taille plis vocaux en fonction du genre chez les enfants. Bien que la recherche continue à développer des modélisations de l'appareil vocal pré-pubère de plus en plus précises et que l'existence de certains dimorphismes sexuels n'est pas exclue; il est à ce jour largement considéré que la différenciation de l'appareil vocal en fonction du sexe apparaît graduellement à partir du début de la puberté. De plus, les dissemblances physiologiques qui ont pu être mises à jour ne permettent pas d'expliquer l'ensemble des différences acoustiques observées dans les voix d'enfants. D'après l'étude acoustique de bases de données de voix d'enfants, on peut considérer que les transformations de la voix liées à la puberté débutent en moyenne à 11 ans chez les filles et à 12 ans chez les garçons et qu'en moyenne, iels atteignent leur

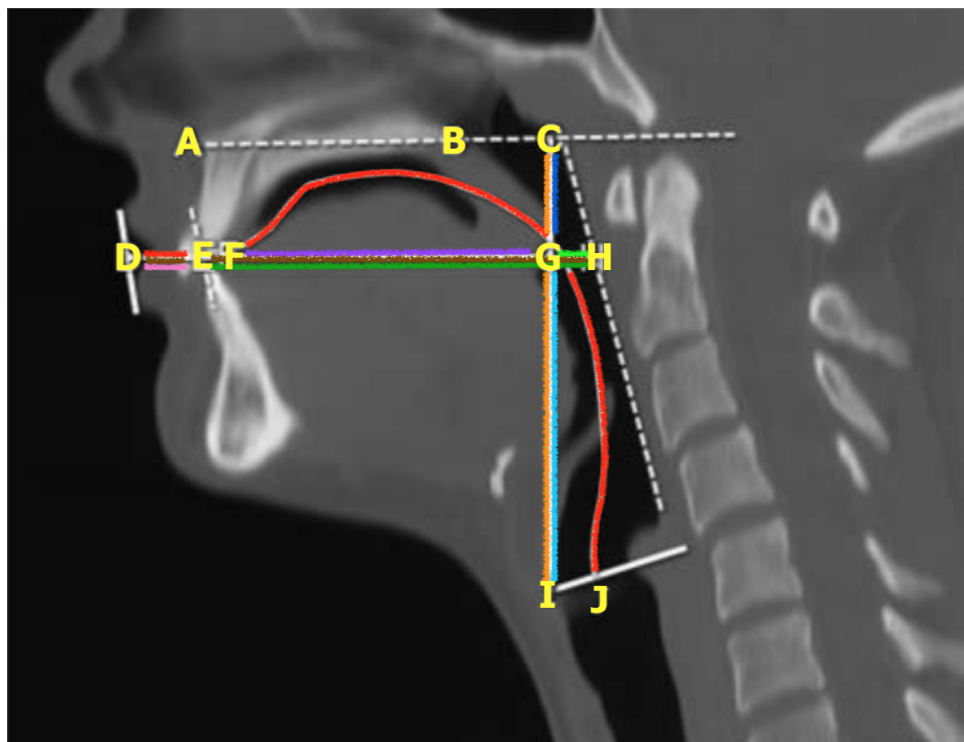


FIGURE 8 – Image sagittale du conduit vocal (obtenue par TDM) sur laquelle sont indiqués les repères anatomiques utilisés pour relever les mesures des variables étudiées. Les 9 variables sont définies ainsi : la longueur du conduit vocal (VTL) correspond à un tracé curviligne entre les points D et J, indiqué en rouge ; la dimension VT-V correspond au segment [CI], en orange ; PCL, au segment [IG], en bleu clair ; NPhL, au segment [GC], en bleu foncé ; VT-H, au segment [DH], en marron ; LTh, au segment [DE] en rose ; ACL, au segment [FG], en violet ; OPhW, au segment [GH], en vert clair et enfin VT-O, au segment [EH], en vert clair (Vorperian et al. 2009).

fréquence fondamentale d'adulte respectivement à 14 et 16 ans (Maturio et al. 2012).

1.4 La F0 en fonction du genre chez les enfants

1.4.1 État de la question

La question de la valeur de la F0 en fonction du genre avant la puberté ne fait pas consensus. Dans une étude de la F0 du discours semi-spontané de 29 garçons et 37 filles de 5 et 6 ans, Weinberg et al. (1971) n'observent pas de différences significatives entre les F0 des deux groupes. Bennett (1983) a mené une étude longitudinale avec des enfants âgé-e-s de 7 à 8 ans, au début de l'expérience, et de 10 à 11 ans, à la fin. Elle a analysé

la même phrase déclarative récitée par chaque enfant, enregistrée à 12 mois d'intervalle pendant 3 ans. Elle n'observe pas de différence significative entre les F0 des garçons et des filles à aucun des âges étudiés. Elle note également que la F0 des participant·e·s baisse au cours des années de l'expérience, de la même manière pour les garçons et les filles. Busby et al. (1994) arrive à une conclusion similaire dans une étude menée sur des mots monosyllabiques usuels prononcés par des enfants âgé·e·s de 5 à 11 ans. Sorenson (1989) ne note pas non plus de différence de F0 entre les garçons et les filles d'un groupe paritaire de 30 enfants âgé·e·s de 5 à 10 ans. Mais lui n'observe aucune évolution significative de la F0 en fonction de l'âge. S. Lee et al. (1999) ne note pas non plus de différence significative de F0 en fonction du genre chez des enfants de 5 à 12 ans lors de la prononciation de phrases et de mots monosyllabiques extraits d'un contexte de phrases lues. Perry et al. (2001) obtiennent le même résultat en étudiant un corpus similaire prononcé par des enfants de 4 à 12 ans, de même pour l'étude de Linders et al. (1995) menée sur des enfants de 7 à 15 ans.

Mais Glaze et al. (1988) arrivent à des conclusions contradictoires. En étudiant la production de voyelles isolées de 59 garçons et 62 de filles de 5 à 11 ans, iels observent une F0 significativement plus élevée chez les filles que chez les garçons. Ces résultats sont concordants avec ceux de Hasek et al. (1980) qui mènent une étude sur un corpus de voix préexistant (Brown et al. 1978) constitué de voyelles isolées prononcées par 6 groupes paritaires composés de 30 enfants chacun, âgé·e·s de 5 à 10 ans. Iels observent une différence significative de F0 entre les garçons et les filles à partir de 7 ans. Cette différence est liée à une diminution de la F0 des garçons entre 7 et 10 ans qui ne se semble pas avoir lieu chez les filles. Vuorenkoski et al. (1978) observent iels aussi une diminution significative de la F0 entre un groupe de garçons de 8 ans et un groupe de garçons de 10 ans, cette différence de F0 ne s'observe pas dans des groupes d'âge équivalent composés de filles. Pourtant l'étude ne met pas en évidence de différence significative de F0 entre les garçons et les filles ni à 8 ans ni à 10 ans, en dépit de l'évolution de la F0 avec l'âge parmi les garçons. Précisons que cette dernière étude a été menée à la fois sur des phrases récitées et lues. S. P. Whiteside et al. (2000) s'intéressent à des phrases issues de description d'images de trois groupes d'enfants de 6, 8 et 10 ans (constitués de 6 à 8 enfants seulement). Elles ne trouvent pas de différence de F0 dans au sein des deux premiers groupes mais mettent en évidence une différence significative pour le groupe d'enfants de 10 ans et leurs résultats montrent que la F0 diminue avec l'âge d'une façon significativement différente en fonction du genre. Nicollas et al. (2008), quant à elles, observent que les garçons ont une F0 moyenne significativement plus basse que celle des

filles du même âge et remarquent que la F0 diminue en fonction de l'âge. Leur étude concerne la production isolée de la voyelle [a] pour 32 filles et 38 garçons français·e·s réparti·e·s dans six groupes d'âge entre 6 et 12 ans. Notons tout de même l'exception des groupes de 6 et 10 ans, dans lesquels la F0 moyenne des filles est plus basse que celle des garçons. Comme nous l'avons vu dans la partie **I.3**, une telle différence, si elle est avérée, n'est *a priori* pas due à un dimorphisme anatomique lié au sexe. Amir et al. suggèrent que ce phénomène pourrait s'expliquer par des pratiques articulatoires culturelles liées à la représentation du genre.¹

« It was suggested that boys attempt to produce a more masculine voice and speech pattern by adopting a lower pitch than their optimal pitch, lowering their jaw, and modifying the extent of lip rounding » (Amir et al. 2012, p. 320)

Notons que l'ensemble des études précédemment citées, à l'exception de la dernière, concernent l'anglais. Soltani et al. (2014) et Ghorbani et al. (2010) ne mettent en évidence aucune différence significative de F0 en fonction du genre chez des enfants respectivement de 4 à 7 ans et de 9 à 13 ans en langue perse. En français, Courbis et al. (2014) ne remarquent pas non plus de différence significative de F0 entre les productions de phrases lues et de discours semi-spontanés de 35 garçons et 36 filles âgé·e·s de 8 à 10 ans. Ces résultats corroborent ceux de Cornut et al. (1971) sur la voix parlée d'enfants de 5 à 9 ans en français.

Mentionnons deux études aux résultats inattendus et contradictoires avec ce que nous avons vu précédemment : Sachs et al. (1973) trouvent que les garçons ont une F0 significativement supérieure à celles des filles. Toutefois ces résultats doivent être considérés comme anecdotiques étant donné le faible échantillon testé par rapport à la grande étendue des âges étudiés (14 garçons et 12 filles de 4 à 14 ans comparés par paire). Une étude brésilienne a obtenu des résultats similaires sur 50 enfants âgé·e·s de 6 à 8 ans. La F0 des garçons était significativement supérieure à celle des filles dans chacune des trois tranches d'âges observées. Mais les explications avancées par les auteur·ice·s pour expliquer ces résultats inattendus sont farfelues et sexistes : « L'étude a mis en évidence des valeurs de F0 distinctes pour les deux sexes. Ce résultat peut avoir une explication d'origine physiologique, en raison du rythme différent de maturation sexuelle pour les garçons et les filles, ou comportementales, car les garçons pratiquent des activités physiques et sportives qui demandent beaucoup d'énergie et qui génèrent de l'hypertension cervicale, provoquant

1. « Il a été suggéré que les garçons tentent de produire une voix et façon de parler plus masculines en adoptant une tonalité plus basse que leur tonalité optimale, en abaissant leur mâchoire et en modifiant l'ampleur de l'arrondi des lèvres. » (Traduction de l'autrice)

ainsi une contraction musculaire excessive des muscles de la phonation. Cette contraction influence la qualité vocale et caractérise une voix avec une fréquence plus aiguë.» (Braga et al. 2009, p. 124¹)

Les résultats *a priori* contradictoires que nous avons énumérés sont peut-être à imputer à la nature de la tâche phonatoire exécutée par les enfants. En effet, toutes les études qui mettent en avant une différence significative de F0 en fonction du genre, à l'exception de S. P. Whiteside et al. (2000), reposent sur l'étude de voyelles isolées. En général, de telles données sont obtenues en demandant aux enfants de prononcer une voyelle hors de tout contexte phonologique pendant deux secondes avec leur « voix normale ». Soltani souligne pourtant que « *in clinical setting, speaking fundamental frequency (Sf0) explored from reading and spontaneous speech has been recommended more than F0 explored from sustained vowel* » (Soltani et al. 2014, p. 276). Il est possible que lorsqu'on demande aux enfants de prononcer une voyelle isolée, les petits garçons aient tendance à placer leur voix plus bas et les petites filles, plus haut. Cela pourrait correspondre à une observation que font Cornut et al. (1971). En étudiant la note sur laquelle les enfants entonnent spontanément la mélodie populaire *Frère Jacques*, ils remarquent, qu'entre 7 et 9 ans, les filles posent leurs voix plus haut que les garçons. On peut donc penser que, pour les tâches phonatoires autres que la parole, les garçons utilisent spontanément une voix plus grave que les filles mais que, lors de l'exercice de la parole, les filles et les garçons adoptent une F0 similaire, cohérente avec la zone de confort de leur appareil vocal ; ou encore que la plage de variation de la F0 de la parole entraîne un recouvrement important des F0 des garçons et des filles.

1.4.2 Influence de l'intensité

Brockmann-Bauser et al. (2015) montrent dans une étude de 2005 que l'intensité du son produit peut avoir un impact significatif sur des paramètres spectraux chez les enfants. Iels analysent la production de voyelles isolées de 39 filles et 29 garçons âgé·e·s de 5 à 9 ans vivant à Zurich. Iels testent les voyelles une première fois avec un niveau sonore spontané (76,2 dBA en moyenne dans les enregistrements effectués) et une seconde

1. « *No estudo realizado, encontrou-se valores distintos de F0 nos dois sexos. Este achado pode ser explicado com base em fatos orgânicos, devido ao ritmo diferente de amadurecimento sexual para meninos e meninas ou comportamentais, pois meninos utilizam atividades lúdicas e desportivas bastante energéticas que geram hipertensão cervical provocando, como consequência, contração muscular excessiva dos músculos da fonação. Esta contração influencia a qualidade vocal caracterizando a voz com um pitch mais agudo.* » (Traduit par l'auteurice.)

fois en demandant aux enfants de parler avec un niveau sonore supérieur à 80 dBA (les enregistrements obtenus ont un niveau moyen de 91,4 dBA). Avec le premier type de phonation, la F0 moyenne du groupe de filles est supérieure à celle du groupe de garçons de 15,2 Hz. Cette différence disparaît pour les voyelles d'intensité supérieure à 80 dBA, il n'existe plus qu'1 Hz de différence entre les F0 moyennes de deux groupes. Ces résultats peuvent laisser penser que parler à un niveau sonore supérieur au niveau de confort tend à brouiller certains indices spectraux du genre dans la voix des enfants.

1.5 Les formants en fonction du genre chez les enfants

Dans une étude pionnière, Sachs et al. (1973) trouvent que les deux premiers formants des voix de garçons sont significativement inférieurs à ceux des voix de filles pour deux des trois voyelles testées (les voyelles [i] et [u]). Une étude ultérieure plus complète obtient des résultats concordants : Bennett (1981) se penche sur la position des formants de cinq voyelles extraites de phrases de l'anglais américain, prononcées par 42 enfants de 7 et 8 ans. Elle trouve que, toutes voyelles confondues, les valeurs formantiques des filles sont plus élevées que celles des garçons respectivement de 9,7%, 9,8%, 8,5% et 10,0% pour F1, F2, F3 et F4. Elle détaille ses résultats concernant les deux premiers formants dans la figure 9.

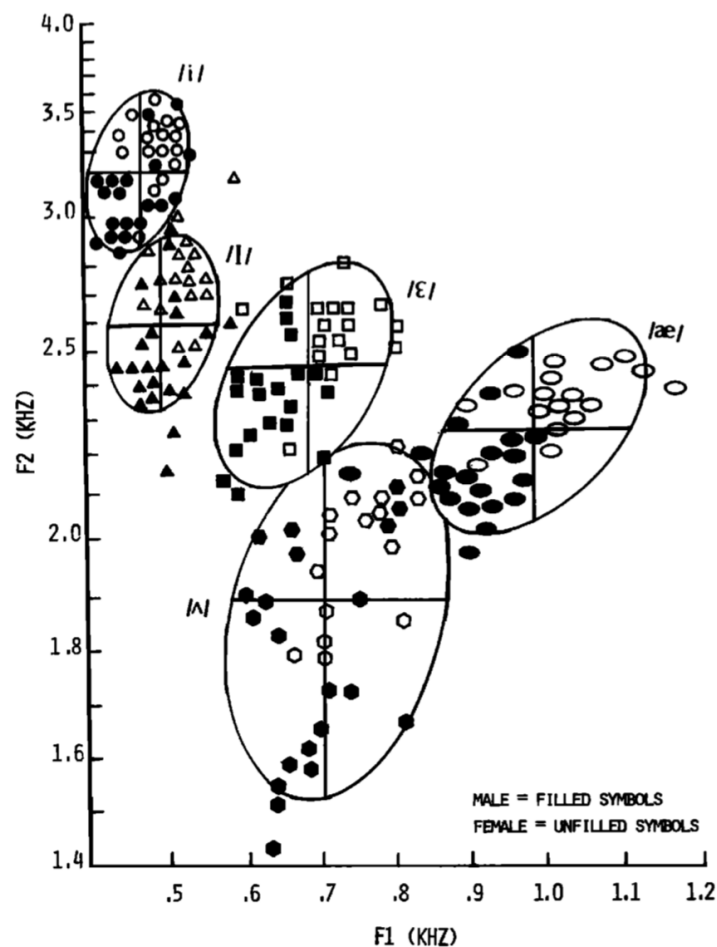


FIGURE 9 – Répartition de F1 et F2 de filles (symboles creux) et de garçons (symboles pleins) (Bennett 1981)

On observe que, pour chaque voyelle, les valeurs des deux premiers formants se regroupent dans des zones qu'on peut raisonnablement distinguer en fonction du genre, bien que certain·e·s valeurs isolées soient relativement éloignées des tendances attendues. L'autrice trace des ellipses qui délimitent les zones où se trouve la majorité des valeurs formantiques de chaque voyelle. Au sein de chacune de ces ellipses, elle trace des lignes qui marquent le mieux visuellement la différence de position des formants en fonction du genre. La ligne verticale indique la démarcation entre les voix de fille et de garçon pour la valeur de F1 et la ligne horizontale, pour celle de F2. Les positions de formants observées pour les voix de filles ont tendance à se regrouper dans le quart supérieur droit des ellipses et inversement pour les garçons ; on peut ainsi se rendre compte visuellement du taux de recouvrement des valeurs de F1 et de F2 pour chaque voyelle entre les voix de garçon et de fille. L'autrice remarque que l'écart moyen entre les formants des filles et des garçons dépend largement de la voyelle étudiée, en particulier en ce qui concerne F1. En effet, il varie de 2,5% pour la voyelle fermée [i], à 16,2% pour la voyelle pré-ouverte [æ]. Si bien que l'autrice conclut que l'existence d'une distinction de position des formants en fonction du genre n'est pas avérée pour les voyelles ouvertes puisque les différences mises en évidence sont inférieures à la marge d'erreur potentielle de mesure. Les différences rapportées sont analogues à celles observées chez les adultes mais elles sont moins marquées. Par exemple, les formants des femmes sont plus élevés en moyenne de 18% que ceux des hommes d'après les travaux de Peterson et al. (1952) sur la langue anglaise, contre 10% chez les enfants étudié·e·s par Bennett (1981). L'autrice a cherché à estimer l'influence de la taille du conduit vocal sur la position des formants et sur les différences constatées entre les genres. Pour cela, elle a collecté la taille en position debout et assise, le poids et la circonférence du cou de chaque enfant enregistré·e. Elle constate que l'ensemble de ces facteurs contribuent à la variance de la position des formants à hauteur de 26% à 75% en fonction du formant et de la voyelle considérés. Bennett explique que les comportements articulatoires des locuteur·ice·s pourraient également être à l'origine d'une partie des différences spectrales observées entre les genres. Elle suggère, par exemple, que lors de la prononciation de la voyelle [æ], les locuteurs masculins pourraient avoir tendance à accentuer l'arrondissement de leurs lèvres pour allonger la taille de leur conduit vocal total et ainsi diminuer la hauteur des formants de la voyelle.

Dans l'ensemble, ces résultats ont été confirmés par des études ultérieures. Busby et al. (1994) étudient quatre groupes paritaires composés de 10 enfants chacun, âgé·e·s respectivement de 5, 7, 9 et 11 ans ; les enfants prononcent des mots familiers monosyllabiques qui contiennent onze voyelles différentes de l'anglais australien. Les auteurs observent qu'à

tous les âges, les valeurs de F2 des filles sont significativement supérieures à celles de garçons pour 10 des voyelles étudiées ; pour F1 c'est le cas pour seulement 5 des voyelles étudiées. Ils observent également que les valeurs des trois premiers formants diminuent avec l'âge d'une façon significativement différente en fonction du genre. S. P. Whiteside et al. (2000) font des observations similaires ; elles étudient des phrases issues de descriptions d'images d'enfants de 6, 8 et 10 ans et constatent que les formants des filles sont significativement plus élevés que ceux des garçons et que le genre a un impact significatif sur l'évolution de ces valeurs avec l'âge. Perry et al. (2001) étudient des voyelles prononcées par des enfants de 4, 8, 12 et 16 ans. Ils observent qu'à 4 ans, il n'existe qu'une petite différence de position pour les deux premiers formants en fonction du genre et que le troisième formant est plus bas chez les garçons que chez les filles ; à partir de 8 ans, tous les formants sont significativement plus bas chez les garçons. Ils observent aussi que la valeur des formants de tous les enfants diminue avec l'âge mais que l'écart entre les valeurs formantiques des filles et des garçons se creuse à mesure qu'ils grandissent. Leurs résultats concordent avec ceux de Bennett (1981) puisqu'ils trouvent qu'à 8 ans, la différence moyenne entre les formants des filles et des garçons est de 9% et que cette différence persiste même après avoir pris en compte l'effet des différences de taille et de poids. Ils concluent que cette dichotomie de genre est liée soit à une différence de longueur du conduit vocal qui ne serait pas reflétée par la taille et le poids des enfants, soit par des pratiques articulatoires distinctes en fonction du genre.

Seuls S. Lee et al. (1999) obtiennent des conclusions différentes. En effet, ils étudient une population d'enfants âgé·e·s de 5 à 17 ans et concluent que la différenciation des valeurs formantiques en fonction du genre s'opère, tout comme la différenciation de la F0, à partir de 11 ans. Pourtant ils mentionnent que, chez les jeunes enfants, la valeur du premier formant est significativement différente entre les garçons et les filles, malheureusement, sans préciser les âges concernés. D'après les données présentées dans leur étude, on calcule que F1 est plus élevé dans les voix de filles de 7,75% en moyenne à l'âge de 8 ans.

Enfin, citons Cartei et al. (2014) qui étudient le positionnement relatif des formants chez 15 garçons et 19 filles de 6 à 9 ans. Pour cela, iels définissent la variable « *formant spacing* » que nous traduisons par espacement formantique. Ce paramètre est dérivé de la différence entre les positions fréquentielles des trois premiers formants, il est proportionnel à la longueur apparente du conduit vocal ; plus le conduit vocal est long, plus l'espacement formantique est réduit. Dans le cadre de leur étude, iels constatent que l'espacement formantique est significativement plus restreint pour les garçons et sa valeur a tendance à se réduire avec l'âge mais pas de manière significative sur la population testée.

En ce qui concerne le suédois, White (1999) constate aussi que les filles ont des fréquences formantiques plus élevées que les garçons pour des voyelles prononcées hors de tout contexte phonologique. Guzman et al. (2014), étudient le spectre de voyelles isolées, d'un texte lu et d'une chanson, prononcés par un groupe paritaire de 30 enfants de 7 à 10 ans. Ils ne constatent aucune différence significative de position des formants en fonction du genre à l'exception de F3 pour la voyelle [i :].

1.6 Autres paramètres acoustiques des voix d'enfants

En ce qui concerne les autres paramètres fréquentiels, ils ont été peu étudiés chez les enfants en fonction du genre. Citons tout de même Nissen et al. (2005) qui se sont intéressés à la zone de bruit des consonnes fricatives sourdes de l'anglais américain [f], [θ], [s] et [ʃ], produites par des enfants âgé-e-s de 3 à 6 ans. Ils étudient, entre autres, le gradient spectral, qui est un indicateur de la distribution de l'énergie spectrale ; le kurtosis spectral, qui est un indicateur de l'aplatissement de la distribution de l'énergie spectrale et la moyenne spectrale. Ils trouvent que, tous âges confondus, les petites filles ont un gradient spectral moyen significativement plus élevé que ceux des garçons pour toutes les fricatives étudiées sauf le [s]. Une différence significative du kurtosis spectral en fonction du genre a pu être montrée uniquement pour la consonne [s]. Quant à la moyenne spectrale, elle est significativement plus élevée chez les filles pour la voyelle [ʃ] et significativement plus élevée chez les garçons pour les consonnes les plus antérieures ([f], [θ], [s]), ce qui n'est pas le cas dans les voix adultes. En ce qui concerne la plage de variation de la F0, Pribuisiene et al. (2011) constatent qu'entre 6 et 13 ans, le genre n'a pas d'impact sur l'étendue de la tessiture des enfants. L'acquisition de l'intonation se fait graduellement tout au long du développement de l'enfant ; si les principaux traits intonatifs sont acquis dès 5 ans en moyenne, la maîtrise de l'intonation continue de se développer jusqu'à 8 ou 9 ans (ibid.). On constate donc une grande variabilité intra-sujet de la pratique intonative ce qui rend difficile l'étude de l'influence du genre sur ce paramètre chez les enfants.

Au sujet des paramètres purement temporels, S. P. Whiteside et al. (2004) étudient le VOT des 6 plosives¹ dans deux contextes vocaliques différents de l'anglais britannique chez des enfants de 5 à 13 ans. Les VOT se sont avérés significativement plus longs chez les filles des groupes d'âges de 7, 9 et 13 ans et le phénomène est plus accentué dans les groupes les plus âgés. Ajoutons que la différence de VOT liée au genre est impactée par le contexte vocalique, par exemple elle est plus accentuée pour une plosive suivie de la voyelle [i]. S. Lee et al. (1999) observent que le genre n'a aucun effet sur la durée des énoncés ni sur la durée des voyelles.

1. Plosive est une autre appellation pour les consonnes occlusives.

1.7 La perception du genre des enfants par leur voix

1.7.1 État des lieux

Weinberg et al. (1971) demandent à des adultes d'identifier le genre de 66 enfants âgé·e·s de 5 à 6 ans à partir d'enregistrements de leurs discours semi-spontanés. Iels rapportent une identification correcte à 74%. Marshal, dans une étude non publiée datant de 1972, citée par Weinberg et al. (1979b) font état d'une identification correcte à 69% du genre d'enfants de 8 et 9 ans. Sachs et al. (1973), observent une identification correcte à 81% au sein d'un échantillon restreint d'enfants âgé·e·s de 4 à 14 ans. Meditch (1975) étudie la perception de la voix de 6 filles et 5 garçons âgé·e·s de 3 à 5 ans et constate une identification correcte du genre de 79%. Weinberg et al. (1979b) comparent les taux d'identifications correctes du genre de 73 enfants de 6 et 7 ans, en fonction du stimuli présenté. Iels observent une perception correcte à 65,3% à l'écoute de voyelles isolées voisées ; de 65,9% à l'écoute de voyelles isolées chuchotée et de 68,1% pour une phrase récitée. Ingrisano et al. (1980) rapportent un score de 70,7% à l'écoute de phrases répétées par 36 enfants âgé·e·s de 4 et 5 ans. Perry et al. (2001) observent qu'à 4 ans, les garçons sont correctement identifiés à 67% contre 62% pour les filles ; à 8 ans, ils le sont à 74% contre 56% pour les filles ; à 12ans, à 82% contre 56% pour les filles et enfin à 16 ans, à 99,7% contre 95% pour les filles¹. Amir et al. (2012) étudient la perception des voix d'enfants de 8 à 18 ans². Entre 8 et 10 ans, le genre des enfants a été correctement identifié à 73,7%, à raison de 81% pour des phrases récitées en hébreu et 66,1% pour des voyelles isolées. Iels observent que les filles sont correctement identifiées en proportion significativement supérieure que les garçons entre 8 et 12 ans, cette tendance s'inverse entre 14 et 18 ans. La reconnaissance des voix de filles est constante de 8 à 18 ans, mais les voix de garçons passent d'un taux d'identification correcte d'environ 60% entre 8 et 12 ans, à environ 80% entre 14 et 18 ans. Les auteur·ice·s en concluent : « *the voices of boys after puberty carry acoustic information that enables listeners to identify their gender with absolute certainty* » (ibid., p. 319)³. Guzman et al. (2014) rapportent une identification correcte du genre de 84,55% à l'écoute d'un texte lu ; de 80,88% à l'écoute d'un passage chanté et de 63,23% pour des voyelles isolées, prononcés par 30 enfants de 7 à 10 ans. Enfin, Barreda et al. (2021) observent une reconnaissance du genre de 63% à l'écoute de

1. L'étude est menée sur des groupes de 20 enfants par tranche d'âge.

2. L'étude est menée sur 120 enfants, répartis en groupe de 20 par tranche d'âge.

3. « les voix des garçons après la puberté portent des informations acoustiques qui permettent aux auditeur·ice·s d'identifier leur genre avec une certitude absolue » (Traduction de l'autrice)

voyelles isolées prononcées par des enfants de 5 à 8 ans et constatent que ce pourcentage augmente avec l'âge des locuteur·ice·s.

Toutes les études précédemment citées concernent la langue anglaise, à l'exception de Amir et al. (2012). Notons que les méthodologies adoptées diffèrent largement d'une étude à l'autre : l'âge étudié, la répartition par tranche d'âge et la nature du stimuli écouté (phrase lue ou répétée, voyelle isolée extraite d'un contexte de phrase ou prononcée hors de tout contexte phonologique, type de voyelle choisie ou discours semi-spontané). Ceci rend la comparaison de leurs résultats peu pertinente.

Weinberg et al. (1979b) note que les moyennes présentées masquent les résultats sur le plan individuel ; certains enfants sont systématiquement correctement identifié·e·s tandis que d'autres sont perçu·e·s comme ambigu·ë·s. Iels expliquent¹ :

« The fact that approximately 25% of these subjects were identified with rates of accuracy in excess of 97% correct in each of the vowel conditions provides convincing evidence to suggest that the sexual identity of some preadolescent boys and girls is as clearly defined as that observed in adult populations »
(Weinberg et al. 1979b, p. 187)

Barreda et al. (2021) font la même observation que Weinberg et Bennett, ils constatent que seules certaines voix d'enfants sont perçues comme ambiguës.

Dans une recherche particulièrement originale, Edwards (1979) s'intéresse à l'identification du genre des enfants en fonction de leur classe sociale. Il étudie des phrases lues par deux groupes paritaires de 20 enfants chacuns, l'un composé d'enfants de la classe ouvrière et l'autre, d'enfants de la classe moyenne. Le genre des enfants est correctement identifié à plus de 80% mais l'auteur s'intéresse à la répartition des erreurs. Il observe que, pour les enfants de la classe ouvrière, les auditeur·ice·s ont tendance à faire plus d'erreurs d'identification parmi les filles ; à l'inverse iels identifient plus souvent à tort les garçons de la classe moyenne comme des filles. De plus, les enfants issu·e·s de la classe ouvrière sont perçu·e·s comme ayant des voix plus graves et plus masculines que celles des enfants de la classe moyenne, de manière significative. L'auteur remarque² :

1. « Le fait qu'environ 25% de ces sujets aient été identifiés avec des taux de bonnes réponses supérieurs à 97% dans chaque voyelle constitue une preuve convaincante que l'identité de genre de certain·e·s préadolescent·e·s est aussi clairement définie que celle observée dans les populations adultes. » (Traduction de l'autrice)

2. « Ces résultats sont liés aux travaux démontrant une association entre la masculinité et l'accent de classe de la classe ouvrière, et au « prestige secret » qui s'attache à cet accent en conséquence. » (Traduction de l'autrice)

« *These findings are related to work demonstrating an association between masculinity and working-class speech, and to the 'covert prestige'^a attaching to such speech as a result.* » (Edwards 1979, p. 121)

a. Cette notion, qui se traduit littéralement par *prestige secret*, est définie comme « un phénomène selon lequel des usages non standards de la langue commune ou des dialectes sont considérés comme prestigieux par les membres d'une communauté linguistique » (définition issue de *Wikipedia*, traduite par l'autrice : https://en.wikipedia.org/wiki/Covert_prestige). Ici elle concerne l'association d'un l'accent de classe populaire à la notion de masculinité, le prestige associé à l'impression de virilité attirerait les locuteurs masculins qui auraient tendance à adopter cette façon de parler afin de jouir de l'effet qu'elle produit sur leurs interlocuteur·ice·s.

L'auteur précise que les auditeur·ice·s ayant participé à l'étude sont elles-eux-mêmes issu·e·s de la classe moyenne et qu'il serait intéressant de mener une étude comparable avec des participant·e·s issu·e·s de classes populaires. À notre connaissance, aucune étude similaire n'a été menée en langue française.

1.7.2 Les indices utilisés pour identifier le genre

Les études qui se sont attachées à comparer les performances d'identification en fonction du stimulus présenté trouvent que les scores sont plus élevés à l'écoute de phrases que de voyelles isolées (Sachs et al. 1973, Amir et al. 2012 et Guzman et al. 2014) ce qui pourrait corroborer l'hypothèse de Barreda et al. (2021), évoquée partie **I.2.2**, selon laquelle l'identification du genre s'appuierait sur un critère moyenné des données spectrales plutôt que sur la valeur discrète de certains paramètres fréquentiels. Cela pourrait aussi indiquer que les auditeur·ice·s utilisent des indices d'ordre temporels pour identifier le genre des enfants : contour intonatif, VOT ou durée des énoncés par exemple. C'est ce qu'Amir et al. (2012) suggèrent en conclusion de leur étude³ :

3. « Il semble que le signal vocal des jeunes garçons porte des informations acoustiques que les auditeur·ice·s peuvent utiliser pour identifier leur genre. Ces informations acoustiques ne sont pas présente (ou moins prononcée) dans les voyelles isolées. Le signal vocal fournit des indices temporels de coarticulation et de prosodie, au-delà de la valeur absolue de la fréquence fondamentale et des indices supraglottaux non-temporels, qui sont présents dans les voyelles isolées. Par conséquent, on suppose que les garçons effectuent des ajustements de leur conduit vocal et de leur articulation au cours de la parole, ce qui permet aux auditeur·ice·s de les identifier correctement. » (Traduction de l'autrice)

« Apparently, the young boys' speech signal carries acoustic information that listeners can use for identification of their gender. This acoustic information is not presented (or less pronounced) in isolated vowels. The speech signal provides coarticulation and prosody dynamic cues, beyond the fundamental frequency and the static supraglottal cues, which are provided in isolated vowels. Therefore, it is assumed that boys are making vocal tract and articulatory adjustments during speech, which enables listeners to identify them correctly. »
(Amir et al. 2012, p. 320)

Amir et al. remarquent que la taille des locuteur·ice·s est associée très significativement à l'identification du genre mais que l'âge, le poids et la circonférence du cou n'ont pas d'interaction notable avec la perception du genre. Weinberg et al. (1971) examinent les F0 des 20 garçons et 29 filles qui ont été systématiquement correctement identifié·e·s. Elles se recouvrent largement entre les deux groupes ce qui mène les auteur·ice·s à conclure que la F0 n'est pas un indice que les auditeur·ice·s utilisent pour identifier le genre. Ingrisano et al. (1980) choisissent également de mener une analyse acoustique des voix de 4 garçons et 3 filles dont le genre a été correctement identifié à plus de 90%. Ils observent une F0 moyenne plus basse et une plage de variation de la F0 plus réduite chez les garçons ; un débit de parole plus lent et plus variable chez les garçons ; des F1 et F2 plus élevés et un triangle formantique plus étendu chez les filles et enfin une différence de contour d'amplitude. Bien que l'amplitude moyenne soit similaire en fonction du genre, les garçons étudiés ont tendance à augmenter l'intensité de leur parole au cours de la phrase alors que la tendance inverse s'observe chez les filles. Les auteurs en concluent que les paramètres acoustiques qui jouent le plus probablement un rôle dans la perception du genre des enfants sont le débit de parole, la position des formants et le contour intonatif.

Weinberg et al. (1979b) comparent les performances d'identification pour des voyelles isolées normales et chuchotées, la différence entre les scores obtenus à l'écoute des deux modes de phonation n'est pas significative ce qui leur laisse penser que la perception du genre repose en grande partie sur la position des formants (voir la partie **I.7.2**). Weinberg et al. (1979a), reprenant les données de leur propre étude (Weinberg et al. 1979b), analysent la relation entre les données perceptives obtenues et certains paramètres spectraux des voix enregistrées. En ce qui concerne l'écoute des voyelles isolées voisées, les valeurs de F0, F1 et F2 se sont toutes avérées significativement liées à l'estimation du genre des locuteur·ice·s ; F2 ayant l'impact de plus important. Les auteur·ice·s ont mené une analyse de régression par étape avec pour variable dépendante, le nombre d'occurrences de la réponse « garçon » pour chaque stimulus. La position de F2 s'avère contribuer

à 38% de la variance totale des réponses des participant·e·s au test perceptif; la F0 moyenne y contribue à 22% et les variations de la valeur de F1 n'ont pas significativement amélioré la performance des participant·e·s. Les auteur·ice·s répartissent les locuteur·ice·s en trois catégories perceptives : enfants perçu·e·s comme des garçons, comme des filles ou ambigu·ë·s. Iels expliquent qu'une partie des F0 de chaque groupe se recouvrent pour chaque combinaison de groupe mais que, tout de même, au sein des deux groupes non ambigus, environ un tiers des F0 ne se recourent clairement pas avec les valeurs de F0 des deux autres groupes. Cette répartition explique, selon les auteur·ice·s que la F0 moyenne ne soit pas un très bon indicateur indépendant du genre d'une voix pré-pubère mais qu'elle participe malgré tout significativement au processus d'identification du genre. En ce qui concerne l'écoute des phrases, la position des formants est significativement liée à l'évaluation du genre des locuteur·ice·s mais pas la F0 moyenne et, cette fois-ci, c'est la valeur de F1 qui a le plus de poids dans le choix des participant·e·s au test perceptif. En effet, F1 cause 48% de la variance totale contre seulement 6% pour F2. Les auteur·ice·s en déduisent que les formants vocaliques constituent un indice majeur pour l'identification du genre et que la F0 moyenne ainsi que certains paramètres intonatifs (comme la plage de variation de la F0, le taux de variation de fréquence lors d'une inflexion de F0 ou la durée moyenne d'une inflexion de F0) peuvent s'avérer significatifs mais sont secondaires lorsqu'il s'agit de déterminer le genre d'un·e locuteur·ice pré-pubère.

Perry et al. (2001) arrivent à une conclusion similaire¹ :

« Adults' perception of gender from children's speech and voice is strongly related to formant frequencies for children aged 4 to 12 years, whereas f0 plays a key role by 16 years. » (Perry et al. 2001, p. 2997)

En effet leurs résultats montrent que les auditeur·ice·s s'appuient considérablement sur la valeur de F3 pour identifier le genre des enfants de 4 ans; sur celles de F1, F2 et F3 pour identifier celui des enfants de 8 à 12 ans, en particulier F1 en ce qui concerne les enfants de 12 ans et enfin qu'ils utilisent les informations de la F0 ainsi que des trois premiers formants pour assigner un genre aux locuteur·ice·s de 16 ans.

Quant à Cartei et al. (2013), iels entreprennent une étude pionnière dans laquelle iels opèrent une resynthèse de la voix de deux filles et deux garçons de 8 ans prononçant des mots monosyllabiques usuels, en modifiant notamment la position de leurs formants. Iels cherchent à étudier l'impact de l'espacement formantique sur la perception du genre

1. « La perception du genre par les adultes à l'écoute des voix d'enfants est fortement liée aux fréquences des formants pour les enfants âgés de 4 à 12 ans, alors que la F0 joue un rôle clé à partir de 16 ans. » (Traduction de l'autrice)

dans les voix pré-pubères. Pour cela, iels normalisent d’abord les F0 de chaque enregistrement à 260Hz et aplatissent le contour intonatif autour de cette valeur. Les valeurs des trois premiers formants sont ensuite modifiées, de sorte à créer plusieurs resynthèses d’un même échantillon dont les espaces formantiques s’échelonnent par pas de 2%. Puis iels font écouter les voix hybrides obtenues à des adultes. Dans un premier temps, iels leur demandent de classer le genre des locuteur·ice·s entre *garçon* et *fille*. Dans un second temps, iels leur demandent d’évaluer l’expression de genre perçue sur une échelle discrète de 0 à 7 labellisée de « *garçon masculin* » à « *fille féminine* ».

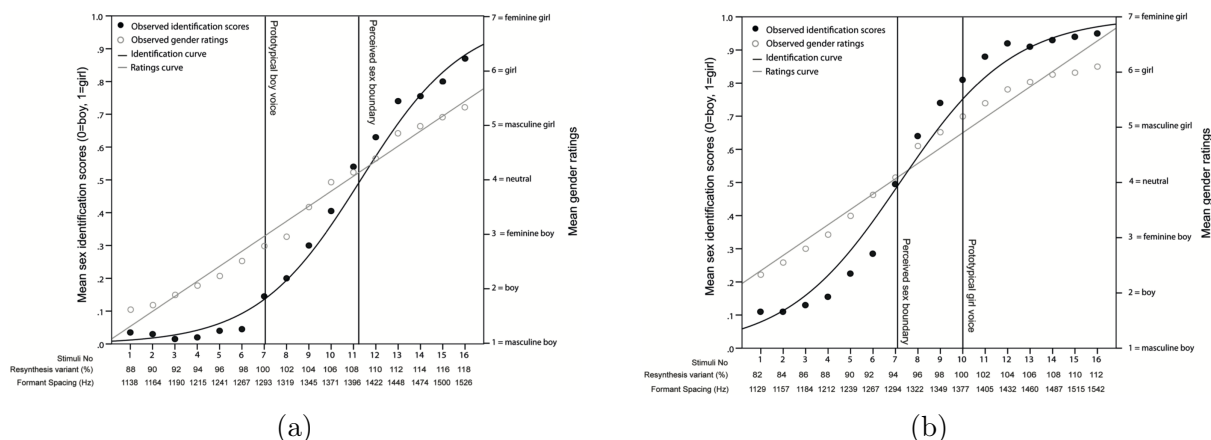


FIGURE 10 – Identification du genre et évaluation de l’expression de genre des voix resynthétisées de garçons (a) et de filles (b) en fonction de l’espace formantique

Les figures 10¹ présentent les résultats des deux expériences, obtenus respectivement à l’écoute des voix de garçons (à gauche) et des voix de filles (à droite) resynthétisées.

1. La figure (a) est légendée ainsi dans l’étude originale : « Scores were averaged across listeners on voice stimuli (numbered 1–16 on the x-axis) for the boys’ exemplars. The mean identification scores are plotted from 0 = male to 1 = female (left y-axis) and fitted with the logistic curve (black line). The vertical lines illustrate the location of the estimated sex boundary (where 50% of the listeners rate the stimuli as female) and the location of the prototypical boy voice stimulus (100%). The percentage of stimuli identified as female follows an S-shaped pattern along the continuum of resynthesis variants. The sex identification curve is characterised by a lower plateau for stimuli 1 to 6 (DFs of 1138–1267 Hz), where less than 10% of the stimuli are identified as female, indicating that stimuli variant with the lowest DF are mostly identified as male. The percentage of stimuli identified as female then increases gradually and linearly, and while no upper plateau is reached, average scores for stimuli 14 to 16 (DFs of 1474–1526 Hz) varied from 76% to 85%, indicating that boys’ voices with the highest DF are mostly classified as female. Average gender rating scores are plotted from 1 = masculine boy (or girl) to 7 = feminine boy (or girl) (right y-axis) and fitted with a linear function (straight grey line). Mean gender ratings of male voices ranged from 1.78 (SE = .07) for the lowest DF variants to 5.36 (SE = .08) for the highest DF variants. » (Cartei et al. 2013, p. 4). La figure (b) a une description analogue, le plateau bas de la courbe d’identification du genre s’étend entre les stimuli 1 et 3 (DFs of 1129–1184 Hz), et, pour ce plateau entre 10% et 15% des stimuli ont été reconnus comme produits par une fille. Le plateau haut, lui, s’étend entre les stimuli 12 et 16 (DFs of 1432–1542 Hz), et, pour ce plateau entre 92% to 95% des stimuli on été reconnus comme

En ce qui concerne la première expérience perceptive, les résultats sont remarquablement linéaires. Dans les deux groupes, la proportion de la réponse *fille* semble quasiment proportionnelle à l'augmentation de la valeur de l'espacement formantique. La frontière perceptive de « voix de garçon » à « voix de fille », valeur à laquelle 50% des auditeur·ice·s désignent l'échantillon comme une voix de fille, se situe entre les resynthèses 11 et 12 pour les voix de garçons modifiées et entre les resynthèses 7 et 8 pour les voix de filles modifiées. Les auteur·ice·s identifient des voix de garçon et de fille « prototypiques » qui ont respectivement un espacement formantique d'environ 1293 Hz et de 1377 Hz. Le facteur « variation du stimuli présenté » s'est révélé avoir un impact significatif sur le genre perçu à la fois pour les resynthèses de voix de garçons et pour celles des voix de filles. De même, l'évaluation de l'expression de genre passe de *garçon masculin* à *fille féminine* à mesure que l'espacement formantique augmente et les deux variables ont un lien significatif. Les auteur·ice·s concluent que l'espacement formantique est un indice majeur pour la discrimination du genre dans les voix d'enfants. Toutefois, iels remarquent que les effets de la resynthèse sur la perception du genre n'ont pas été exactement les mêmes pour les voix des filles et des garçons, en particulier la frontière perceptive entre « voix de garçon » et « voix de filles » est plus basse pour les voix de filles modifiées que pour les voix de garçons. Ceci laisse à penser que d'autres indices de genre subsistent malgré les resynthèses, comme la qualité de voix. De plus, bien que la décision d'uniformiser la variation de la F0 soit compréhensible, le choix d'aplatir le contour intonatif est contestable. En effet, Weinberg et al. (1979b) ont montré qu'une phrase prononcée avec un contour intonatif plat est plus souvent perçue comme ayant été prononcée par un garçon par rapport à un phrase ayant été prononcée avec le contour intonatif spontané du·de la locuteur·ice (ces dernier·e·s ayant entre 6 et 7 ans). Il est donc possible que dans l'étude de Cartei *et al*, le choix d'aplatir le contour intonatif ait eu pour effet d'augmenter la proportion d'identification *garçon*.

1.7.3 Rôle de l'âge

Dans un article datant de 2018 à propos de la capacité à identifier l'âge d'un·e locuteur·ice à l'écoute de sa voix, Barreda et Assmann ont constaté que la perception de l'âge est conjointe à une estimation du genre du·de la locuteur·ice (Barreda et al. 2018). Ils

produits par une fille. Les scores moyens d'évaluation d'expression de genre varient de 2,33 (SE = 0,02) pour l'espacement formantique le plus bas, à 6,10 (SE = 0,06) pour le plus haut.

expliquent¹ :

« Results suggested that both sets of listeners used acoustic cues in a gender-dependent manner, regardless of whether they were provided with information about talker gender. In cases where talker gender was not explicitly provided, listeners appeared to ‘guess’ the gender of the talker and adjust their use of acoustic cues. » (Barreda et al. 2021, p. 3950)

Dès lors on peut se demander si la réciproque est vraie. Dans l'expérience que Barreda et Assmann mènent en 2021, ils constatent que le fait d'indiquer l'âge des locuteur·ice·s aux participant·e·s avant de leur demander d'évaluer leur genre fait baisser légèrement leurs performances (identifications correctes moins fréquentes) mais le phénomène n'est pas très marqué. Les auteurs concluent² :

« Providing listeners with information about talker age did not have a strong effect on listener sensitivity or bias ; however, all listeners appeared to use talker age information in their judgments of talker gender. » (Barreda et al. 2021, p. 3956)

1.7.4 Le degré de certitude

Le critère de la confiance de l'auditeur·ice dans son identification du genre n'est malheureusement que très peu étudié. Sachs (1975) indique tout de même que les participant·e·s au test perceptif se plaignent d'avoir le sentiment de deviner au hasard, ce qui trahit un degré de certitude très faible dans l'examen du genre. Perry et al. (2001) demandent aux participant·e·s de leur test perceptif d'évaluer le genre des enfants sur l'échelle suivante :

1. « Les résultats suggèrent que les deux groupes d'auditeur·ice·s utilisent les indices acoustiques en fonction du genre, qu'ils aient reçu ou non des informations sur le genre du·de la locuteur·ice. Dans les cas où le genre du·de la locuteur·ice n'était pas explicitement indiqué, les auditeurs semblaient " deviner " le genre et adapter leur utilisation des indices acoustiques. » (Traduction de l'autrice)

2. « Le fait de fournir aux ·ice· des informations sur l'âge du·de la locuteur·ice n'a pas eu un effet important sur la sensibilité ou le biais des auditeur·ice·s ; cependant, tou·te·s les auditeur·ice·s ont semblé utiliser les informations sur l'âge du·de la locuteur·ice dans leur examen sur le genre. » (Traduction de l'autrice)

1. « assurément une femme
2. semble être une femme
3. incertain, peut-être une femme
4. incertain, peut-être un homme
5. semble être un homme
6. assurément une femme »

(Perry et al. 2001, p. 2993-2994)

Les moyennes des réponses obtenues respectivement pour les garçons et les filles dans chaque tranche d'âge de 4 à 12 ans sont comprises entre 3 et 4,5. Les auditeur·ice·s semblent donc mal assuré·e·s des réponses qu'ils donnent. Les moyennes de réponses obtenues pour les garçons et les filles de 16 ans se rapprochent quant à elles de l'extrémité de l'échelle proposée par Perry et al, ce qui indique un degré de certitude supérieur de la part des participant·e·s pour cette tranche d'âge. Les données obtenues pour des voix d'enfants semblent donc se distinguer de ce que l'on observe pour l'identification du genre pour des voix pubères. En effet, lorsqu'il s'agit de voix adultes, les participant·e·s témoignent d'une confiance très solide dans leur jugement. Pepiot2013, par exemple, trouve un degré de certitude moyen de 6,54 (sur une échelle discrète allant de 0 à 7) à l'écoute d'une voyelle isolée produite par des locuteur·ice·s pubèr·e·s ; ce score monte à 6,92 à l'écoute d'un mot dissyllabique.

En conclusion, bien que les voix d'enfants semblent perçues comme plus ambiguës en ce qui concerne le genre que les voix d'adultes, les auditeur·ice·s sont capables d'identifier le genre des locuteur·ice·s dans des proportions qui dépassent la chance. Bien que la position relative des formants vocaliques semble jouer un rôle prépondérant dans l'identification du genre par la voix dès le plus jeune âge, les indices acoustiques sur lesquels s'appuie la discrimination du genre avant la puberté semblent multiples et ne sont pas encore identifiés avec certitude.

1.8 La conscience des rôles de genre et la perception de la voix par rapport au genre chez les enfants

De nombreuses recherches ont mis en évidence l'élaboration d'une constellation de représentations stéréotypiques de genre chez les enfants avant l'âge de 5 ans, qu'ils sont ensuite susceptibles de s'appliquer à elles-eux-même et aux autres (Halim et al. 2013). Dès l'âge de 8 ans, les enfants démontrent une capacité à réaliser des inférences reposant sur ces stéréotypes. Par exemple le fait qu'un enfant de genre inconnu aime s'amuser avec un jouet « kit d'outils » les amène à déduire que cet enfant sera susceptible de préférer porter un costume cravate (Martin et al. 1990). La façon dont la voix s'insère dans ces représentations de genre que construisent les enfants au cours de leur développement n'a, pendant longtemps, pas été étudiée. Ces dernières années, l'autrice Valentina Cartei, avec plusieurs de ses collègues, a développé un travail remarquable sur le sujet en anglais britannique qui n'a, à notre connaissance, pas encore été poursuivi par d'autres auteur·ice·s.

Cartei et al. (2021) montrent que des enfants âgé·e·s de 8 à 10 ans infèrent la capacité à occuper des postes jugés typiquement masculins, féminins ou neutres et la compétence en fonction du degré de masculinité ou de féminité perçu dans des voix d'adultes. Les voix utilisées pour mener cette expérience sont des resynthèses de voix de femme et d'homme dont la F0 moyenne a été augmentée et diminuée ; les participant·e·s sont préalablement renseigné·e·s sur le genre de chaque locuteur·ice·s. Dans une autre étude, Cartei et al. (2019a) observent que des enfants âgé·e·s de 7 à 8 ans ont tendance à associer des voix pré-pubères qui contiennent des indices acoustiques typiques du genre féminin ou masculin (ici l'espacement vocalique) à des personnages fictionnels auxquels on attribue des caractéristiques respectivement typiquement féminines et typiquement masculines (centre d'intérêt et fréquentation).

Cartei et al. (2014) suggèrent que les locuteur·ice·s utilisent certains paramètres acoustiques tels que la F0 ou l'espacement vocalique comme une façon d'exprimer leur genre à travers leur voix. Iels cherchent à caractériser la capacité qu'ont les enfants à altérer certains paramètres acoustiques de leur voix afin de performer des stéréotypes de « voix de garçons » et de « voix de filles ». Pour cela iels enregistrent 15 garçons et 19 filles âgé·e·s de 6 à 9 ans qui énoncent des mots monosyllabiques usuels de l'anglais hors d'un contexte de phrases ; une première fois avec leur voix naturelle et ensuite à nouveau, deux fois, après qu'on leur a demandé de sonner le plus possible comme un garçon puis comme une fille. Iels observent que les enfants ne modifient pas significativement leur F0 quand

iels caricaturent leur propre genre ; en revanche les filles abaissent significativement leur F0 lorsqu'elles imitent une voix de garçon et les garçons utilisent une F0 significativement plus élevée pour imiter une voix de fille. De plus, les enfants, quel que soit leur genre, réduisent significativement leur espacement vocalique lorsqu'ils contrefont une voix de garçon et, à l'inverse, l'augmentent significativement en imitant une voix de fille. Notons que, face au même exercice, les adultes adoptent une stratégie similaire vis-à-vis de l'espacement vocalique (Cartei et al. 2012). Les auteur·ice·s remarquent¹ :

« Despite the confirmed absence of sex differences in the fundamental frequency of pre-pubertal children's natural voices, both boys and girls adjusted this parameter when imitating the opposite sex in line with the sex differences present in adults. » (Cartei et al. 2014, p. 103)

Dans une étude ultérieure, Cartei et al. (2019b) obtiennent des résultats similaires au cours d'une expérience plus élaborée. Iels demandent à 36 filles et 34 garçons de 6 à 10 ans de donner leurs voix à des personnages fictionnels, en répétant trois phrases courtes. Les personnages sont du même genre que les enfants et on décrit leur caractère comme plus ou moins masculin ou féminin. Lorsqu'iels incarnent des personnages aux attributs virils, les filles et les garçons adoptent une F0 significativement plus basse et un espacement vocalique significativement plus réduit que pour les personnages aux attributs neutres et féminins. En incarnant des personnages caractérisés avec un haut degré de féminité, les filles augmentent significativement leurs F0 et élargissent significativement leurs espacements vocaliques mais ces phénomènes ne sont pas significatifs chez les garçons.

Ainsi, il semble clair non seulement que, dès l'âge de 6 ans, la voix a une place dans les représentations stéréotypées du genre dont les enfants se servent pour raisonner mais aussi que ces dernier·e·s sont capables d'altérer leur propre voix afin de se conformer aux attributs acoustiques d'un genre lorsqu'on leur demande de l'imiter. Il reste encore à déterminer s'iels sont également capables de contrôler ces paramètres dans leur voix au quotidien afin d'offrir une performance de genre à la fois identifiable pour leurs interlocuteur·ice·s et qui corresponde à l'image qu'iels souhaitent renvoyer.

1. « Malgré l'absence confirmée de différence de F0 en fonction du genre dans les voix naturelles des enfants pré-pubères, les garçons et les filles ont ajusté ce paramètre lors de l'imitation du genre opposé, conformément aux différences entre les genres présentes chez les adultes. » (Traduction de l'autrice)

Conclusion partielle

Ce chapitre nous a permis d'aborder la notion du genre dans la voix. Nous avons pu voir qu'à l'âge adulte, les voix d'hommes et de femmes se distinguent sur de nombreux paramètres acoustiques comme la F0, la position des formants, l'intonation, le VOT ou encore la qualité de voix. Ces différences caractéristiques s'expliquent en partie par un dimorphisme de l'appareil vocal mais elles semblent aussi liées à des pratiques culturelles puisqu'elles varient en fonction de la langue parlée et ce, même chez les locuteur·ice-s bilingues. L'accumulation de ces différences permettent aux auditeur·ice-s d'identifier avec une très grande fiabilité le genre d'une voix, y compris à l'écoute d'un signal très court ou dépourvu d'un certain nombre d'informations spectrales. Les informations de la F0 et de la position des formants, lorsqu'elles sont disponibles, semblent être les indices principaux permettant la discrimination du genre. Plusieurs études en langue anglaise montrent que les auditeur·ice-s sont susceptibles de distinguer les voix de garçons des voix de filles au delà de la chance. Pourtant, avant la puberté, aucune différence significative de la taille des plis vocaux ni de la taille du conduit vocal en fonction du genre n'est constatée. L'hypothèse d'une différence significative de la F0 usuelle de la parole des garçons et des filles avant la puberté, n'est pas non plus formellement démontrée. D'autres paramètres, comme la position des formants, semblent se différencier en fonction du genre en anglais, avant même la puberté. Il est possible que ces différences s'expliquent par des stratégies articulatoires que les enfants assimilent en même temps qu'ils développent des représentations stéréotypiques de voix féminines et masculines, qui font partie de modèles de compréhensions sociales plus larges. La question du développement de la production vocale par rapport au genre dans l'enfance, tant sur le plan physique que sur le plan des pratiques articulatoires, reste encore largement à explorer, notamment en comparant les données disponibles entre plusieurs langues. Par la suite, nous nous intéresserons à ces problématiques à travers la constitution et l'étude d'un corpus de voix d'enfant en langue française.

PARTIE 2

Enregistrement et analyse acoustique préliminaire d'un corpus de
voix d'enfants

Problématiques et hypothèses

Au cours des deux prochains chapitres, nous nous pencherons sur les profils spectraux des voix de filles et de garçons francophones âgé·e·s de 8 à 10 ans puis sur le processus d'identification du genre à l'écoute des voix de ces mêmes enfants. Nous décrirons comment, à l'occasion de ce travail, nous avons enregistré un corpus de voix d'enfants, nommé *CoVEP* (Corpus de Voix d'Enfants Prépubères), destiné à l'étude socio-phonétique des questions de l'expression et de la perception du genre dans la voix enfantine. Nous chercherons à apporter de premiers éléments de réponses aux questions restées en suspens à l'issue de l'étude bibliographique présentée dans le chapitre précédent : existe-t-il des différences acoustiques entre les voix de garçon et de fille francophones avant le début de la puberté ? Les auditeur·ice·s francophones ont-iels la faculté d'identifier avec fiabilité le genre des enfants à la seule écoute de leurs voix, comme cela a été démontré en langue anglaise ? Sur quels indices acoustiques s'appuie l'examen du genre des voix pré-pubères ? Dans la suite de ce travail, nous chercherons à tester les trois hypothèses suivantes :

1. Il existe, dans le corpus de voix étudiées, une différence significative de F0 entre les garçons et les filles.
2. À l'écoute des voix d'enfants âgé·e·s de 8 à 10 ans, les auditeur·ice·s francophone sont capables de reconnaître le genre au-delà du seuil de chance.
3. La F0 joue un rôle dans l'identification du genre à l'écoute des voix d'enfants francophones âgé·e·s de 8 à 10 ans.

2.1 Méthodologie

2.1.1 Enfants enregistré·e·s

Nous avons enregistré 28 enfants âgé·e·s de 8 à 10 ans. Tou·te·s ont été enregistré·e·s lors d'un temps périscolaire, dans une salle d'école qui leur était familière. Iels ont été enregistré·e·s deux par deux afin de réduire au maximum l'effet du stress. Tou·te·s les enfants vivent au Blanc-Mesnil ou aux alentours, habitent en Seine-Saint-Denis depuis au moins trois ans et sont issu·e·s d'un milieu social populaire. Seule une petite fille a signalé des problèmes d'élocution, son profil a donc été écarté. La plupart des enfants enregistré·e·s sont bilingues. Nous avons choisi de conserver les enfants qui comptent le français parmi leurs langues maternelles et qui parlent le français, entre autres langues, à la maison. Considérant que leur niveau de français égale celui d'un enfant ne parlant que le français et qu'iels correspondent à la définition d'individu·e·s bilingues, nous estimons qu'iels sont concerné·e·s par les conclusions évoquées partie **I.2.3**, selon lesquelles les locuteur·ice·s bilingues ont intégré les codes vocaux de genre propres à une langue et sont susceptibles d'adopter des stratégies articulatoires qui correspondent aux normes de genre de la langue dans laquelle iels s'expriment, au même titre que les locuteur·ice·s monolingues. Les enfants qui nous ont indiqué ne pas parler le français dans leur cadre familial ont été écarté·e·s. Les sujets de cette étude sont des locuteur·ice·s du français dit parisien, aucun accent régional autre que l'accent francilien n'a été détecté. En revanche il est à noter que nous avons pu déceler chez la plupart d'entre elles-eux un accent dit populaire. Si cela ne pose pas problème puisque l'origine sociale est homogène dans le groupe étudié, cet élément sera à prendre ne compte dans l'analyse des résultats puisque nous avons pu voir que l'accent de classe est susceptible de changer la perception du genre dans la voix (voir **I.7.2**).

Notre choix s'est porté sur la tranche d'âge de 8 à 10 ans afin d'étudier la période de l'enfance la plus proche possible du début de la puberté. Notre objectif étant d'étudier des enfants ayant à la fois largement développé des représentations stéréotypiques du genre, ayant intégré les codes vocaux relatifs à celles-ci et ayant acquis la meilleure maîtrise possible de leur propre appareil vocal. Nous voulions aussi que les enfants enregistré·e·s soient à l'aise avec l'exercice de la lecture à haute voix. Malgré tout, la maîtrise de la lecture, en particulier de l'intonation des phrases en fonction de la ponctuation, n'est pas toujours acquise à cet âge. Les niveaux de lecture étaient donc inégaux au sein du groupe enregistré; notons tout de même que les disparités d'aisance de production vocale lue ne

semblent pas liées au genre.

Il est intéressant de noter que spontanément, certains enfants ne se sont pas défini.e.s de manière binaire. On peut faire le choix d'analyser leur voix sur le plan acoustique mais il ne nous est pas possible de les intégrer au test perceptif, car leur identité ne correspond pas à la structure de test que nous avons choisi d'adopter. Ceci est regrettable et met en évidence un écueil de la conception des tests phonétiques concernant le genre. Ce problème est révélateur des limites des protocoles dont nous disposons pour étudier le genre en socio-phonétique qui sont mal adaptés aux notions qu'ils prétendent décrire. C'est une limite que pointe également Arnold (2015) à laquelle il consacre une sous-partie dans sa thèse.

Après avoir écarté les profils des enfants présentant des difficultés de lecture trop importantes, de celle-ux n'ayant pas le français comme langue maternelle ou présentant des traces d'accents de leur seconde langue qui pourrait fausser les mesures acoustiques ainsi que de celles-eux ayant rencontré des problèmes techniques lors de l'enregistrement, le corpus obtenu est composé des voix de 8 garçons et de 10 filles. Les garçons et les filles sont relativement également réparti.e-s sur la tranche d'âge étudié (5 garçons et 4 filles sont âgé.e-s de 8 ans ; 2 garçons et 2 filles sont âgé.e-s de 9 ans et 1 garçon et 4 filles sont âgé.e-s de 10 ans). Les filles étant en moyenne plus âgées que les garçons, si des valeurs spectrales significativement plus basses chez les garçons que chez les filles sont observées, la différence ne pourra pas être imputée à une différence de taille de leurs conduits vocaux.

2.1.2 Matériau linguistique enregistré

Le CoVEP est constitué de cinq parties distinctes. La première est une série de 33 pseudo-mots dissyllabiques prononcés au sein d'une « phrase capsule », toujours identique. Le recours à une « phrase capsule » permet d'éviter l'écueil de l'effet liste. En effet, à la lecture d'une liste de mot, on aura tendance à adopter un contour intonatif particulier : la F0 monte à la lecture de chaque mot pour redescendre sur le dernier item de la liste. La « phrase capsule » garantit que les mots du corpus soient prononcés avec une prosodie constante.

Nous avons choisi de tester la combinaison de plusieurs phonèmes : les trois voyelles du français [a], [u], [i] en début de mot ou bien précédées d'une des dix consonnes suivantes : [s], [z], [ʃ], [ʒ], [t], [d], [k], [g], [p], [b]. Chacune de ces différentes combinaisons était suivie de la syllabe /pi/ pour conclure le pseudo-mot dissyllabique. Les trois voyelles choisies constituent les trois pôles du triangle formantique, elles sont donc toutes indiquées

pour étudier l'étendue du triangle formantique des locuteur·ice·s. Les consonnes fricatives sourdes [s] et [ʃ] permettent d'étudier l'impact des zones de bruits en l'absence de voisement et le corpus permet de les comparer avec les deux fricatives voisées correspondantes. En effet, la position des articulateurs est identique lors des prononciations respectives de [s] et [z] et de [ʃ] et [ʒ] ; seule la présence de voisement permet de distinguer ces phonèmes deux à deux. L'étude des occlusives sourdes et voisées ([t], [d], [k], [g], [p] et [b]) permet, quant à elle, d'obtenir des informations sur le VOT. Les 33 pseudo-mots suivants ont donc été retenus :

/sipi/, /sapi/, /supi/, /zipi/, /zapi/, /zupi/, /ʃipi/, /ʃapi/, /ʃupi/, /ʒipi/, /ʒapi/, /ʒupi/, /pipi/, /papi/, /pupi/, /bipi/, /bapi/, /bupi/, /tipi/, /tapi/, /tupi/, /dipi/, /dapi/, /dupi/, /kipi/, /kapi/, /kupi/, /gipi/, /gapi/, /gupi/, /ipi/, /api/, /upi/

Ils correspondent aux transcriptions orthographiques suivantes :

sipi, sapi, soupi, zipi, zapi, zoupi, chipi, chapi, choupi, jipi, japi, joupi, pipi, papi, poupi, bipi, bapi, boupi, tipi, tapi, toupi, dipi, dapi, doupi, kipi, kapi, koupi, guipi, gapi, goupi, ipi, api, oupi

Les phrases présentées aux enfants étaient déclinées sur le modèle suivant :

« Il a dit sapi deux fois. »

La deuxième partie du CoVEP est composée de 12 phrases, dont 9 déclaratives, 2 interrogatives et 1 exclamative. Ce sont des phrases ordinaires comme « J'espère juste qu'un jour on pourra en parler. », « Il l'observa pendant une minute entière » ou encore « Où as-tu trouvé ce livre ? » Notons que, comme la première partie du CoVEP, ces phrases sont reprises du corpus utilisé par Pépiot et al. (2021). Elles ont donc été rédigées avec un niveau de langue approprié pour des adultes, ce qui a pu déstabiliser autant les enfants chargé·e·s de les lire, que les participant·e·s du test perceptifs puisque les mots ne sonnent pas toujours très « naturels » dans une bouche d'enfant. Il aurait sûrement été préférable d'adapter ces phrases avec un vocabulaire plus infantin. Les troisième et quatrième parties du CoVEP sont respectivement identiques aux deux premières mais il a été demandé aux enfants de les prononcer en chuchotant.

Enfin, la dernière partie du CoVEP est constitué de trentaine de secondes de discours semi-spontané des enfants, à qui nous avons demandé de répondre à la question suivante : « Est-ce que tu peux me raconter, avec le plus de détails possible, ce que tu as fait hier entre le moment où tu t'es réveillé·e et celui où tu t'es couché·e ? ». Cette partie du corpus est prononcée sans chuchoter, avec un mode de voix normal.

2.1.3 Matériel et procédure d'enregistrement

Les enregistrements ont été réalisés avec un microphone de mesure Behringer ECM8000, un enregistreur Sound Devices MixPre-6 paramétré à une fréquence d'échantillonnage de 192 kHz sur une quantification de 24 bits. Ils se sont déroulés dans une salle de classe aux dimensions standards, équipée de double vitrage et avec un bruit de fond inférieur à 40 dBA. Lors de l'enregistrement, les enfants se tenaient debout sur un tapis, à une distance fixe de 30 cm du micro et étaient entouré-e-s par des panneaux acoustiques Clearsonic S5-2D. Il leur était demandé de lire deux fois les phrases présentées sur un panneau en face d'elles-eux. En cas de difficulté de lecture, la phrase posant problème était lue par une voix féminine et il leur était demandé de la répéter. Chaque phrase était toujours prononcées deux fois par sécurité, afin de laisser la possibilité ultérieure de sélectionner l'occurrence la plus satisfaisante acoustiquement. Si la réponse à la question correspondant à la dernière partie du corpus était trop courte, l'enfant était relancé-e par une question lui demandant des précisions sur sa première réponse. Lors des passages de la voix normale à la voix chuchotée et inversement, la distance au microphone restait la même mais un réglage de gain était effectué.

2.1.4 Réduction du corpus pour les analyses acoustiques

Afin de réduire le nombre de données à traiter, nous avons fait le choix de limiter à 10 le nombre de phrases qui seront étudiées, par enfant, dans la suite de cette étude. Nous avons décidé de garder les deux phrases présentant une intonation interrogative ainsi que la phrase exclamative. Les phrases les plus courtes et qui avaient été les plus facilement lues sans erreur par les enfants ont été privilégiées. Voici donc les cinq phrases qui ont été retenues, chacune ayant été prononcée avec un mode de voix normal puis avec une voix chuchotée :

1. « Steven a vendu son vélo hier après-midi. »
2. « Ma sœur m'a dit qu'elle allait passer demain. »
3. « Si tu refais ça, j'appelle la police ! »
4. « Est-ce que tu veux aller au cinéma ce soir ? »
5. « Où as-tu trouvé ce livre ? »

2.1.5 Méthodologie des analyses acoustiques

Les sons ont été découpés en utilisant le logiciel *Pro Tools*. Les pseudo-mots dissyllabiques des parties 1 et 3 du CoVEP ont été extraits des phrases capsules. Pour chaque phrase des parties 2 et 4 du CoVEP, des deux occurrences enregistrées, nous avons choisi celle dont l'intonation était la plus proche de celle indiquée par la ponctuation et dont la prosodie était la plus fluide. En ce qui concerne le discours semi-spontané, nous avons systématiquement sélectionné les trente premières secondes de la réponse en coupant la voix de la personne faisant passer le test si une question supplémentaire avait été posée. Nous avons procédé à une égalisation RMS du niveau de tous les sons en imposant un niveau maximal de -3 dBFS grâce un script *Python*. Les F0 des stimuli des deux premières parties du corpus ont été obtenues grâce à la fonction *Get Pitch* de *Praat* qui donne une valeur moyenne de la F0 sur la durée sélectionnée (voir la figure 11¹). De même, les valeurs de la déviation standard de la F0 (STD)² ont été obtenues grâce à la fonction *Get standard deviation* des valeurs prises par le *Pitch*, elle a été calculée sur la durée totale des stimuli. La collecte de ces données a été automatisée par un script *Praat* et les valeurs collectées ont été vérifiées ultérieurement.

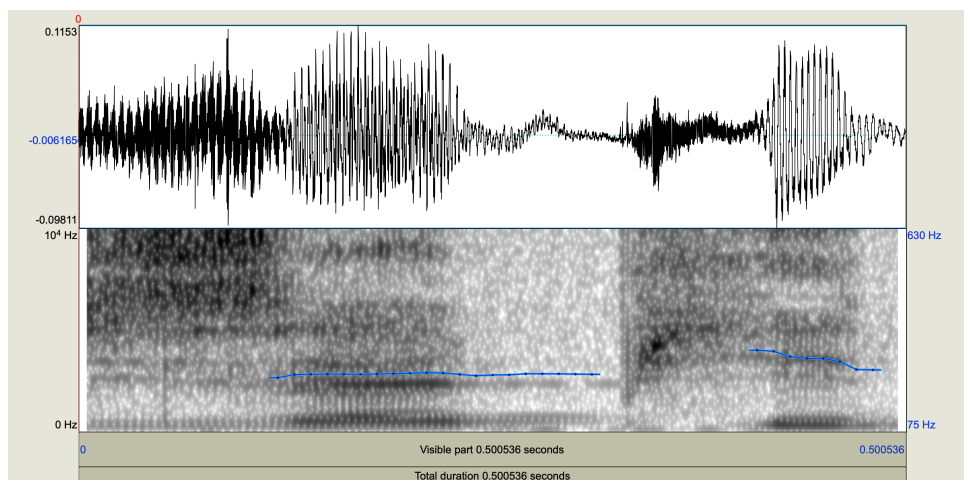


FIGURE 11 – Visualisation du pseudo-mot *sapi* sur le logiciel *Praat*. La partie supérieure représente l'onde sonore ; la partie inférieure, le spectrogramme et la courbe de F0 (en bleu) calculée par le logiciel.

Les calculs des moyennes des différentes données ont été effectués sur *Google Sheet*.

1. Le fichier son correspondant est audible via ce [lien SoundCloud](#).
2. la STD est la mesure de la dispersion des différentes valeurs prises par F0 autour de la F0 moyenne du stimulus étudié

Nous avons réalisé, grâce à un script *Python* dédié, plusieurs analyses de variance à un facteur¹, calcul que nous désignerons par la suite par l'acronyme ANOVA.

1. L'analyse de variance à un facteur est un modèle statistique qui permet de déterminer si les moyennes de multiples groupes provenant d'une même population présentent une différence significative. Ce test s'applique lorsqu'on cherche à déterminer l'influence d'une variable explicative catégorielle, aussi appelée variable indépendante, sur la loi de distribution d'une variable continue, dite variable dépendante ; ici, la répartition de la F0 en fonction du genre par exemple.

2.2 Analyse des enregistrements

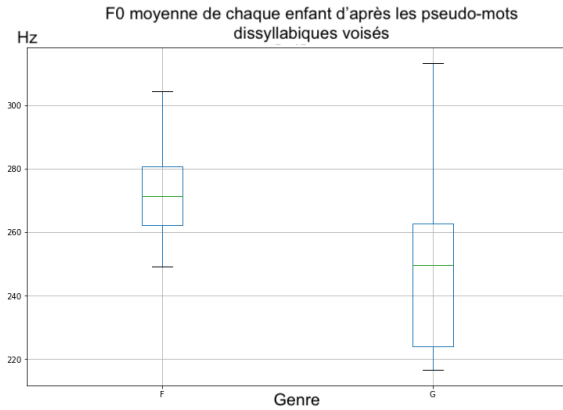
2.2.1 Présentation des résultats

Enfant	Pseudo-mots dissyllabiques	Phrases	Discours libre
F 10	262	258	257
F 11	279	281	264
F 12	263	249	236
F 14	282	278	271
F 17	273	265	250
F 18	249	247	234
F 20	281	268	242
F 23	254	254	251
F 26	270	257	240
F 28	304	302	293
G 06	313	274	265
G 07	261	281	236
G 08	217	215	218
G 15	220	217	213
G 16	240	259	249
G 19	226	222	198
G 22	268	224	201
G 24	259	260	254
Moyenne des filles	272	266	254
Moyenne des garçons	250	244	229

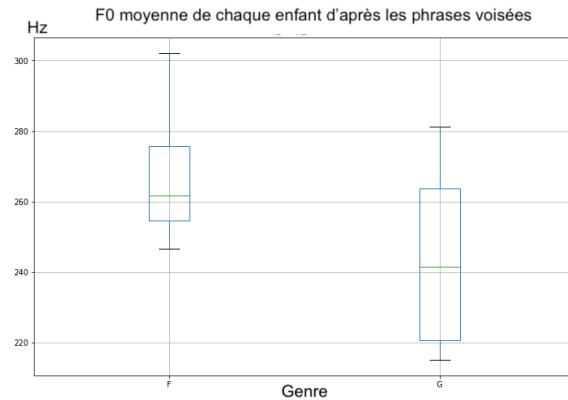
FIGURE 12 – Tableau présentant les F0 (en Hz) de chaque enfant ainsi que les F0 moyennes (en Hz) de chaque genre, calculées pour les trois types de stimuli voisés du corpus.

Les valeurs moyennes de la F0 de chaque enfant obtenue à partir des parties 1, 2 et 5 du corpus sont présentées dans la figure 12. On remarque que la F0 moyenne du groupe des filles est plus élevée que celle du groupe des garçons, quelle que soit la nature des stimuli étudiés et que l'écart augmente de 3 Hz quand on passe des stimuli lus au discours semi-spontané.

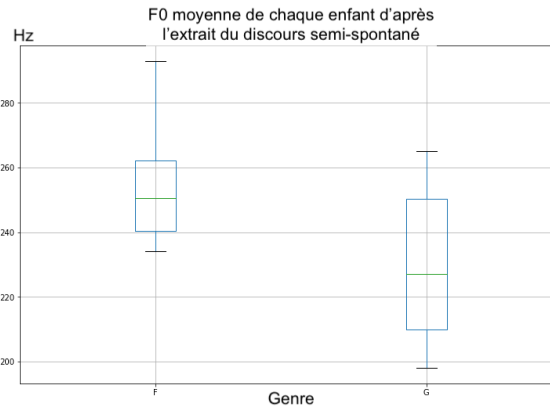
Les diagrammes en boîte des F0 moyennes des deux groupes *garçon* et *fille* sont présentés dans la figure 13 pour chacun des trois types de stimuli étudiés. On observe que les F0 des deux groupes se recouvrent largement pour les trois types de stimuli.



(a) Répartition des F0 moyennes de chaque enfant calculées à partir des pseudo-mots dissyllabiques, en fonction du genre



(b) Répartition des F0 moyennes de chaque enfant calculées à partir des phrases, en fonction du genre



(c) Répartition des F0 moyennes de chaque enfant calculées à partir du discours semi-spontané, en fonction du genre

FIGURE 13 – Répartition des F0 moyennes de chaque enfant en fonction du genre pour les trois parties voisées du corpus. La barre verte représente la médiane ; les côtés de la boîte, le 1^{er} et le 2^e quartile et les extrémités du segment indiquent les extrema des données.

Pour chacune des colonnes du tableau présenté en figure 12 nous avons réalisé une ANOVA, avec pour variable dépendante, la F0 et pour variable indépendante, le genre du-de la locuteur-ice.

La différence des F0 moyennes entre les groupes *garçon* et *fille* ne s'est pas avérée significative pour les deux premières catégories de stimuli ($F(1,16)=3,3$ avec $p=0,087$ pour les pseudo-mots dissyllabiques et $F(1,16)=4,3$ avec $p=0,054$ pour les phrases). En revanche, la différence de F0 moyennes entre les deux groupes s'est révélée significative pour le discours semi-spontané ($F(1,16)=5,7$ avec $p=0,029$).

La figure 14 illustre la répartition des F0 moyennes des discours semi-spontanés des 18 enfants étudiés.

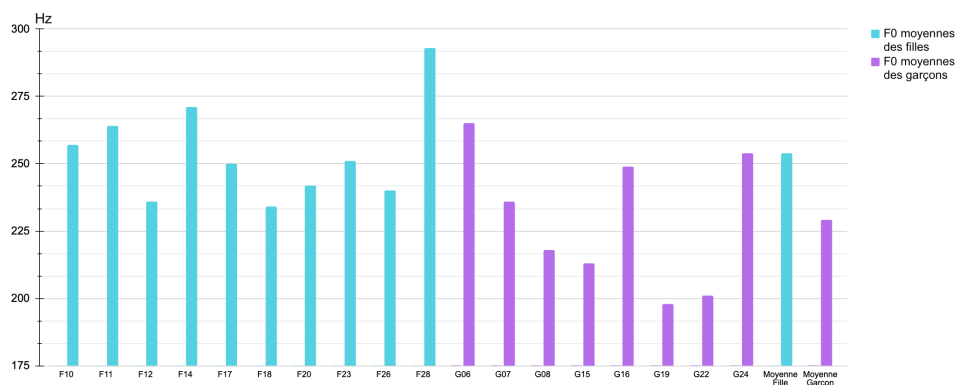


FIGURE 14 – Présentation des F0 (en Hz) de chaque enfant et des moyennes par genre, calculées à partir des extraits de discours semi-spontané (en bleu pour les filles et en violet pour les garçons)

Nous avons également mesuré la déviation standard de la F0 (STD) pour chaque stimulus ; c'est-à-dire la mesure de la dispersion des différentes valeurs prise par la F0 autour de la F0 moyenne du stimulus étudié. Les STD moyennes des garçons sont significativement supérieures à celle des filles pour les 3 types de stimuli considérés. La figure 15 présente la F0 moyenne de chaque enfant modulo sa STD moyenne pour les extraits de discours semi-spontané.

2.2.2 Discussion

À notre connaissance, une différence significative de la F0 en fonction du genre dans la voix parlée avant la puberté n'avait jamais été mise à jour en français. Nicollas et al. (2008) ont constaté une différence significative de la F0 entre un groupe de garçons et de

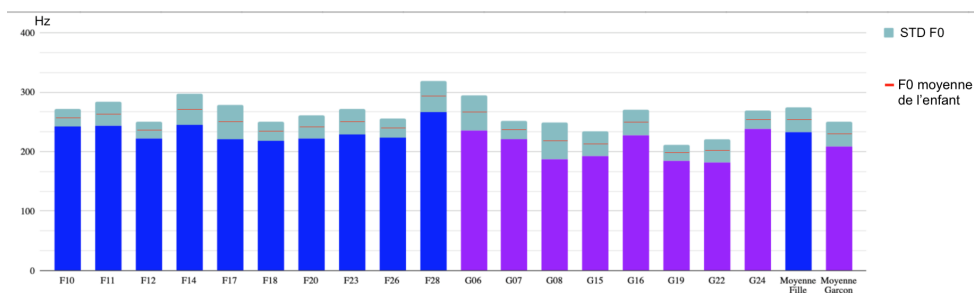


FIGURE 15 – Présentation des STD (en Hz) de chaque enfant et des moyennes par genre, calculées à partir des extraits de discours semi-spontané. La F0 moyenne est indiquée par la barre rouge et la zone grise illustre l'étendue de la STD autour de la valeur de la F0. Les profils de filles sont en bleu et ceux des garçons en violet

filles âgé-e-s de 6 à 12 ans lors de la production isolée du phonème [a] pendant plusieurs secondes. Cependant la tranche d'âge étudiée est susceptible de recouvrir le début de la puberté. De plus, comme nous l'avons vu dans la **I.4.1**, il est possible que les enfants placent leur voix différemment lors de la phonation soutenue d'une voyelle isolée et lors de l'exercice de la parole. Nos résultats sont en contradiction avec ceux obtenus par Courbis et al. (2014). En effet, les autrices n'avaient constaté aucune différence significative de la F0 moyenne en fonction du genre dans un corpus de paroles semi-spontanées prononcées par des enfant âgé-e-s de 8 à 10 ans. Nos résultats concernant le discours semi-spontané, s'ils sont reproduits, seraient donc particulièrement intéressants. Il serait souhaitable de reconduire une expérience semblable sur un échantillon plus important d'enfants du même âge afin de conforter ces résultats.

Sachant que la taille des plis vocaux n'est pas significativement différente en fonction du genre dans l'enfance (Rogers et al. 2014) et que les filles du groupe étudié sont plus grandes que les garçons en moyenne¹, il est peu probable que la différence de F0 moyenne mise en évidence soit liée à des disparités physiologiques entre les genres. Il est envisageable que la différence constatée soit liée à des pratiques articulatoires spécifiques à des normes de genre que les enfants assimilent progressivement au cours de leur développement. Si tel est le cas, les résultats présentés ici pourraient ne pas être en contradiction avec les études citées dans la partie **I.4.1** qui ne constatent pas de différence de la F0 moyenne de la parole chez les enfants anglophones avant la puberté (Bennett 1983, Busby et al. 1994, Sorenson 1989, S. Lee et al. 1999, Perry et al. 2001, Linders et al. 1995). La différence de F0 moyenne mise en évidence dans notre étude pourrait donc constituer un phénomène

1. La taille des plis vocaux est liée à la taille de l'individu. Si les filles de cette étude sont plus grandes, elles auront *a priori* des plis vocaux plus longs que les garçons.

culturel propre à langue française, qui ne serait pas à l'œuvre dans la langue anglaise.

Cartei et al. (2013) ont montré qu'en langue anglaise, les filles abaissent leur F0 pour imiter une voix de garçon et les garçons augmentent leur F0 pour imiter une voix de fille. Les enfants miment une différence entre les genres caractéristique des voix adultes alors même que l'existence de cette différence avant la puberté n'est pas établie en langue anglaise. Les enfants intègrent donc la différence de F0 moyenne entre les voix d'homme et de femme comme constitutive de l'expression de genre. Il est possible que les enfants observé-e-s dans cette étude aient appris à placer la F0 usuelle de leur voix parlée en partie en fonction de leur genre. Il faudrait toutefois tester nos observations sur un groupe plus large d'enfants, idéalement en obtenant l'information précise de la taille du conduit vocal de chaque enfant, ou en confrontant plusieurs groupes d'enfants de même âge et de tailles identiques pour pouvoir valider une telle hypothèse.

Notre hypothèse n°1 n'a pu être validée que pour la partie constituée de parole libre du CoVEP ; elle n'est pas validée pour les extraits de voix lue étudiés. Le fait que la différence de F0 moyenne entre les garçons et les filles ne soit pas significative pour les parties lues du corpus pourrait indiquer que les enfants ne placent pas leur voix de la même manière lors de l'exercice de la lecture que lors de la parole libre. Pour tou-te-s les enfants étudié-e-s, la F0 moyenne est plus basse lors de la parole libre que lors de la lecture, à l'exception de deux garçons (G08 et G16). En moyenne, les garçons ont davantage abaissé leur F0 en passant de la lecture à la parole que les filles, cela est peut-être lié au stress généré par l'exercice de la lecture à voix haute.

Conclusion partielle

Les résultats obtenus montrent que, bien que les F0 moyennes se recouvrent largement entre les genres entre 8 et 10 ans, il existe déjà une tendance à ce que les garçons adoptent une voix plus grave que les filles. Il est probable que cette différence soit liée au développement de stratégies articulatoires spécifiques au genre plutôt qu'à des disparités physiologiques. Il est impératif que ces observations soient reproduites au travers d'études ultérieures sur un échantillon plus important d'enfants francophones pour valider les hypothèses et résultats avancés dans ce travail.

Test perceptif autour d'une partie du corpus enregistré

3.1 Procédure expérimentale

3.1.1 Participant·e·s

Un groupe de 44 personnes a participé au test perceptif, il était constitué de 15 femmes, 27 hommes et deux personnes non-binaires, âgé·e·s de 20 à 59 ans. Une très large majorité d'entre elles·eux sont des étudiant·e·s de l'ENS Louis-Lumière et de l'Université Paris VIII. Tou·te·s les participant·e·s résident en Île-de-France et comptent le français parmi leurs langues maternelles ; aucun·e·s d'entre elles·eux n'a déclaré avoir de problèmes d'audition.

3.1.2 Matériau linguistique sélectionné

Afin de réduire la durée du test à un laps de temps raisonnable, nous avons sélectionné une fraction du CoVEP afin de constituer notre test perceptif. Nous avons d'abord choisi d'écartier arbitrairement les profils de deux petites filles afin de présenter un corpus paritaire aux participant·e·s. Les profils des filles n°10 et n°11 ont donc été écartés au hasard. Voici les principales caractéristiques des enfants sélectionné·e·s :

Test perceptif autour d'une partie du corpus enregistré

1. Garçon n°6 (G06) : 8 ans ; langue(s) maternelle(s) : français, haïtien ; F0 moyen : 304 Hz ; taille : 1,35 m
2. Garçon n°7 (G07) : 8 ans ; langue(s) maternelle(s) : français ; F0 moyen : 242 Hz ; taille : 1,44 m
3. Garçon n°8 (G08) : 8 ans ; langue(s) maternelle(s) : français ; F0 moyen : 221 Hz ; taille : 1,33 m
4. Fille n°12 (F12) : 8 ans ; langue(s) maternelle(s) : français ; F0 moyen : 263 Hz ; taille : 1,38 m
5. Fille n°14 (F14) : 8 ans ; langue(s) maternelle(s) : français, haïtien ; F0 moyen : 287 Hz ; taille : 1,26 m
6. Garçon n°15 (G15) : 8 ans ; langue(s) maternelle(s) : français ; F0 moyen : 219 Hz ; taille : 1,31 m
7. Garçon n°16 (G16) : 9 ans ; langue(s) maternelle(s) : français ; F0 moyen : 247 Hz ; taille : 1,29 m
8. Fille n°17 (F17) : 10 ans ; langue(s) maternelle(s) : français ; F0 moyen : 269 Hz ; taille : 1,40 m
9. Fille n°18 (F18) : 9 ans ; langue(s) maternelle(s) : français ; F0 moyen : 252 Hz ; taille : non mesurée
10. Garçon n°19 (G19) : 10 ans ; langue(s) maternelle(s) : français ; F0 moyen : 226 Hz ; taille : 1,53 m
11. Fille n°20 (F20) : 10 ans ; langue(s) maternelle(s) : français, espagnol ; F0 moyen : 283 Hz ; taille : 1,44 m
12. Garçon n°22 (G22) : 9 ans ; langue(s) maternelle(s) : français, cap verdien ; F0 moyen : 270 Hz ; taille : 1,35 m
13. Fille n°23 (F23) : 9 ans ; langue(s) maternelle(s) : français ; F0 moyen : 253 Hz ; taille : 1,48 m
14. Garçon n°24 (G24) : 8 ans ; langue(s) maternelle(s) : français ; F0 moyen : 269 Hz ; taille : 1,32 m
15. Fille n°26 (F26) : 10 ans ; langue(s) maternelle(s) : français, kabyle ; F0 moyen : 262 Hz ; taille : 1,43 m
16. Fille n°28 (F28) : 10 ans ; langue(s) maternelle(s) : français, ukrainien ; F0 moyen : 306 Hz ; taille : 1,49 m

Nous avons ensuite choisi de ne pas faire écouter les discours semi-spontanés des enfants qui représentaient à eux seuls presque dix minutes d'écoute sur la durée totale du test. Nous avons conservé les 10 phrases par enfant sélectionnées à partir des analyses acoustiques.

Enfin, des parties 1 et 3 de l'expérience, nous avons choisi d'écarter les pseudo-mots dissyllabiques commençant par des consonnes occlusives afin de resserrer notre attention sur les paramètres fréquentiels. Il serait intéressant de mener un test perceptif semblable, incluant les extraits écartés dans une étude ultérieure. Le test ainsi obtenu a une durée moyenne de 45 minutes et jusqu'à 1 heure pour les participant·e·s les plus lent·e·s.

3.1.3 Déroulement du test

Le test perceptif est réalisé sous le logiciel *Praat*. La collecte de données obtenues en réponse au test est également automatisée grâce à un script *Praat*. Dans un premier temps, chaque participant·e·s renseigne ses informations personnelles comme son genre, ses langues maternelles et autres langues parlées couramment, le nombre d'années vécues en Île-de-France, d'éventuels problèmes d'audition ou encore s'ils sont en contact régulier avec des enfants de moins de 12 ans. Avant le début du test, les participant·e·s sont informé·e·s que toutes les voix du test appartiennent à des enfants âgé·e·s de 8 à 10 ans et s'étant spontanément défini·e·s de manière binaire comme un «garçon» ou une « fille».

L'écoute des voix se fait au casque ouvert dans une salle calme. Le test à proprement parler consiste en un choix forcé pour chaque stimulus entre les catégories *garçon* et *fille* et à l'évaluation de la certitude sur une échelle discrète allant de 0 à 5 (voir la figure 16). L'expérience commence par trois stimuli tests qui permettent de s'habituer au fonctionnement du test et de régler le niveau d'écoute. Les échantillons test étaient les voix de deux garçons et une fille n'ayant pas été retenu·e·s pour l'expérience, car ne parlant pas le français dans leur cadre familial. Le test est ensuite subdivisé en quatre expériences numérotées de 1 à 4 qui désignent respectivement : l'écoute des pseudo-mots dissyllabiques voisés ; l'écoute des phrases voisées ; l'écoute des pseudo-mots dissyllabiques chuchotés et l'écoute des phrases chuchotées. Les quatre expériences se succèdent toujours dans le même ordre, mais l'écoute des stimuli au sein d'une expérience est randomisée. Les participant·e·s n'ont pas la possibilité de réécouter un son ou de revenir sur leur choix et leur temps de réponse est également collecté par le script.

Nous avons choisi de faire écouter les pseudo-mots dissyllabiques systématiquement avant les phrases en nous inspirant du paradigme du *gating* qui consiste à dévoiler progres-

1/3

Identifiez-vous une voix de garçon ou de fille ?

Fille

Garçon

0 (au hasard) 1 2 3 4 5 (certain·e)

FIGURE 16 – Fenêtre de réponse extraite du test perceptif. Elle présente le choix forcé entre les catégories *garçon* et *fille* et l'échelle permettant d'indiquer le degré de certitude entre 0 et 5

sivement des indices acoustiques supplémentaires d'une même voix afin de déterminer un seuil à partir duquel les indices disponibles permettent d'identifier le genre avec fiabilité et avec un degré de certitude élevé (Pépiot 2013). Selon cette méthode, il aurait été pertinent de commencer par les expériences 3 et 4 voire d'organiser le test selon l'ordre suivant : 3, 1, 4, 2. Nous n'avons pas retenu ces options, car nous avons estimé que l'écoute des pseudo-mots dissyllabiques chuchotés en début d'expérience aurait pu déstabiliser voire décourager les participant·e-s du test. L'écoute des pseudo-mots dissyllabiques voisés nous a paru une bonne manière d'introduire les participant·e-s à l'exercice qui leur était demandé.

3.1.4 Méthodologie de l'analyse des résultats

Les pourcentages de bonnes réponses ainsi que les temps de réponse moyens ont été calculés sur Google Sheet. Les temps de réponse supérieurs à 20 secondes n'ont pas été pris en compte afin de ne pas fausser les moyennes. En effet, la conception du test ne permettant pas de faire une pause au sein d'une même expérience, la plupart des participant·e-s ont fait des pauses d'elles-eux-même, allongeant ainsi exagérément le temps de réponse pour certains stimuli.

3.2 Analyses des résultats du test perceptif

3.2.1 Présentation des résultats

Scores d'identification

Le pourcentage d'identification correcte du genre de chaque enfant en fonction du type de stimulus écouté est présenté dans les figures 17 et 18. À l'écoute des pseudo-mots voisés, les filles ont été correctement identifiées à hauteur de 53% contre 65% pour les garçons ; à l'écoute des phrases voisées, les filles ont été identifiées à 65% et les garçons à 70% ; à l'écoute des pseudo-mots chuchotés, les filles ont été reconnues à 55% et les garçons à 57% ; enfin, à l'écoute des phrases chuchotées, les filles ont été reconnues à 64% et les garçons à 66%.

Enfant	Pourcentage d'identification correcte à l'expérience n°1	Pourcentage d'identification correcte à l'expérience n°2	Pourcentage d'identification correcte à l'expérience n°3	Pourcentage d'identification correcte à l'expérience n°4	Pourcentage d'identification correcte totale
G06	68,79%	67,73%	72,88%	66,36%	69,89%
G07	76,82%	89,09%	57,73%	82,73%	71,93%
G08	48,64%	56,82%	41,82%	52,73%	33,92%
G15	92,58%	91,82%	73,33%	92,27%	85,23%
G16	32,88%	34,09%	25,76%	37,73%	30,97%
G19	91,82%	81,82%	84,39%	65,91%	84,55%
G22	70,00%	76,36%	60,45%	76,82%	68,07%
G24	41,21%	63,64%	37,73%	55,45%	44,49%
F12	26,67%	45,91%	45,91%	31,82%	33,92%
F14	93,33%	85,00%	80,15%	82,27%	85,97%
F17	60,15%	77,27%	45,91%	50,45%	55,74%
F18	25,15%	48,64%	51,82%	85,91%	45,68%
F20	48,79%	60,91%	46,82%	66,36%	51,76%
F23	25,30%	34,55%	34,55%	31,82%	31,53%
F26	52,42%	69,55%	68,18%	84,09%	64,43%
F28	91,21%	96,36%	62,88%	76,70%	77,50%
Total Fille	52,88%	64,77%	54,53%	63,68%	55,82%
Total Garçon	65,34%	70,17%	56,76%	66,25%	61,13%
Total	59,11%	67,47%	55,64%	64,96%	58,47%

FIGURE 17 – Pourcentages de bonnes réponses obtenues sur les 44 participations, présentés pour chaque enfant en fonction du type de stimulus écouté.

Test perceptif autour d'une partie du corpus enregistré

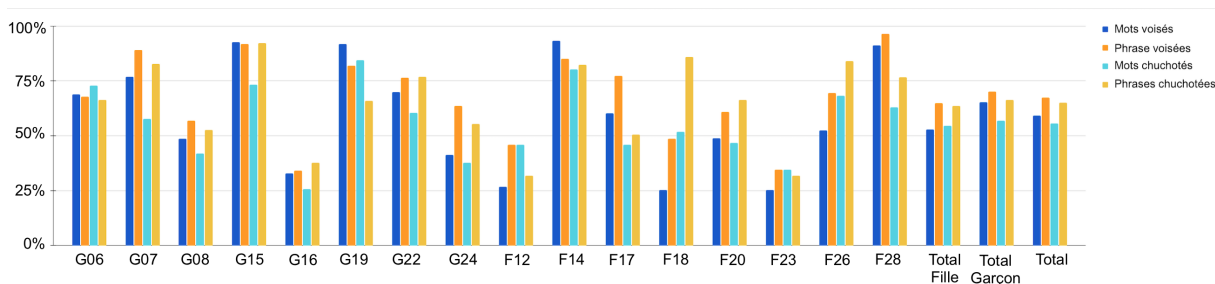


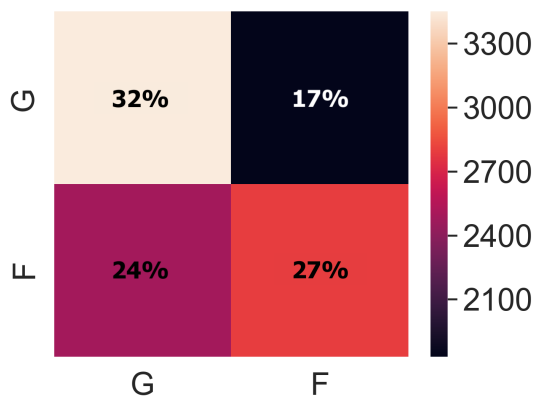
FIGURE 18 – Pourcentages de bonnes réponses obtenues sur les 44 participations, présentés pour chaque enfant en fonction du type de stimulus écouté. Le bleu foncé correspond à l'expérience n°1 (écoute des pseudo-mots dissyllabiques voisés) ; l'orange, à l'expérience n°2 (écoute des phrases) ; le bleu clair, à l'expérience n°3 (écoute pseudo-mots dissyllabiques chuchotés) et enfin le jaune, à l'expérience n°4 (écoute des phrases voisées)

Biais de réponse

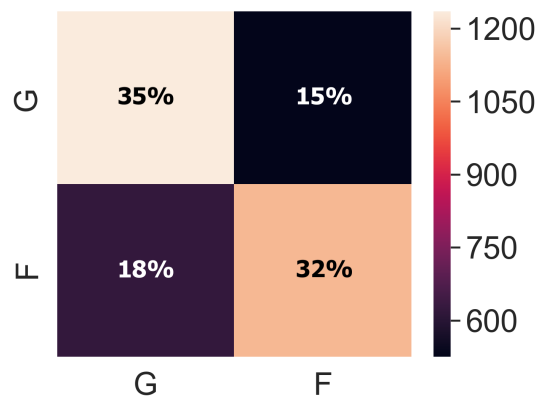
Nous avons établi une matrice de confusion pour chacune des quatre expériences constituant le test perceptif. Elles présentent la répartition des garçons identifiés comme des garçons (case supérieure gauche), des garçons identifiés comme des filles (case supérieure droite), des filles identifiées comme des filles (case inférieure droite) et des filles identifiées comme des garçons (case inférieure gauche) sur la totalité des réponses comptabilisées pour chaque expérience. Le code couleur indique le nombre absolu de réponses données pour chacune des quatre catégories sus-citées, et le pourcentage de réponses données que représente chaque catégorie est également indiqué.

Ces matrices permettent de visualiser à la fois la performance des participant·e·s au test et d'éventuels biais de réponse. En effet, si les participant·e·s sont performants, la diagonale de la matrice sera nettement plus claire car les filles sont majoritairement perçues comme des filles et les garçons, comme des garçons. Si, en revanche, les participant·e·s ont tendance à systématiquement privilégier une des deux réponses, c'est une des deux colonnes qui sera nettement plus claire, les enfants auront été identifié·e·s systématiquement selon le biais, indépendamment de leur genre véritable.

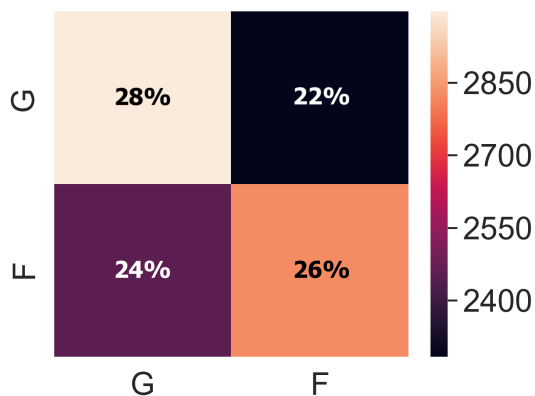
On observe que les auditeur·ice·s sont davantage performant·e·s dans la discrimination du genre à l'écoute des phrases qu'à l'écoute des pseudo-mots dissyllabiques, quel que soit le mode de voix utilisé. On constate également que dans les 4 expériences menées, les participant·e·s ont présenté un biais favorisant la réponse « garçon », en particulier à l'écoute des stimuli voisés.



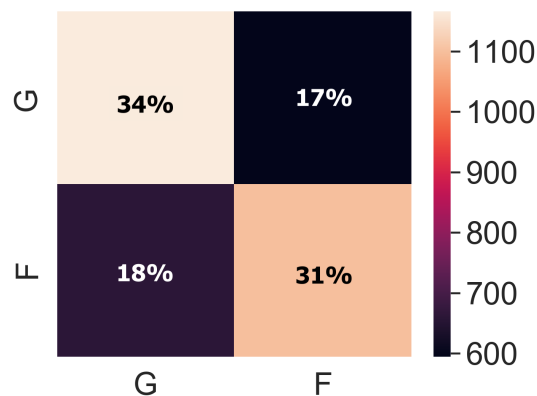
(a) Matrice de confusion de l'expérience 1



(b) Matrice de confusion de l'expérience 2



(c) Matrice de confusion de l'expérience 3



(d) Matrice de confusion de l'expérience 4

FIGURE 19 – Matrices de confusion des 4 expériences composant le test perceptif. Les colonnes indiquent le genre identifié et les lignes, le genre véritable. L'échelle de couleur indique le nombre d'occurrences de l'association de la réponse *garçon* à un stimulus ayant été produit par un garçon, pour la case supérieure gauche et ainsi de suite pour les autres cases de la matrice. Il est également indiqué dans les cases, le pourcentage que représente ce nombre sur le total des réponses données par les 44 participant·e·s pour chaque expérience.

Degré de certitude

Sur l'échelle de 0 à 5 proposée, le degré de certitude moyen de tou·te·s les participant·e·s au test est de 3,35 à l'écoute des pseudo-mots dissyllabiques voisés ; de 3,47 à l'écoute des phrases voisées ; de 2,47 à l'écoute des pseudo-mots dissyllabiques chuchotés et de 2,98 à l'écoute des phrases chuchotées.

Temps de réponse

Enfin, les temps de réponse moyens de toutes les participant·e·s au test sont de 2,6 s pour l'expérience n°1; 4,2 s pour l'expérience n°2; 2 s pour l'expérience n°3 et 3,8 s pour l'expérience n°4. Lors des deux premières expériences, le temps de réponse est très significativement plus long lorsque la réponse *garçon* a été donnée ($F(1,10519)=15,1$ avec $p<0,001$ pour l'expérience 1 et $F(1,3505)=27,0$ avec $p<0,001$ pour l'expérience 2); pour les expériences 3 et 4, il est significativement plus long lorsque la réponse *garçon* a été donnée.

3.2.2 Discussion

Capacité à identifier le genre

Les scores d'identification du genre à l'écoute des pseudo-mots dissyllabiques voisés et chuchotés (respectivement 59% et 56%) sont significativement supérieurs au score obtenu si l'on répond au hasard (de l'ordre de 50%). En effet, observons le nombre de réponses correctes obtenues pour les voix de filles dans l'expérience n°1, qui est le plus bas score obtenu sur l'ensemble du test perceptif. La probabilité s'obtenir un tel score (2 792 bonnes réponses sur 5 280 essais effectués avec des voix de filles), si les participant·e·s avaient répondu au hasard, est infime ($P(2\ 792)=0,000\ 002$ d'après la loi binomiale¹).

Nous pouvons en conclure que les participant·e·s sont capables de discriminer le genre dans des proportions significativement supérieures à la chance. Toutefois ces scores sont considérablement plus bas que ceux obtenus à l'écoute de voix d'adultes. En ce qui concerne l'expérience n°1, bien que le score de réponses correctes avoisine 60% pour les voix de garçons, les participant·e·s ont eu largement tendance à répondre davantage *garçon*. Si bien que les garçons ont été correctement classifiés dans une proportion relativement large alors que les filles, elles, n'ont été reconnues qu'à hauteur de 53%. Il semble donc que le score obtenu pour l'expérience ne corresponde pas à une réelle aptitude de discrimination du genre à l'écoute de pseudo-mots dissyllabiques et trahisse plutôt un biais favorisant la réponse « garçon ». Ceci nous amène à conclure que l'écoute de pseudo-mots dissyllabiques voisés suffit à reconnaître le genre dans des proportions significatives.

Ces résultats correspondent à ce que notre étude bibliographique laissait supposer. En

1. Selon la loi binomiale, on a : $P(x) = \frac{N!}{x!(N-x)!} \pi^x (1-\pi)^{N-x}$

avec $x = 2\ 792$, le nombre de bonnes réponses obtenues,

$N = 5\ 280$, le nombre d'essais effectués,

$\pi = 0,5$, la probabilité d'obtenir une bonne réponse si les participant·e·s répondent au hasard.

effet, Perry et al. (2001) constatent que les participant·e·s de leur étude sont capables d'identifier le genre de voix pré-pubères au-delà de la chance à l'écoute d'une syllabe de structure CVC¹. Sachs et al. (1973) et Weinberg et al. (1979b) font une observation similaire concernant l'écoute de voyelles isolées. Toutefois c'est la première fois, à notre connaissance que de tels résultats sont produits pour la langue française. Nos résultats pourraient indiquer que les paramètres spectraux présent dans les pseudo-mots dissyllabiques, comme la F0 ou la position des formant, constituent des indices pour l'identification le genre à l'écoute de voix d'enfants. Les francophones, comme les anglophones, s'appuieraient en partie sur la F0 ou la position des formants pour discriminer le genre des enfants, à l'écoute de leur voix.

Les scores obtenus à l'écoute des phrases voisées et chuchotées (respectivement 67% et 65%), se détachent également nettement de la chance. Weinberg et al. (ibid.) et Marshal (1972²) obtiennent des scores similaires en faisant écouter des phrases voisées prononcées par des enfants âgé·e·s respectivement de 6 et 7 ans et de 8 et 9 ans. Les scores obtenus pour les expériences 2 et 4 sont largement supérieurs que ceux obtenus aux expériences 1 et 3. Il est possible que, pour l'identification du genre à l'écoute des phrases, les participant·e·s se soient appuyé·e·s sur des indices qui n'étaient pas présents dans les pseudo-mots dissyllabiques, et qui leur ont permis d'obtenir de meilleurs scores. Amir et al. (2012) et Guzman et al. (2014) ont elles-eux aussi constaté des scores nettement plus élevés à l'écoute de phrases qu'à l'écoute de voyelles isolées. On peut émettre l'hypothèse que l'intonation a joué un rôle important dans la perception que les participant·e·s ont eu du genre des enfants. Il est également possible qu'à l'écoute des phrases, le cerveau analyse les valeurs relatives de certains paramètres spectraux entre plusieurs phonèmes comme le \bar{G}_{talker} proposé par Assmann et al. (2008), ce qui permettrait de discriminer le genre plus efficacement qu'à l'écoute d'un seul pseudo-mot dissyllabique.

Si l'on compare deux à deux les scores recueillis à l'expérience n°1 par rapport à la n°3 et à l'expérience n°2 par rapport à la n°4, on constate que les stimuli voisés obtiennent un meilleur taux d'identification que les stimuli chuchotés mais que les différences de scores de sont pas très importantes. Il semble donc que la F0 joue un rôle dans la perception du genre. Toutefois on peut affirmer qu'un certain nombre d'autres paramètres acoustiques permettent aux auditeur·ice·s de discriminer les voix de garçons et de filles, ce qui est cohérent avec le large taux de recouvrement entre les F0 moyennes des filles et des garçons examiné·e·s dans cette étude.

1. le sigle CVC indique un enchaînement de type Consonne – Voyelle – Consonne

2. Dans une étude non publiée citée par Weinberg et al. (1979b)

La réponse *Garçon* par défaut

Perry et al. (2001), eux aussi, ont rapporté une forte disproportion entre les réponses *garçon* et *fille*. Dans leur étude de 2001, 74% des garçons ont été correctement identifiés contre seulement 56% des filles. De même, Meditch (1975), indique que les auditeur·ice·s sont significativement plus performant·e·s pour identifier les garçons (85% d'identification correcte pour les garçons contre 74% pour les filles). Il est possible que, face à un stimulus ambigu, les participant·e·s aient tendance à répondre *garçon* par défaut. Dans les résultats de notre étude, le phénomène est beaucoup moins marqué pour les stimuli chuchotés. On peut supposer que la propension à répondre davantage *garçon* soit compensée par une tendance à associer les voix chuchotées à des voix féminines. En effet, en voix chuchotée, les formants se décalent vers le haut du spectre par rapport au mode voix normal (Fux 2012), ce qui pourrait influencer la perception du genre. De plus, le souffle, caractéristique du chuchotement, pourrait être associé au mode de voix « *breathy* », plus souvent adopté par les femmes en langue française (Pépiot 2013). Toutefois, une étude plus poussée sur la perception du genre en voix chuchotée serait nécessaire pour valider cette hypothèse.

Quant à l'éventuel biais favorisant la réponse *garçon* en voix modale, il trouve peut-être son origine dans le phénomène du « masculin comme norme » (Arnold 2015, p. 225) abordé par Arnold dans sa thèse. Il y décrit un paradigme culturel qui façonne un masculin « dégenré et universalisé » (ibid., p. 231), contribuant ainsi à ériger le masculin comme un neutre par défaut dans l'imaginaire collectif. Dès lors, un profil anonyme, sans caractéristiques particulières, évoque davantage le masculin, puisque le féminin constitue, par essence, une figure d'altérité. Selon cette approche, en l'absence d'indices acoustiques concluants sur le genre, les auditeur·ice·s pourraient être amené·e·s à choisir la réponse *garçon* par défaut. Cette hypothèse concorde avec la mesure des temps de réponse. En effet, le temps de réponse est significativement plus long lorsque la réponse *garçon* est donnée, en particulier pour les stimuli voisés. Les voix perçues comme ambiguës en matière de genre, pour lesquelles les auditeur·ice·s hésitent avant de choisir une réponse, seraient plus souvent catégorisées comme des voix de garçon, par défaut, ce qui expliquerait la différence de temps moyens observés en fonction de la réponse donnée.

Il serait intéressant de tester cette hypothèse dans une étude ultérieure. On peut imaginer, par exemple, un dispositif expérimental dans lequel on demande aux auditeur·ice·s d'identifier le genre, d'évaluer le degré de masculinité/féminité, puis de donner une brève description de l'enfant qu'ils imaginent pour chaque stimulus. L'étude des réponses données permettrait de vérifier si les profils « neutres » sont davantage classés comme des

garçons par les participant·e·s.

Un autre élément de réponse possible à la prépondérance de la réponse *garçon* est la classe sociale des enfants enregistrés. Edwards (1979) a montré que les auditeur·ice·s issu·e·s d'une classe sociale moyenne ont tendance à percevoir les enfants issus de la classe ouvrière davantage comme des garçons et les enfants de classe sociale moyenne, comme des filles. Dans la présente étude, les participant·e·s du test perceptif sont issu·e·s d'une classe sociale aisée et les enfants enregistrés, d'une classe sociale populaire, il est donc possible que le biais mis en évidence par Edwards se soit appliqué. Une étude comparant la perception du genre par la voix entre un groupe d'enfants francophones issu·e·s de la classe populaire et un groupe d'enfants issu·e·s d'une classe aisée permettrait d'évaluer l'influence de la classe sociale dans la perception du genre en langue française.

Notons qu'il est probable que le biais observé résulte d'une combinaison des deux hypothèses proposées ici et éventuellement d'autres phénomènes que nous ne sommes pas parvenue à identifier.

Rôle de la F0

Lorsque l'on compare deux à deux les degrés de certitude moyens obtenus aux expériences 1 et 3 et 2 et 4, on constate que le degré de certitude moyen diminue d'environ 1 point lorsqu'on passe de l'écoute des stimuli voisés à l'écoute des stimuli chuchotés. Il semble donc qu'en retirant la F0, on prive les auditeur·ice·s d'un indice important. La figure 20 présente la répartition des F0 moyennes des stimuli de l'expérience n°1 classés comme *garçon* par rapport à celles des stimuli classés comme *fille*. Bien que la F0 soit significativement supérieure en moyenne pour les stimuli pour lesquels la réponse *fille* a été donnée ($F(1,10558)=499,7$ avec $p<0,001$), on observe que les F0 des deux groupes se recouvrent largement. Il est donc peu probable que la F0 constitue l'indice principal sur lequel les auditeur·ice·s se reposent pour identifier le genre. On peut supposer que les auditeur·ice·s utilisent une combinaison de la F0 moyenne et d'autres paramètres spectraux comme la position des formants et l'intonation pour estimer le genre des enfants.

Étude de profils particuliers

Dans la figure 18, certains profils émergent. Nous avons choisi d'étudier les cas de 4 profils quasi-systématiquement bien identifié·e·s (G07, G15, G19, F23 et F28) et 3 profils quasi-systématiquement mal identifié·e·s (G16, F18 et F23). La répartition des valeurs moyennes des F0, F1 et F2 de chaque stimuli de l'expérience n°1 sont exposées dans

Test perceptif autour d'une partie du corpus enregistré

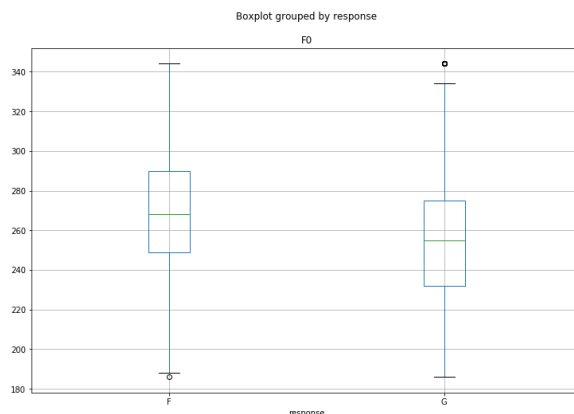
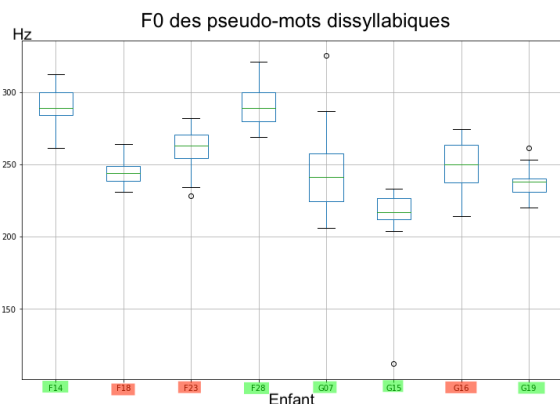


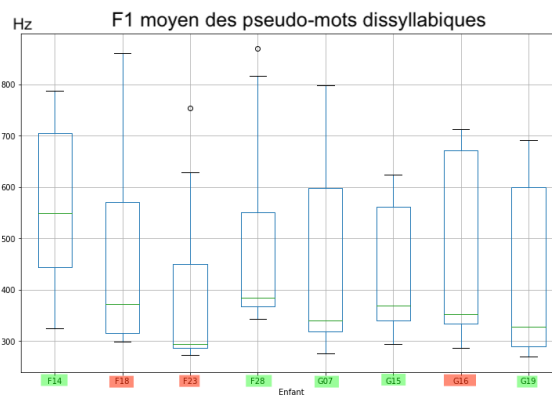
FIGURE 20 – Répartition des F0 moyennes des stimulus de l'expérience n°1 en fonction de la réponse *Garçon* ou *Fille* qui a été donnée. La barre verte représente la médiane ; les côtés de la boîte, le 1^{er} et le 2^e quartile et les extrémités du segment indiquent les extrema des données.

la figure 21 pour les 7 enfants sélectionné-e-s. On constate effectivement que les filles bien identifiées ont une F0 nettement supérieure aux garçons bien identifiés. Toutefois les enfants mal identifié-e-s ne présentent pas des F0 très différentes des leurs camarades bien identifié-e-s. Par exemple, G07 et G16 ont des plages de F0 moyennes assez semblables, pourtant G07 est largement correctement identifié et G15 majoritairement mal identifié. En ce qui concerne la position de F1 et F2, aucune différence claire ne se dégage de l'étude de ces profils particuliers.

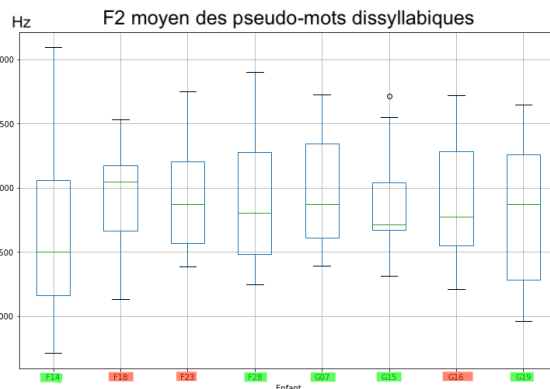
L'étude de cas des enfants les mieux et moins bien identifié-e-s ne nous permet de tirer aucune conclusion définitive. Pour obtenir des résultats plus concluants, il serait intéressant de mener une étude acoustique de chaque stimulus individuellement, en fonction du score l'identification correcte obtenue. La figure 22 présente le nombre d'identifications correctes par enfant pour les 44 participant-e-s au test perceptif, pour l'exemple du stimulus voisé « sapi ». Les figures correspondant aux 39 autres stimuli constituant le test perceptif sont présentées en annexe de ce document (voir figures 24 et 25). L'analyse de certains paramètres acoustiques en fonction de ces figures donnerait une compréhension plus complète du processus d'identification du genre par la voix des enfants.



(a) Répartition de la F0 des pseudo-mots dissyllabiques pour chaque profil particulier



(b) Répartition des valeurs de F1 des pseudo-mots dissyllabiques pour chaque profil particulier



(c) Répartition des valeurs de F2 des pseudo-mots dissyllabiques pour chaque profil particulier

FIGURE 21 – Répartition des valeurs de F0, F1 et F2 des enfants F14, F18, F23, F28, G07, G15, G16 et G19, calculées à partir des pseudo-mots dissyllabiques voisés. Le vert indique les enfants quasi-systématiquement correctement bien identifié-e-s; le rouge, les enfants quasi-systématiquement correctement mal identifié-e-s. La barre verte représente la médiane; les côtés de la boîte, le 1^{er} et le 2^e quartile et les extrémités du segment indiquent les extrema des données.

Test perceptif autour d'une partie du corpus enregistré

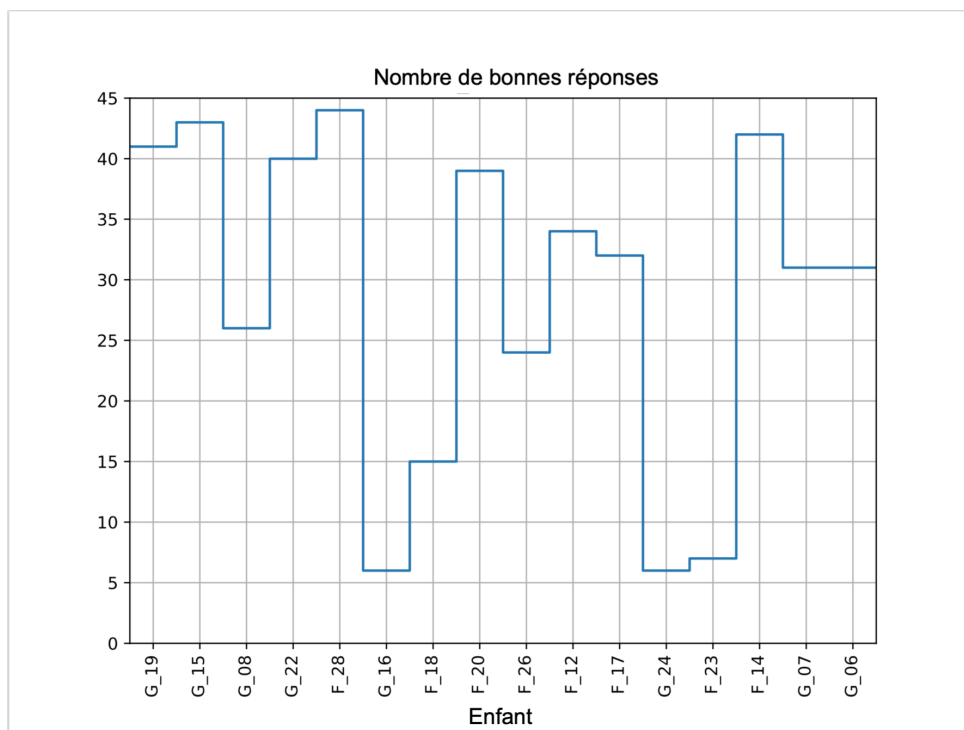


FIGURE 22 – Nombre d'identification correcte de chaque enfant, sur les 44 participations au test, à l'écoute du stimulus voisé « sapi »

Conclusion partielle

Notre test perceptif a permis de valider notre hypothèse n°2, à savoir que les auditeur·ice·s francophone sont capables de reconnaître le genre d'enfants âgé.e.s de 8 à 10 ans à l'écoute de leurs voix au-delà de la chance). Les participant·e·s ont été plus performant·e·s à l'écoute de phrases que des stimuli plus courts.

L'analyse de nos résultats ne permet pas de trancher définitivement la question soulevée par notre hypothèse n°3, à savoir que la F0 joue un rôle dans l'identification du genre). Il semble que la F0 joue un rôle dans la perception qu'ont eu les participant·e·s du genre mais les résultats présentés ici ne permettent pas de démontrer le caractère décisif de la F0.

Une étude des autres paramètres susceptibles de jouer un rôle dans la perception du genre pourrait permettre d'accéder à une meilleure compréhension de l'articulation des différents critères sur lesquels s'appuient les auditeur·ice·s pour arrêter leur jugement sur le genre des voix d'enfants.

Conclusion

Notre étude bibliographique nous a permis de constater que la voix joue un rôle fondamental dans le *travail de genre*¹. La façon de parler représente non seulement un outil de l'expression de genre mais aussi un indice pour la perception du genre des autres. Des études dans de multiples langues ont montré que les voix d'hommes et femmes se distinguent sur de nombreux aspects acoustiques. Ce phénomène trouve ses origines à la fois dans des caractéristiques morphologiques propres à chaque sexe et dans des stratégies articulatoires qui relèvent de la pratique culturelle.

L'étude des voix d'enfants est un enjeu important dans la compréhension du développement de l'appareil vocal et de l'émergence de ces pratiques articulatoires. Bien que des modélisations de plus en plus précises du conduit vocal voient le jour, il est encore aujourd'hui largement considéré que l'appareil vocal ne se différencie en fonction du sexe qu'à partir de la puberté. En particulier, on ne constate aucune différence significative des dimensions des plis vocaux dans l'enfance, or la F0 est étroitement liée à ce paramètre. Dès lors, on pourrait supposer que les petits garçons n'ont pas des voix plus graves que les petites filles avant l'adolescence. La question d'une différence significative F0 entre les voix de garçons et de filles dans l'enfance a souvent été étudiée, essentiellement en anglais, mais ne fait pas consensus.

L'existence d'une telle différence n'a pas pu être démontrée avec des extraits de voix lue ou parlée. En revanche, il a été montré que d'autres paramètres acoustiques se différencient significativement en fonction du genre chez des enfants anglophones, notamment

1. Le terme réfère ici à la notion de *gender work* telle qu'elle est définie par Kessler et al. (1978, p. 136)

CONCLUSION

la position des deux premiers formants, qui constituent une partie importante du timbre de la voix ou encore le VOT, qui décrit un aspect temporel du voisement. De plus, il a été montré que la voix fait partie intégrante des représentations stéréotypiques du genre assimilées par les enfants au cours de leur développement, sur lesquelles ces dernier·e·s s'appuient lors de leurs interactions sociales. Dès l'âge de 6 ans, les enfants d'une étude anglophone à qui l'on a demandé d'imiter des « voix de fille » et des « voix de garçon », ont instinctivement contrefait des différences spectrales que l'on observe entre les voix des femmes et des hommes adultes. Si, au cours de leur développement, les enfants intègrent certaines caractéristiques des « voix de femme » et « voix d'homme » comme des attributs de la féminité et de la masculinité, iels pourraient être amené·e·s à imiter celles-ci aux fins de leur propre présentation de genre. Ce phénomène social pourrait expliquer une partie des différences acoustiques observées entre les voix de garçon et de filles.

Notre étude d'un corpus de voix de 10 filles et 8 garçons âgé·e·s de 8 à 10 ans, a montré que les garçons ont adopté une F0 significativement plus basse que les filles lors d'un exercice de parole semi-spontanée d'une trentaine de secondes. Toutefois l'observation d'une telle différence n'a pu être mise en évidence formellement pour les extraits de voix lues produites par les mêmes enfants. En outre, le groupe d'enfants étudié est trop restreint pour que nos conclusions puissent raisonnablement s'inférer à une population plus large. Les résultats préliminaires présentés dans notre étude demandent donc à être validés par des études ultérieures.

La réalisation d'un test perceptif d'identification du genre à partir d'une partie du CoVEP, a permis de montrer que les auditeur·ice·s sont susceptibles de reconnaître le genre d'un·e enfant à l'écoute de sa voix dans des proportions significativement supérieures au hasard. Toutefois cette capacité existe dans une proportion largement moindre que pour des voix d'adultes. Il semble que dans notre expérience, la F0 aie bel et bien joué un rôle dans la perception du genre des enfants mais il est peu probable qu'elle ait constitué l'indice principal sur lequel les auditeur·ice·s ont fondé le jugement.

Le corpus enregistré à l'occasion de ce travail n'a pas été exploité dans son intégralité. Une étude acoustique plus exhaustive de celui-ci pourrait donner accès à une compréhension plus complète du développement de l'expression du genre dans la voix au cours de l'enfance. L'étude de la position des formants, des zones de bruit des consonnes fricatives, de l'intonation, du VOT ou de la qualité de voix, pour ne citer que ces paramètres, permettrait de mieux comprendre l'interaction des différents paramètres acoustiques dans le processus d'identification du genre par le cerveau. Il serait également pertinent de ne pas se limiter à l'étude des moyennes et des variances de valeurs absolues des paramètres spec-

traux étudiés. Des analyses temporelles de ces mêmes paramètres seraient très éclairantes, d'autant plus que l'analyse de la voix par le cerveau semble être intégrative.

L'expression et la perception du genre dans la voix infantile ont été peu étudiées en langue française. Les résultats présentés offrent de premiers éléments de réponses à des questions qui n'avaient pas encore été explorées en français. Au-delà des analyses présentées, ce mémoire aura été l'occasion de créer un corpus d'enregistrement de voix d'enfants précieux pour d'éventuelles études ultérieures. L'étude des voix d'enfants est peu accessible en raison de la difficulté de constituer des bases de données exploitables. En français, l'enregistrement de voix d'enfants selon une méthodologie rigoureuse est souvent réservé aux travaux d'orthophonie ou à l'étude de pathologies liées à l'appareil vocal. La constitution d'un corpus exploitable de voix d'enfants, construit autour de la question du genre, est un apport important aux ressources destinées à l'étude socio-phonétique du français. La structure du CoVEP a été pensée pour donner la possibilité d'analyser plusieurs paramètres acoustiques supplémentaires à ceux abordés dans ce mémoire. L'étude de ces critères permettrait d'apporter de nombreux éléments de réponses aux questions restées en suspens au terme de cette étude préliminaire. Il serait également intéressant de mener d'autres expériences perceptives à partir du CoVEP, notamment à partir des extraits de discours semi-spontané. L'enregistrement d'un corpus complémentaire, constitué de voix d'enfants issu-e-s d'une classe sociale aisée, permettrait non seulement de conforter ou de nuancer nos résultats, mais aussi d'évaluer l'impact de la classe sociale des enfants sur la perception de leur genre en confrontant les deux corpus dans un même test perceptif.

Malheureusement la structure du CoVEP présente certaines limites. En premier lieu, le nombre trop restreint d'enfants enregistré-e-s mais aussi la question du niveau de langue des phrases proposées, la disparité des niveaux de lecture ou encore de l'impact de l'exercice de la lecture sur la façon de poser la voix. Il aurait été intéressant d'enregistrer une partie du corpus en demandant aux enfants de décrire des images standardisées pour les amener à prononcer certains mots *balises*, par exemple « tapis », « toupie » et « tipi ». Ces enregistrements auraient permis de s'affranchir des problèmes de lecture, ils auraient pu être confrontés à une seconde série d'enregistrements semblables à la première partie du CoVEP afin d'étudier l'impact de la lecture sur le spectre de la voix en fonction des enfants.

Le nombre trop restreint d'enfants enregistrés représente également un écueil pour le test perceptif que nous avons élaboré. Plusieurs participant-e-s aux tests perceptifs ont eu l'impression de reconnaître des voix d'enfants d'une expérience à l'autre, leurs réponses ont donc pu être influencé-e-s pendant le déroulement du test. Il aurait été pertinent

CONCLUSION

d'introduire un pourcentage de stimuli apparaissant à deux reprises au cours du test, pour évaluer le taux reproduction des réponses de chaque participant·e-s.

Enfin, le fait d'aborder la perception du genre de façon binaire présente des limites évidentes. Les travaux élaborés par Valentina Cartei et ses collaborateur·ice-s proposent des méthodes de travail autour du genre prometteuses. Il serait intéressant de prolonger les questions abordées dans ce travail en mettant en regard l'identification du genre des enfants avec la perception de la masculinité et de la féminité de la voix ou en examinant le rapport des enfants à l'expression de genre, aussi bien à travers l'étude de leurs pratiques de l'imitation, que de leurs représentations archétypales des « voix de garçon » et des « voix de filles ».

Annexe 1 : Schéma du système phonatoire supérieur

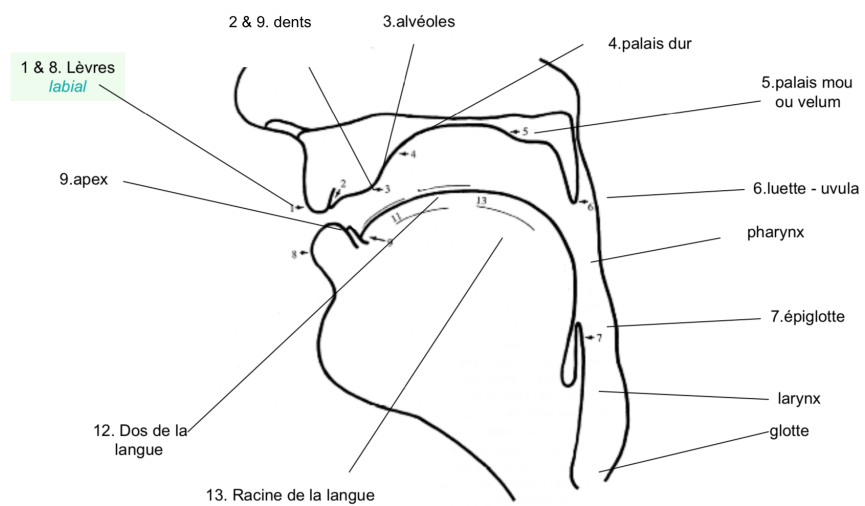


FIGURE 23 – Schéma du système phonatoire supérieur (Bogliotti 2010)

Annexe 2 : Corpus lu

1. Il a dit « sapi » deux fois.
2. Il a dit « soupi » deux fois.
3. Il a dit « sipi » deux fois.
4. Il a dit « chapi » deux fois.
5. Il a dit « choupi » deux fois.
6. Il a dit « chipi » deux fois.
7. Il a dit « zapi » deux fois.
8. Il a dit « zoupi » deux fois.
9. Il a dit « zipi » deux fois.
10. Il a dit « japi » deux fois.
11. Il a dit « joupi » deux fois.
12. Il a dit « jipi » deux fois.
13. Il a dit « papi » deux fois.
14. Il a dit « poupi » deux fois.
15. Il a dit « pipi » deux fois.
16. Il a dit « tapi » deux fois.
17. Il a dit « toupi » deux fois.
18. Il a dit « tipi » deux fois.
19. Il a dit « kapi » deux fois.
20. Il a dit « koupi » deux fois.
21. Il a dit « kipi » deux fois.
22. Il a dit « bapi » deux fois.
23. Il a dit « boupi » deux fois.
24. Il a dit « bipi » deux fois.
25. Il a dit « dapi » deux fois.
26. Il a dit « doupi » deux fois.
27. Il a dit « dipi » deux fois.
28. Il a dit « gapi » deux fois.
29. Il a dit « goupi » deux fois.
30. Il a dit « guipi » deux fois.
31. Il a dit « api » deux fois.
32. Il a dit « oupi » deux fois.
33. Il a dit « ipi » deux fois.

1. Steven a vendu son vélo hier après-midi.
2. Quand il fait froid et qu'il pleut, je préfère rester chez moi.
3. Les conservateurs, qui ont toujours soutenu cette mesure, sont satisfaits.
4. Ma sœur m'a dit qu'elle allait passer demain.
5. Si tu refais ça, j'appelle la police !
6. J'espère juste qu'un jour on pourra en parler.
7. Il alluma une cigarette et observa son visage.
8. Franny se réveilla en sursaut.
9. Il l'observa pendant une minute entière.
10. Il y avait un silence étrange à l'autre bout du fil.
11. Est-ce que tu veux aller au cinéma ce soir ?
12. Où as-tu trouvé ce livre ?

Annexe 3 : Template du questionnaire de profil pour les enfants

Questionnaire n°

Prénom :

Nom :

Date de naissance :

Âge :

Genre :

Langue(s) parlée(s) :

Langue(s) parlée(s) à la maison :

Langue(s) parlée(s) à l'école :

Profession du parent 1 :

Profession du parent 2 :

(préciser le secteur d'activité en cas de chômage ou de retraite)

Lieu d'habitation actuel :

Depuis combien de temps :

Lieux d'habitation précédents :

Taille :

École :

Classe :

Est-ce que tu lis beaucoup en français en dehors des cours ?

Est-ce que tu as déjà suivi :

Des cours de chant :

Si oui, combien de temps ? À quelle fréquence ?

Des cours de théâtre :

Si oui, combien de temps? À quelle fréquence?

Niveau de stress ressenti sur une échelle de 0 à 5

Est-ce que tu as des troubles d'élocution? (déjà vu un·e phoniatre ou un·e orthophoniste?)

Est-ce que tu fumes?

Retours libres sur l'expérience :

Commentaires de l'expérimentateur.ice :

Annexe 4 : Détails des scores obtenus par stimulus



FIGURE 24 – Nombre d'identification correcte de chaque enfant, sur les 44 participations au test, à l'écoute des stimuli voisés

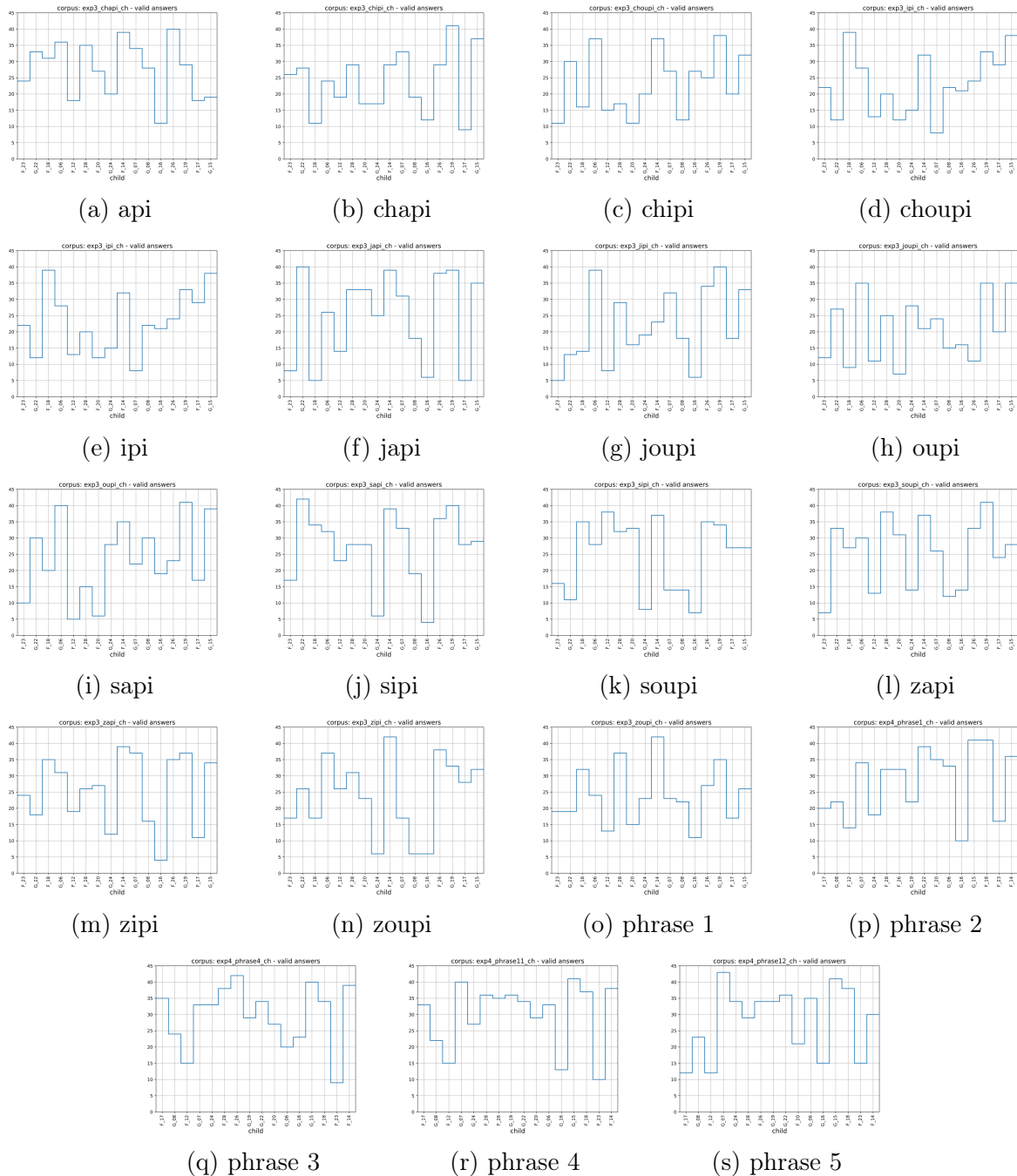


FIGURE 25 – Nombre d'identification correcte de chaque enfant, sur les 44 participations au test, à l'écoute des stimuli chuchotés

Table des figures

1	Triangle vocalique représentant les valeurs fréquentielles moyennes (en Hz) de F1 (en ordonnée) et F2 (en abscisse) des voix de femmes (triangles rouges) et d’hommes (carrés bleus) à partir des données de Calliope (1989) (Arnold 2015)	10
2	Différences entre les valeurs moyennes de F1, F2 et F3 (en bark) chez les femmes et chez les hommes, dans 26 langues (Johnson 2005)	11
3	Degré d’influence estimé des paramètres acoustiques pouvant jouer un rôle dans l’identification du genre par la voix à partir de mots dissyllabiques, pour les auditeurs francophones et anglophones américains. Le symbole « \emptyset » indique l’absence d’influence, les autres indications utilisent une échelle graduée allant de « * » (très faible influence) à « ***** » (très forte influence). (Pépiot 2013)	18
4	Triangles formantiques des voyelles de l’anglais [i :] [æ] and [u :] (gauche), et des voyelles du français [i], [a], [u] (droite) prononcées par des locutrices (bleu) et des locuteurs (rouge) bilingues (Pépiot et al. 2021)	19
5	Conséquences de la variation de l’ouverture glottique sur le spectre d’une voyelle, calculées à partir d’une simulation du conduit vocal (Morris 2003)	21
6	Spectres moyens de 24 voyelles voisées et chuchotées prononcées par des femmes (a) et des hommes (b) (Ito et al. 2005)	22
	a	22
	b	22

7	Comparaison des spectres de la voyelle [a] voisée (haut) et chuchotée (bas) d'un même locuteur (Sharifzadeh et al. 2009)	23
8	Image sagittale du conduit vocal (obtenue par TDM) sur laquelle sont indiqués les repères anatomiques utilisés pour relever les mesures des variables étudiées. Les 9 variables sont définies ainsi : la longueur du conduit vocal (VTL) correspond à un tracé curviligne entre les points D et J, indiqué en rouge ; la dimension VT-V correspond au segment [CI], en orange ; PCL, au segment [IG], en bleu clair ; NPhL, au segment [GC], en bleu foncé ; VT-H, au segment [DH], en marron ; LTh, au segment [DE] en rose ; ACL, au segment [FG], en violet ; OPhW, au segment [GH], en vert clair et enfin VT-O, au segment [EH], en vert clair (Vorperian et al. 2009).	26
9	Répartition de F1 et F2 de filles (symboles creux) et de garçons (symboles pleins) (Bennett 1981)	31
10	Identification du genre et évaluation de l'expression de genre des voix resynthétisées de garçons (<i>a</i>) et de filles (<i>b</i>) en fonction de l'espacement formantique	41
	a	41
	b	41
11	Visualisation du pseudo-mot <i>sapi</i> sur le logiciel <i>Praat</i> . La partie supérieure représente l'onde sonore ; la partie inférieure, le spectrogramme et la courbe de F0 (en bleu) calculée par le logiciel.	55
12	Tableau présentant les F0 (en Hz) de chaque enfant ainsi que les F0 moyennes (en Hz) de chaque genre, calculées pour les trois types de stimuli voisés du corpus.	57
13	Répartition des F0 moyennes de chaque enfant en fonction du genre pour les trois parties voisées du corpus. La barre verte représente la médiane ; les côtés de la boîte, le 1 ^{er} et le 2 ^e quartile et les extrémités du segment indiquent les extrema des données.	58
	a Répartition des F0 moyennes de chaque enfant calculées à partir des pseudo-mots dissyllabiques, en fonction du genre	58
	b Répartition des F0 moyennes de chaque enfant calculées à partir des phrases, en fonction du genre	58
	c Répartition des F0 moyennes de chaque enfant calculées à partir du discours semi-spontané, en fonction du genre	58

14	Présentation des F0 (en Hz) de chaque enfant et des moyennes par genre, calculées à partir des extraits de discours semi-spontané (en bleu pour les filles et en violet pour les garçons)	59
15	Présentation des STD (en Hz) de chaque enfant et des moyennes par genre, calculées à partir des extraits de discours semi-spontané. La F0 moyenne est indiquée par la barre rouge et la zone grise illustre l'étendue de la STD autour de la valeur de la F0. Les profils de filles sont en bleu et ceux des garçons en violet	60
16	Fenêtre de réponse extraite du test perceptif. Elle présente le choix forcé entre les catégories <i>garçon</i> et <i>fille</i> et l'échelle permettant d'indiquer le degré de certitude entre 0 et 5	66
17	Pourcentages de bonnes réponses obtenues sur les 44 participations, présentés pour chaque enfant en fonction du type de stimulus écouté.	67
18	Pourcentages de bonnes réponses obtenues sur les 44 participations, présentés pour chaque enfant en fonction du type de stimulus écouté. Le bleu foncé correspond à l'expérience n°1 (écoute des pseudo-mots dissyllabiques voisés) ; l'orange, à l'expérience n°2 (écoute des phrases) ; le bleu clair, à l'expérience n°3 (écoute pseudo-mots dissyllabiques chuchotés) et enfin le jaune, à l'expérience n°4 (écoute des phrases voisées)	68
19	Matrices de confusion des 4 expériences composant le test perceptif. Les colonnes indiquent le genre identifié et les lignes, le genre véritable. L'échelle de couleur indique le nombre d'occurrences de l'association de la réponse <i>garçon</i> à un stimulus ayant été produit par un garçon, pour la case supérieure gauche et ainsi de suite pour les autres cases de la matrice. Il est également indiqué dans les cases, le pourcentage que représente ce nombre sur le total des réponses données par les 44 participant·e·s pour chaque expérience.	69
a	Matrice de confusion de l'expérience 1	69
b	Matrice de confusion de l'expérience 2	69
c	Matrice de confusion de l'expérience 3	69
d	Matrice de confusion de l'expérience 4	69

20	Répartition des F0 moyennes des stimulus de l'expérience n°1 en fonction de la réponse <i>Garçon</i> ou <i>Fille</i> qui a été donnée. La barre verte représente la médiane ; les côtés de la boîte, le 1 ^{er} et le 2 ^e quartile et les extrémités du segment indiquent les extrema des données.	74
21	Répartition des valeurs de F0, F1 et F2 des enfants F14, F18, F23, F28, G07, G15, G16 et G19, calculées à partir des pseudo-mots dissyllabiques voisés. Le vert indique les enfants quasi-systématiquement correctement bien identifié-e-s ; le rouge, les enfants quasi-systématiquement correctement mal identifié-e-s. La barre verte représente la médiane ; les côtés de la boîte, le 1 ^{er} et le 2 ^e quartile et les extrémités du segment indiquent les extrema des données.	75
	a Répartition de la F0 des pseudo-mots dissyllabiques pour chaque profil particulier	75
	b Répartition des valeurs de F1 des pseudo-mots dissyllabiques pour chaque profil particulier	75
	c Répartition des valeurs de F2 des pseudo-mots dissyllabiques pour chaque profil particulier	75
22	Nombre d'identification correcte de chaque enfant, sur les 44 participations au test, à l'écoute du stimulus voisé « sapi »	76
23	Schéma du système phonatoire supérieur (Bogliotti 2010)	83
24	Nombre d'identification correcte de chaque enfant, sur les 44 participations au test, à l'écoute des stimuli voisés	89
	a api	89
	b chapî	89
	c chipî	89
	d choupî	89
	e îpî	89
	f japî	89
	g jîpî	89
	h joupî	89
	i oupî	89
	j sapî	89
	k sipî	89
	l soupî	89
	m zapî	89

n	zipi	89
o	zoupi	89
p	phrase 1	89
q	phrase 2	89
r	phrase 3	89
s	phrase 4	89
t	phrase 5	89
25	Nombre d'identification correcte de chaque enfant, sur les 44 participations au test, à l'écoute des stimuli chuchotés	90
a	api	90
b	chapi	90
c	chipi	90
d	choupi	90
e	ipi	90
f	japi	90
g	joupi	90
h	oupi	90
i	sapi	90
j	sipi	90
k	soupi	90
l	zapi	90
m	zipi	90
n	zoupi	90
o	phrase 1	90
p	phrase 2	90
q	phrase 3	90
r	phrase 4	90
s	phrase 5	90

Bibliographie

- Altenberg, E. P. et C. T. Ferrand (2006). « Fundamental frequency in monolingual English, bilingual English/Russian and bilingual English/Cantonese young adult women ». In : *Journal of Voice* 20, p. 89-96.
- Amir, O., M. Engel, E. Shabtai et N. Amir (2012). « Identification of children's gender and age by listeners ». In : *The Voice Foundation* 26.3, p. 313-321.
- Arnold, A. (2015). « La voix genrée, entre idéologies et pratiques - Une étude sociophonétique ». Thèse de doct. Université Sorbonne Paris Cité.
- Assmann, P. F., T. M. Neary et S. Bharadwaj (2008). « Analysis of a vowel database ». In : *Canadian Acoustics* 36.3, p. 148-149.
- Barreda, S. et P. F. Assmann (2018). « Modeling the perception of children's Modeling the perception of children's age from speech acoustics ». In : *Journal of the Acoustical Society of America* 143.5, EL361-EL366.
- Barreda, S. et P. F. Assmann (2021). « Perception of gender in children's voices ». In : *The Journal of the Acoustical Society of America* 150.5, p. 3949-3963.
- Bennett, S. et L. Montero-Diaz (2010). « Children's perception of speaker sex ». In : *Journal of Phonetics*, p. 113-121.
- Bennett, S. (1981). « Vowels formant frequency characteristics of preadolescent males and females ». In : *The Journal of the Acoustical Society of America* 69.1, p. 231-238.
- Bennett, S. (1983). « A 3-year longitudinal study of school-aged children's fundamental frequencies ». In : *Journal of speech and hearing research* 26.137-142.

- Blandon, R., C. G. Henton et J. B. Pickering (1984). « Towards an auditory theory of speaker normalization ». In : *Language and Communication* 4, p. 59-69.
- Bogliotti, C. (2010). « Cours de Phonétique - Université Paris Ouest ». In : *Module Langue et Littérature Sciences du Langage*. URL : https://master-fldl.parisnanterre.fr/medias/fichier/phone-utique-1_1532524252625-pdf.
- Borsel, J. V. et J. J. ans M. De Bodt (2009). « Breathiness as a Feminine Voice Characteristic : A Perceptual Approach ». In : *Journal of Voice* 23.3, p. 291-294.
- Braga, J. N., D. S. F. de Oliveira et T. M. M. Sampaio (2009). « Fundamental frequency of children's voice ». In : 11.1, p. 119-126.
- Brockmann-Bausser, M., D. Beyer et J. E. Bohlender (2015). « Reliable acoustic measurements in children between 5;0 and 9;11 years : Gender, age, height and weight effects on fundamental frequency, jitter and shimmer in phonations without and with controlled voice SPL ». In : *International Journal of Pediatric Otorhinolaryngology* 79.2035-2042.
- Brown, W. B., T. Murry, S. Singh et C. S. Hasek (1978). « Multidimensional analysis of voice in male and female children ». In : *Annual convention, American Speech-Language-Hearing Association*.
- Busby, P. A. et G. L. Plant (1994). « Formant frequency values of vowels produced by preadolescent boys and girls ». In : *The Journal of the Acoustical Society of America* 97.4, p. 2603-2606.
- Butler, J. (1990). *Gender Trouble : Feminism and the Subversion of Identity*.
- Calliope (1989). *La Parole et son traitement automatique*. Paris : Masson.
- Cartei, V., R. Banerjee, L. Hardouin et D. Reby (2019a). « The role of sex-related voice variation in children's gender-role stereotype attributions ». In : *British Journal of Developmental Psychology*.
- Cartei, V., H. W. Cowles et D. R. 2012 (2012). « Spontaneous voice gender imitation abilities in adult speakers ». In : *Plos One* 7.2, p. 1-12.
- Cartei, V., W. Cowles, R. Banerjee et D. Reby (2014). « Control of voice gender in prepubertal children ». In : *British Journal of Developmental Psychology* 32, p. 100-106.
- Cartei, V., A. Garnham, J. Oakhill, R. Banerjee, L. Roberts et D. Reby (2019b). « Children can control the expression of masculinity and femininity through voice ». In : *Royal Society Open Science Journal* 6.
- Cartei, V., J. Oakhill, A. Garnham, R. Banerjee et D. Reby (2021). « Voice cues influence children's assessments of adults' occupational competence ». In : *Journal of Nonverbal Behavior* 45, p. 281-296.

- Cartei, V. et D. Reby (2013). « Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices ». In : *Plos One* 8.12.
- Cheng, A. (2020). « Cross-linguistic f0 differences in bilingual speakers of English and Korean ». In : *The Journal of the Acoustical Society of America* 147, EL67-EL73.
- Clopper, C. G. et R. Smiljanic (2011). « Effects of gender and regional dialect on prosodic patterns in American English ». In : *Journal of Phonetics* 39.237-245.
- Cornut, G., V. Riou-Bourret et M. H. Louis (1971). « Contribution à l'étude de la voix parlée et chantée de l'enfant normal de 5 à 9 ans ». In : *Folia phoniat* 23, p. 381-389.
- Courbis, A. et S. Monterymard (2014). « La voix de l'enfant prépubert de 8 à 10 ans : étendue vocale et mécanismes laryngés ». Mém. de mast. Université Claude Bernard Lyon1 Institut des sciences et techniques de réadaptation.
- Diehl, R. L., B. Lindblom, K. A. Hoemeke et R. P. Fahey (1996). « On explaining certain male-female differences in the phonetic realization of vowel categories ». In : *Journal of Phonetics* 24, p. 187-208.
- Ditcheva, M. et J.-Y. Dommergues (2008). « Modèles acoustiques des voyelles du français et de l'anglais – voix de femme ». In : *Ploudiv University "Paissii Hilendarski" – Bulgaria, Scientific Works – Philology* 46, p. 203-210.
- Edwards, J. R. (1979). « Social class differences and the identification of sex in children's speech ». In : *Journal of Child Language* 6.1, p. 121-127.
- Eklund, I. et H. Traunmüller (1997). « Comparative Study of Male and Female Whispered and Phonated Versions of the Long Vowels of Swedish ». In : *Phonetica* 54, p. 1-21.
- Fant, G. (1975). « Non-uniform vowel normalization ». In : *Speech Transmission Laboratory, Quarterly Progress Status Report*.
- Fausto-Sterling, A. (2000). « The Five Sexes, Revisited ». In : *The Sciences* 40.4, p. 18-23.
- Fitch, W. T. et J. Giedd (1999). « Morphology and development of the human vocal tract : a study using magnetic resonance imaging ». In : *The Journal of the Acoustical Society of America* 196.3, p. 1511-1522.
- Fux, T. (2012). « Vers un système indiquant la distance d'un locuteur par transformation de sa voix ». Thèse de doct. Université de Grenoble.
- Ghorbani, A., A. Saffarian, F. Torabinezhad, Y. A. Shavaki et M. R. Keyhani (2010). « The profile of fundamental frequency changes in normal Persian-speaking individuals 9-50 years old ». In : 19.12.
- Glaze, L. E., D. M. Bless, P. Milenkovic et R. D. Susser (1988). « Acoustic characteristics of children's voice ». In : *Journal of Voice* 2.4, p. 312-319.

- Goldstein, U. (1980). « An articulatory model for the vocal tracts of growing children ». Thèse de doct. Cambridge : Massachusetts Institute of Technology.
- Guzman, M., D. Munoz, M. Viverod, N. Marine, M. Ramireze, M. T. Rivera, C. Vidal, J. Gerhard et C. Gonzalez (2014). « Acoustic markers to differentiate gender in prepubescent children's speaking and singing voice ». In : *International Journal of Pediatric Otorhinolaryngology* 78, p. 1592-1598.
- Haan, J. (2002). *Speaking of questions. An exploration of Dutch question intonation*. Utrecht, Pays-Bas : LOT.
- Haan, J. et V. J. V. Heuven (1999). « Male vs. female pitch range in Dutch questions ». In : *Proceedings of the 14th International Congress of Phonetic Sciences*, p. 1581-1584.
- Halim, M., D. Ruble, C. Tamis-LeMonda et P. Shrout (2013). « Rigidity in gender-typed behaviors in early childhood : a longitudinal study of ethnic minority children ». In : *Child Development* 84.4, p. 1269-1284.
- Hasek, C. S. et S. Singh (1980). « Acoustic attributes of preadolescent voices ». In : *The Journal of the Acoustical Society of America*, p. 1262-1265.
- Henton, C. G. (1995). « Cross-language variation in the vowels of female and male speakers ». In : *Proceedings of the 13th International Congress of Phonetic Sciences*. Sous la dir. de Stockholm, p. 420-423.
- Ingemann, F. (1968). « Identification of the speaker's sex from voiceless fricatives ». In : *Journal of the Acoustical Society of America* 44, p. 1142-1144.
- Ingrisano, D., G. Weismer et G. H. Schuckers (1980). « Sex Identification of Preschool Children's Voices ». In : *Folia Phoniatica* 32, p. 61-69.
- Ito, T., K. Takeda et F. Itakura (2005). « Analysis and recognition of whispered speech ». In : *Speech Communication* 45, p. 39-152.
- Johns-Lewis, C. (1986). « Intonation in discourse ». In : Londres, Royaume-Uni : Croom Helm. Chap. Discourse specific pitch behaviour.
- Johnson, K. (2005). « The Handbook of Speech Perception ». In : Oxford, Royaume-Uni : Blackwell Publishers. Chap. Speaker normalization in speech perception, p. 363-389.
- Johnson, K. (2006). « Resonance in an exemplar-based lexicon : the emergence of social identity and phonology ». In : *Journal of Phonetics* 34, p. 485-499.
- Jovičić, S. T. (1998). « Formant feature differences between whispered and voice sustained vowels ». In : *Acustica Acta Acustica* 84, p. 739-743.
- Jovičić, S. T. et Z. Šarić (2008). « Acoustic analysis of consonants in whispered speech ». In : *Journal of voice official journal of the Voice Foundation* 22, p. 263-274.

- Kahane, J. C. (1978). « A Morphological Study of the Human Prepubertal and Pubertal Larynx ». In : *The American Journal of Anatomy* 151, p. 11-20.
- Kessler, S. J. et W. McKenna (1978). *Gender : An Ethnomethodological Approach*. Chicago : University of Chicago Press.
- King, E. W. (1952). « A roentgenographic study of pharyngeal growth ». In : *Angle Orthodontist* 22, p. 23-37.
- Laqueur, T. W. (1990). *Making sex : Body and Gender from the Greeks to Freud*. Cambridge : Harvard University Press.
- Lass, N. J., K. R. Hughes, M. D. Bowyer, L. T. Waters et V. T. Bourne (1976). « Speaker sex identification from voiced, whispered, and filtered isolated vowels ». In : *Journal of the Acoustical Society of America* 59, p. 675-678.
- Lattner, S., M. E. Meyer et A. D. Friederici (2005). « Voice perception : Sex, pitch, and the right hemisphere ». In : *Human Brain Mapping* 24, p. 11-20.
- Lee, B. et D. V. L. Sidtis (2017). « The bilingual voice : Vocal characteristics when speaking two languages across speech tasks ». In : *Speech, Language and Hearing*.
- Lee, S., A. Potamianos et S. Narayanan (1999). « Acoustics of children's speech : developmental changes of temporal and spectral parameters ». In : *Acoustical Society of America* 105.3, p. 1455-1468.
- Linders, B., G. C. Massa, B. Boersma et P. Dejonckere (1995). « Fundamental voice frequency and jitter in girls and boys measured with electroglottography : influence of age and height ». In : *International Journal of Pediatric Otorhinolaryngology* 33, p. 61-65.
- Martin, C. L., C. Wood et J. K. Little (1990). « The development of gender stereotype components ». In : *Child Development*.
- Matsuda, M. et H. Kasuya (1999). « Acoustic Nature of The Whisper ». In : *Proceedings of Eurospeech - Budapest (Hungary)* 1.137-140.
- Mattingly, I. G. (1966). « Speaker variation and vocal-tract size ». In : *Journal of the Acoustical Society of America* 39, p. 1219.
- Maturo, S., C. Hill, G. Bunting, C. Ballif, R. Maurer et C. Hartnick (2012). « Establishment of a Normative Pediatric Acoustic Database ». In : *Archives of Otolaryngology-Head - Neck Surgery* 138.10, p. 956-961.
- Maurer, D., N. Cook, T. Landis et C. D'Heureuse (1992). « Are measured differences between the formants of men, women and children due to f0 differences ? » In : *Journal of the International Phonetic Association* 21, p. 66-79.

- Meditch, A. (1975). « The Development of Sex-specific Speech Patterns in Young Children ». In : *Anthropological Linguistics* 17, p. 421-433.
- Morris, R. W. (2003). « Enhancement and recognition of whispered speech ». Thèse de doct. Georgia institute of technology.
- Mullennix, J. W. (1995). « The perceptual representation of voice gender ». In : *Journal of the Acoustical Society of America* 98, p. 3080-3095.
- Nicollas, R., R. Garrel, M. Ouaknine, A. Giovanni, B. Nazarian et J.-M. Triglia (2008). « Normal voice in children between 6 and 12 years of age : database and nonlinear analyse ». In : *Journal of Voice* 22.6, p. 671-675.
- Nissen, S. L. et R. A. Fox (2005). « Acoustic and spectral characteristics of young children's fricative productions : A developmental perspective ». In : *The Journal of the Acoustical Society of America* 118.4, p. 2570-2578.
- Oakley, A. (1972). *Sex, gender and society*. London ; Maurice Temple Smith Ltd.
- Ordin, M. et I. Mennen (2017). « Cross-Linguistic Differences in Bilinguals' Fundamental Frequency Ranges ». In : *Journal of Speech, Language and Hearing Research*, p. 1-14.
- Pegoraro-Krook, M. I. (1988). « Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis ». In : *Folia Phoniatica* 40, p. 82-90.
- Pépiot, E. (2010). « Sur l'identification du genre par la voix chez des auditeurs anglophones et francophones ». Mém. de mast. Université Paris VIII Vincennes-Saint Denis.
- Pépiot, E. (2013). « Voix de femmes, voix d'hommes : différences acoustiques, identification du genre par la voix et implications psycholinguistiques chez les locuteurs anglophones et francophones ». Thèse de doct. Université Paris VIII Vincennes-Saint Denis.
- Pépiot, E. et A. Arnold (2021). « Cross-gender differences in french/english bilingual speakers : a multiparametric study ». In : *Perceptual and Motor Skills* 128.1, p. 153-177.
- Perry, T. L., R. Ohde et D. H. Ashmead (2001). « The acoustics bases for gender identification from children's voices ». In : *Acoustical Society of America* 109.6, p. 2988-2998.
- Peterson, G. E. et H. L. Barney (1952). « Control methods used in a study of the identification of vowels ». In : *Journal of the Acoustical Society of America* 24, p. 175-184.
- Petrushin, V. A., L. I. Tsirulnik et V. Marakova (2010). « Five dimensions of prosody : intensity, intonation, timing, voice quality, and degree of reduction ». In : sous la dir. de H. H. et R. Mixdorf. TUDpress. Chap. Five dimensions of prosody : intensity, intonation, timing, voice quality, and degree of reduction, p. 6-9.

- Pribuisiene, R., V. ULOZA et V. KARDISIENE (2011). « Voice characteristics of children aged between 6 and 13 years : Impact of age, gender, and vocal training ». In : *Logopedics Phoniatrics Vocology* 36, p. 150-155.
- Ryalls, J. H. et P. Lieberman (1982). « Fundamental frequency and vowel perception ». In : *Journal of the Acoustical Society of America* 72, p. 1631-1643.
- Sachs, J. (1975). « Language and sex : difference and dominance ». In : sous la dir. de B. T. bibinitperiod N. Henley. Rowley, MA : Newbury House Publishers. Chap. Cues in the identification of sex in children's speech.
- Sachs, J., P. Lieberman et D. Erickson (1973). *Language attitude : current trends and prospects*. Sous la dir. de B. T. bibinitperiod B. Thorne et N. Henley. Washington DC University press. Chap. Anatomical and cultural determinants of male and female speech.
- Schwartz, M. F. (1968). « Identification of speaker sex from isolated voiceless fricatives ». In : *Journal of the Acoustical Society of America* 43, p. 1178-1179.
- Schwartz, M. F. et H. E. Rine (1968). « Identification of speaker sex from isolated, whispered vowels ». In : *Journal of the Acoustical Society of America* 44, p. 1736-1737.
- Sharifzadeh, H. R., I. Mcloughlin et F. Ahamdi (2009). « Voiced Speech from Whispers for Post-Laryngectomised Patients ». In : *International Journal of Computer Science* 36.4.
- Sharifzadeh, H. R., I. Mcloughlin et M. Russell (2012). « A Comprehensive Vowel Space for Whispered Speech ». In : *Journal of Voice* 26.2, e49-e56.
- Simpson, A. P. (2000). « Gender-specific differences in the articulatory and acoustic realization of interword vowels sequences in American English ». In.
- Simpson, A. P. (2009). « Phonetic differences between male and female speech ». In : *language and Linguistics Compass* 3, p. 621-640.
- Simpson, A. P. et C. Ericsdotter (2007). « Sex-specific differences in f0 and vowel space ». In : *Proceedings of the 16th International Congress of Phonetic Sciences*. Sous la dir. de Saarbrücken, p. 933-936.
- Smith, D. R. R. (2016). « Speaker-Sex Discrimination for Voiced and Whispered Vowels at Short Durations ». In : *i-Perception* 1.13, p. 1-13.
- Sokhi, D. S., M. D. Hunter, I. D. Wilkinson et P. W. Woodruff (2005). « Male and female voices activate distinct regions in the male brain ». In : *NeuroImage* 27, p. 572-578.

- Soltani, M., H. Ashayeri, Y. Modarresi, M. Salavati et H. Ghomashchi (2014). « Fundamental Frequency Changes of Persian Speakers Across the Life Span ». In : *Journal of Voice* 28.3, p. 274-281.
- Sorenson, D. N. (1989). « fundamental frequency investigation of children ages 6-10 years old ». In : *Journal of communication disorders* 22, p. 115-123.
- Swerdlin, Y., J. Smith et J. Wolfe (2010). « The effect of whisper and creak vocal mechanisms on vocal tract resonances ». In : *Journal of the Acoustical Society of America* 127, p. 2590-2598.
- Trautmüller, H. (1988). « Paralinguistic variation and invariance in the characteristic frequencies of vowels ». In : *Phonetica* 45, p. 1-29.
- Trautmüller, H. et A. Eriksson (1994). « The size of F0 excursions in speech production and perception ». In : *Working Papers of the Department of Linguistics and Phonetics of Lund University* 43, p. 136-139.
- Vercherand, G. (2010). « Production et perception de la parole chuchotée en français : analyse segmentale et prosodique ». Thèse de doct. Université de Paris 7.
- Vorperian, H. K., S. Wang, M. K. Chung, E. M. Schimek, R. B. Durtschi, R. D. Kent, A. J. Ziegert et L. R. Gentry (2009). « Anatomical development of the oral and pharyngeal portions of the vocal tract : An imaging study ». In : *Acoustical Society of America*.
- Vuorenkoski, V., H. L. Lenko, P. Tjernlund, L. Vuorenkoski et J. Perheentupa (1978). « Fundamental voice frequency during normal and abnormal growth, and after androgen treatment ». In : *Archives of Disease in Childhood* 53, p. 201-209.
- Weinberg, B. et S. Bennett (1971). « Speakers sex recognition of 5- and 6-year-old children's voices ». In : *The Journal of the Acoustical Society of America*, p. 1210-1213.
- Weinberg, B. et S. Bennett (1979a). « Acoustic correlates of perceived sexual identity in preadolescent children's voices ». In : *The Journal of the Acoustical Society of America* 66.4, p. 989-1000.
- Weinberg, B. et S. Bennett (1979b). « Sexual characteristics of preadolescent children's voices ». In : *Journal of the Acoustical Society of America* 65.1, p. 179-189.
- White, P. (1999). « Formant Frequency Analysis of Children's Spoken and Sung Vowels Using Sweeping Fundamental Frequency Production ». In : *Journal of Voice* 13.4, p. 570-582.
- Whiteside, S. P. (1998). « Identification of a speaker's sex : a fricative study ». In : *Perceptual and Motor Skills* 86, p. 587-591.
- Whiteside, S. (1998). « Identification of a speaker's sex : a study of vowels ». In : *Perceptual and Motor Skills* 86, p. 579-584.

- Whiteside, S. P., L. Henry et R. Dobbin (2004). « Sex differences in voice onset time : A developmental study of phonetic context effects in British English ». In : *The Journal of the Acoustical Society of America* 116.2, p. 1179-1183.
- Whiteside, S. P. et C. Hodgson (2000). « Some acoustic characteristics in the voices of 6- to 10-year-old children and adults : a comparative sex and developmental perspective ». In : *Logopedics Phoniatrics Vocology* 25.122-132.
- Wu, K. et D. G. Childers (1991). « Gender recognition from speech : Part I. Coarse analysis ». In : *Journal of the Acoustical Society of America* 90, p. 1828-1840.

