



HAL
open science

**Voix de femmes, voix d'hommes : différences
acoustiques, identification du genre par la voix et
implications psycholinguistiques chez les locuteurs
anglophones et francophones**

Erwan Pépiot

► **To cite this version:**

Erwan Pépiot. Voix de femmes, voix d'hommes : différences acoustiques, identification du genre par la voix et implications psycholinguistiques chez les locuteurs anglophones et francophones. Linguistique. Université Paris VIII Vincennes-Saint Denis, 2013. Français. NNT : . tel-00821462

HAL Id: tel-00821462

<https://theses.hal.science/tel-00821462>

Submitted on 10 May 2013

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

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

UNIVERSITE PARIS 8

2 rue de la Liberté – 93526 SAINT-DENIS Cedex – France

ED 31 – Pratiques et théories du sens

EA 1569 – Transferts critiques et dynamique des savoirs
(domaine anglophone)

THESE DE DOCTORAT

Discipline : Sciences du langage

Spécialité : Phonétique / Linguistique anglaise

VOIX DE FEMMES, VOIX D’HOMMES :

DIFFERENCES ACOUSTIQUES, IDENTIFICATION DU GENRE PAR LA VOIX

ET IMPLICATIONS PSYCHOLINGUISTIQUES CHEZ LES LOCUTEURS

ANGLOPHONES ET FRANCOPHONES

Présentée par **M. Erwan PEPIOT**

Sous la direction de **M. Jean-Yves DOMMERGUES**

Membres du jury :

M. Jean-Yves DOMMERGUES

M. Pierre HALLE, pré-rapporteur

M. Philippe MARTIN, pré-rapporteur

M. Roussi NIKOLOV

Mme Jacqueline VAISSIERE

Mme Sophie WAUQUIER

3 avril 2013

REMERCIEMENTS

Je souhaiterais en premier lieu rendre hommage à mon directeur de recherche, Jean-Yves Dommergues, pour m'avoir suivi, encadré et conseillé pendant ces trois années de doctorat, comme il l'avait déjà fait lors de mes deux années de master. Son investissement, sa disponibilité et ses encouragements ont fortement contribué à l'accomplissement de ce travail. Durant ces années, il aura été d'un soutien sans faille et d'une efficacité impressionnante pour répondre à mes nombreuses demandes de relectures. En somme, le directeur de recherche idéal !

Je suis particulièrement reconnaissant envers Roussi Nikolov, qui m'a apporté lui aussi un soutien précieux et continu ainsi que des conseils avisés depuis ma première année de master. Merci à lui d'avoir accepté d'être membre du jury. Je suis très honoré que Jacqueline Vaissière ait elle aussi accepté de figurer parmi les membres du jury. Ses cours ont très largement contribué à l'élargissement de mes connaissances en phonétique expérimentale, sans lesquelles je n'aurais pu mener à terme ce travail. De plus, ses conseils et critiques constructives, prodigués à l'occasion de divers colloques, m'ont grandement rendu service. J'adresse mes sincères remerciements à Sophie Wauquier, Pierre Hallé et Philippe Martin qui ont également accepté d'être membres du jury.

Cette thèse n'aurait certainement pas été achevée aussi rapidement si je n'avais pas obtenu un contrat doctoral, qui m'a permis de me consacrer entièrement à la recherche et à l'enseignement durant ces trois dernières années. Aussi, je remercie l'école doctorale « Pratiques et Théories de Sens » et en particulier sa directrice Laurence Gavarini, pour m'avoir fait confiance en m'accordant ce contrat. Je remercie également l'Equipe d'Accueil 1569 et spécialement tous les membres du groupe LAPS, avec qui j'ai eu grand plaisir à échanger et à travailler pendant mon master et mon doctorat. En outre, je suis très heureux d'avoir pu effectuer ma charge d'enseignement au sein de l'UFR SDL de l'Université Paris 8, merci notamment à Joaquim Brandão de Carvalho de m'avoir laissé cette opportunité.

Le présent travail n'aurait jamais vu le jour sans la bonne volonté des quelques 80 participants qui ont pris part aux enregistrements et aux diverses expériences. Merci d'abord à mes étudiants, qui ont été nombreux à répondre présent pour participer à ces expériences. De même, je remercie Steve Bueno et Adèle, pour avoir recruté leurs

propres étudiants. J'adresse également mes remerciements à l'équipe du service IPT de Paris 8, qui a aimablement mis à ma disposition des salles pour les passations d'expériences, et particulièrement à Brice et Valérian, qui ont de surcroît accepté de participer à plusieurs d'entre elles !

Le recrutement des participants anglophones américains pour l'expérience d'identification du genre par la voix a pu être effectué grâce à la gentillesse et l'investissement personnel de Dominique Mougel à AUP et d'Isabelle Coydon à NYU Paris, qui ont non seulement réussi à trouver plus de 25 volontaires (many thanks to all of them !), mais m'ont également permis d'effectuer des passations d'expérience directement dans les locaux de leur institution respective. Je remercie enfin Elise, qui grâce à ses contacts, notamment au sein de TESOL France, m'a permis de recruter des locuteurs anglophones américains pour les enregistrements.

J'adresse mes remerciements à Carine André, ingénieure au Laboratoire Parole et Langage de l'Université de Provence, pour son aide lors de la création des scripts dans les logiciels Perceval et Lancelot. Merci également au Laboratoire de Phonétique et Phonologie de l'Université Paris 3 de m'avoir laissé utiliser librement leur chambre sourde. Merci enfin à Aurélie pour son soutien, sa patience et ses relectures.

RESUME

Les différences entre les voix de femmes et d'hommes relèvent de problématiques complexes et pluridisciplinaires. Le travail de recherche présenté ici traite des aspects phonétiques et psycholinguistiques de ces différences inter-genres. Le premier chapitre est consacré à une revue de la littérature du domaine. Les différences acoustiques ainsi que leur origine sont abordées, de même que le processus d'identification du genre par la voix chez les auditeurs et les conséquences psycholinguistiques de ces différences acoustiques inter-genres. L'ensemble de ces études sont présentées de manière critique et des perspectives de recherche sont évoquées.

Le deuxième chapitre présente une analyse acoustique de (pseudo-)mots dissyllabiques produits par des locuteurs anglophones du nord-est des Etats-Unis et des francophones parisiens. Les fréquences de résonance, le F0 moyen, la plage de variation de F0, le VOT, la différence d'intensité H1-H2 ainsi que la durée des énoncés ont été mesurés. Des différences inter-genres significatives ont été observées dans les deux langues sur chacun des paramètres testés. D'autre part, d'importantes variations inter-langues ont été constatées pour la plage de variation de F0, les formants vocaliques et la différence H1-H2. Ces résultats suggèrent que les différences acoustiques hommes-femmes sont en partie dépendantes de la langue.

Les enregistrements analysés ont ensuite été utilisés dans le cadre d'une expérience d'identification du genre par la voix. Cette dernière a été menée conjointement sur des auditeurs francophones parisiens avec les stimuli français, et sur des auditeurs anglophones américains avec les stimuli anglais. Les auditeurs avaient pour tâche d'identifier le genre du locuteur ayant produit le stimulus et d'indiquer leur degré de certitude. Chez les locuteurs des deux langues, les pourcentages de bonnes réponses sont significativement supérieurs au seuil de chance dès la présentation d'une consonne sourde initiale de mot et la présentation d'une voyelle initiale suffit pour obtenir des scores proches des 100 %. On constate par ailleurs que les paramètres acoustiques n'ont pas influencé de la même manière les auditeurs des deux langues. Ainsi, le F0 moyen et la qualité de voix (H1-H2) semblent avoir plus fortement influencé les auditeurs anglophones que les francophones, à l'inverse de la fréquence des formants vocaliques et de la plage de variation de F0.

Le quatrième et dernier chapitre est consacré à une expérience de détection de mots, réalisée auprès d'auditeurs francophones. Le but de cette expérience est de tester le temps de réponse des participants en fonction du genre du locuteur ayant produit le mot-cible. Les résultats suggèrent que les voix d'hommes et de femmes sont traitées par l'auditeur à vitesse équivalente, mais néanmoins comme deux entités différentes. D'autre part, aucune corrélation significative entre la fréquence fondamentale moyenne du mot-cible et le temps de réponse des auditeurs n'a pu être mise en évidence.

En conclusion, il apparaît que les différences acoustiques inter-genres tout comme le processus d'identification du genre par la voix sont fortement dépendants de la langue et donc construits socialement. Des perspectives de recherche et des applications pratiques sont proposées.

MOTS-CLES : phonétique, voix d'hommes, voix de femmes, différences acoustiques, identification du genre par la voix, traitement de la parole, différences inter-langues, locuteurs anglophones américains, locuteurs francophones.

ABSTRACT

Female and male voices: acoustic differences, gender identification from speech and psycholinguistic implications in English and French speakers.

Differences between female and male voices are linked to complex and multidisciplinary issues. The current study focuses on phonetic and psycholinguistic aspects. The first chapter is a literature review of the field. Cross-gender acoustic differences and their origin are discussed, as well as listener's gender identification from voice processing. Psycholinguistic implications are also described. All studies are presented from a critical point of view and research perspectives are proposed.

The second chapter presents an acoustic analysis of dissyllabic (pseudo-)words produced by Northeastern American English speakers and Parisian French speakers. Resonant frequencies, mean F0, F0 range, VOT, H1-H2 intensity differences and words' durations were measured. Significant cross-gender differences were obtained for each tested parameter. Moreover, cross-language variations were observed for F0 range, vocalic formants and H1-H2 differences. These results suggest that cross-gender acoustic differences are partly language dependent.

The same recordings were then used in gender identification from speech experiment. It was conducted jointly on Parisian French native listeners with French stimuli, and on American English listeners with English stimuli. Listeners had to identify the speaker's gender and indicate their degree of certainty. In both languages, percentages of correct identifications were significantly above chance for initial voiceless consonants, and close to 100 % with initial vowels. Considering the acoustic analysis performed on the stimuli, it was found that American English and French listeners did not use the same strategies. Thus, mean F0 and voice quality (H1-H2) had more influence on American English listeners' judgements than on French listeners', contrary to vowel formant frequencies and F0 range.

The fourth and last chapter is dedicated to a word spotting experiment, conducted with French native speakers. The aim of this experiment is to test the participants' response time, depending on whether the target word is produced by a male or a female voice. Results suggest that these two types of voice are processed equally fast, even though they seem processed as two different entities. Moreover, no significant correlation was found between mean F0 of the target word and response time.

To conclude, it appears that cross-gender acoustic differences and listeners' strategies in gender identification from voice are strongly language dependent and therefore socially constructed. Research perspectives and practical applications are proposed.

KEYWORDS: phonetics, male voices, female voices, male speech, female speech, acoustic differences, gender identification from speech, speech processing, cross-language differences, American English speakers, French speakers.

TABLE DES MATIERES

REMERCIEMENTS	- 3 -
RESUME	- 5 -
ABSTRACT	- 6 -
TABLE DES MATIERES	- 7 -
LISTE DES SIGLES, SYMBOLES ET ABREVIATIONS	- 11 -
INTRODUCTION	- 12 -
CHAPITRE 1 – VOIX ET GENRE : UN ETAT DE LA QUESTION	- 15 -
RESUME / ABSTRACT	- 16 -
1. SEXE ET GENRE	- 17 -
2. QUELQUES NOTIONS IMPORTANTES CONCERNANT LA PRODUCTION ET LA PERCEPTION DES SONS DE LA PAROLE	- 20 -
2.1 Les sons voisés	- 20 -
2.2 Les sons non-voisés	- 24 -
3. DIFFERENCES ACOUSTIQUES ENTRE LES PRODUCTIONS DES LOCUTEURS FEMININS ET MASCULINS	- 25 -
3.1 Fréquence fondamentale moyenne.....	- 25 -
3.2 Intonation	- 30 -
3.3 Fréquence des formants vocaliques	- 34 -
3.4 Zones de bruit des fricatives sourdes	- 43 -
3.5 Qualité de voix.....	- 46 -
3.6 Voice onset time	- 50 -
3.7 Durée des voyelles	- 55 -
3.8 Durée des énoncés	- 57 -
3.9 Intensité	- 59 -
4. IMPLICATIONS PSYCHOLINGUISTIQUES DES DIFFERENCES ENTRE VOIX D’HOMMES ET VOIX DE FEMMES	- 62 -
4.1 Traitement différencié des voix de femmes et d’hommes par le cerveau humain	- 62 -
4.2 Identification du genre par la voix	- 65 -
4.2.1 <i>Fréquence fondamentale moyenne et position des formants vocaliques : les deux indices majeurs ?</i>	- 65 -
4.2.2 <i>Zones de bruit des fricatives sourdes</i>	- 73 -
4.2.3 <i>Intonation</i>	- 75 -
4.2.4 <i>Qualité de voix</i>	- 76 -
4.2.5 <i>Influence du contexte phonique</i>	- 77 -
4.3 Conséquences des différences entre voix de femmes et voix d’hommes sur le traitement de la chaîne parlée	- 80 -
4.3.1 <i>Frontières perceptives et genre du locuteur</i>	- 80 -
4.3.2 <i>Les voix de femmes plus difficiles à traiter que celles des hommes ?</i>	- 82 -
5. DISCUSSION - CONCLUSION PARTIELLE	- 87 -
HYPOTHESES GENERALES	- 93 -

CHAPITRE 2 – ANALYSE ACOUSTIQUE DE PRODUCTIONS DE FEMMES ET D'HOMMES ANGLOPHONES AMERICAINS ET FRANCOPHONES PARISIENS	- 95 -
RESUME / ABSTRACT.....	- 96 -
1. METHODE	- 97 -
1.1 Matériau linguistique	- 97 -
1.2 Choix des locuteurs pour les enregistrements	- 99 -
1.3 Procédure d'enregistrement	- 101 -
1.4 Extraction des mots.....	- 103 -
2. ANALYSE DES DONNEES.....	- 105 -
2.1 Segmentation et étiquetage	- 105 -
2.2 Mesure du F0 et de la durée de chaque mot.....	- 107 -
2.3 Analyse acoustique des consonnes	- 108 -
2.4 Analyse acoustique des voyelles.....	- 109 -
3. RESULTATS	- 113 -
3.1 Résultats bruts.....	- 113 -
3.1.1 <i>Analyse acoustique des consonnes</i>	- 113 -
3.1.1.1 <i>Pic spectral et centre de gravité</i>	- 113 -
3.1.1.2 <i>VOT</i>	- 119 -
3.1.1.3 <i>F0</i>	- 120 -
3.1.1.4 <i>Durée</i>	- 122 -
3.1.2 <i>Analyse acoustique des voyelles</i>	- 123 -
3.1.2.1 <i>Formants vocaliques (F1, F2, F3)</i>	- 124 -
3.1.2.2 <i>Qualité de voix</i>	- 129 -
3.1.2.3 <i>F0</i>	- 130 -
3.1.2.4 <i>Durée</i>	- 132 -
3.1.3 <i>Analyse acoustique des mots dissyllabiques entiers</i>	- 133 -
3.1.3.1 <i>F0 moyen</i>	- 133 -
3.1.3.2 <i>Plage de variation de F0</i>	- 135 -
3.1.3.3 <i>Durée des mots entiers et répartition temporelle intra-mots</i>	- 137 -
3.2 Analyse statistique	- 141 -
3.2.1 <i>Analyse statistique des données obtenues pour les consonnes</i>	- 141 -
3.2.1.1 <i>Pic spectral et centre de gravité</i>	- 141 -
3.2.1.2 <i>VOT</i>	- 146 -
3.2.1.3 <i>F0</i>	- 148 -
3.2.1.4 <i>Durée</i>	- 149 -
3.2.2 <i>Analyse statistique des données obtenues pour les voyelles</i>	- 151 -
3.2.2.1 <i>Formants vocaliques (F1, F2, F3)</i>	- 151 -
3.2.2.2 <i>Qualité de voix</i>	- 156 -
3.2.2.3 <i>F0</i>	- 158 -
3.2.2.4 <i>Durée</i>	- 159 -
3.2.3 <i>Analyse statistique des données obtenues pour les mots dissyllabiques entiers</i>	- 161 -
3.2.3.1 <i>F0 moyen</i>	- 161 -
3.2.3.2 <i>Plage de variation de F0</i>	- 163 -
3.2.3.3 <i>Durée des mots entiers et répartition temporelle intra-mots</i>	- 165 -
4. DISCUSSION - CONCLUSION PARTIELLE.....	- 168 -

CHAPITRE 3 – EXPERIENCE D'IDENTIFICATION DU GENRE PAR LA VOIX CHEZ DES AUDITEURS ANGLOPHONES ET FRANCOPHONES	- 175 -
RESUME / ABSTRACT.....	- 176 -
1. METHODE	- 177 -
1.1 Matériau linguistique	- 177 -
1.2 Choix des locuteurs pour les enregistrements	- 178 -
1.3 Procédure d'enregistrement	- 180 -
1.4 Création des stimuli	- 181 -
1.4.1 Extraction des mots	- 181 -
1.4.2 Découpage des mots pour l'expérience.....	- 181 -
1.5 Participants	- 184 -
1.6 Procédure expérimentale.....	- 185 -
2. ANALYSE DES DONNEES	- 190 -
3. RESULTATS	- 191 -
3.1 Résultats bruts.....	- 191 -
3.1.1 Tous stimuli	- 191 -
3.1.2 Consonnes initiales isolées.....	- 196 -
3.1.3 Voyelles initiales isolées.....	- 199 -
3.1.4 Combinaisons « consonne + voyelle »	- 201 -
3.1.5 Mots dissyllabiques	- 204 -
3.2 Analyse statistique	- 206 -
3.2.1 Tous stimuli	- 206 -
3.2.2 Consonnes initiales isolées.....	- 211 -
3.2.3 Voyelles initiales isolées.....	- 216 -
3.2.4 Combinaisons « consonne + voyelle »	- 219 -
3.2.5 Mots dissyllabiques	- 224 -
4. DISCUSSION - CONCLUSION PARTIELLE.....	- 230 -
CHAPITRE 4 – SUR LES TEMPS DE TRAITEMENT DES VOIX DE FEMMES ET D'HOMMES.....	- 237 -
RESUME / ABSTRACT.....	- 238 -
1. METHODE	- 239 -
1.1 Matériau linguistique	- 239 -
1.2 Choix des locuteurs pour les enregistrements	- 240 -
1.3 Procédure d'enregistrement	- 241 -
1.4 Obtention des stimuli	- 242 -
1.4.1 Extraction des stimuli	- 242 -
1.4.2 Création d'un fichier son par série de stimuli	- 243 -
1.5 Participants	- 244 -
1.6 Procédure expérimentale.....	- 244 -
2. ANALYSE DES DONNEES	- 253 -
3. RESULTATS	- 254 -
3.1 Résultats bruts.....	- 254 -
3.2 Analyse statistique	- 257 -

3.2.1 Différences de temps de réponse entre auditeurs féminins et masculins	- 257 -
3.2.2 Différences de temps de réponse en fonction du locuteur produisant le mot-cible	- 258 -
3.2.3 Différences de temps de réponse en fonction du moment de présentation du mot-cible...	- 261 -
3.2.4 Le facteur « sujet »	- 262 -
3.2.5 Différences de temps de réponse en fonction des conditions expérimentales.....	- 264 -
4. DISCUSSION - CONCLUSION PARTIELLE.....	- 267 -
CONCLUSION – DISCUSSION GENERALE.....	- 271 -
PERSPECTIVES	- 274 -
REFERENCES BIBLIOGRAPHIQUES.....	- 277 -
REFERENCES SITOGRAPHIQUES	- 297 -
LISTE DES FIGURES	- 298 -
LISTE DES TABLEAUX	- 303 -
ANNEXES	- 307 -
ANNEXE 1 : Téléchargement des fichiers sons analysés et des stimuli utilisés	- 308 -
ANNEXE 2 : Participants anglophones à l'expérience d'identification du genre par la voix	- 309 -
ANNEXE 3 : Participants francophones à l'expérience d'identification du genre par la voix.....	- 310 -
ANNEXE 4 : Participants à l'expérience de détection de mots	- 311 -
ANNEXE 5 : Scripts Praat programmés et utilisés pour l'analyse acoustique	- 312 -
ANNEXE 6 : Scripts utilisé pour les expériences d'identification du genre par la voix dans le logiciel Lancelot.....	- 313 -
ANNEXE 7 : Script utilisé pour l'expérience de détection de mots dans le logiciel Perceval	- 314 -
ANNEXE 8 : Téléchargement des résultats détaillés de l'analyse acoustique	- 315 -
ANNEXE 9 : Téléchargement des résultats détaillés des expériences d'identification du genre et de détection de mots	- 316 -

LISTE DES SIGLES, SYMBOLES ET ABREVIATIONS

F0 = fréquence fondamentale

F1, F2, F3, F4 = premier, deuxième, troisième et quatrième formants vocaliques

H1, H2 = premier et deuxième harmoniques

C = consonne

V = voyelle

IRMf = imagerie par résonance magnétique fonctionnelle

QOG = quotient d'ouverture glottique

VOT / DEV = voice onset time / délai d'établissement du voisement

σ = écart-type

Δ = différence globale

dB = décibel(s)

dt = demi-ton(s)

ERB = equivalent rectangular bandwidth

Hz = hertz

kHz = kilohertz

ms = milliseconde(s)

INTRODUCTION

Les différences entre les voix de femmes et d'hommes relèvent de problématiques complexes et éminemment pluridisciplinaires. Elles renvoient non seulement à des mesures acoustiques (fréquence fondamentale, fréquences de résonance, etc.) et perceptives, mais également à des questions d'ordre anatomique (différences au plan des organes de la phonation), sociologique et même philosophique (construction de l'identité de genre, part de l'inné et de l'acquis). Le travail de recherche présenté ici traite des aspects phonétiques et psycholinguistiques de ces différences inter-genres¹.

D'un point de vue acoustique, la fréquence fondamentale moyenne (associée à la notion perceptive de hauteur mélodique) est communément considérée comme la principale différence entre les voix² d'hommes et de femmes. Elle se situerait respectivement aux environs des 120 Hz et des 200 Hz (Takefuta et al., 1972). De nombreuses études ont mis en évidence d'autres différences : les fréquences de résonance des voyelles, également appelées formants vocaliques (Hillenbrand et al., 1995), les fréquences de résonance des consonnes (Schwartz, 1968), l'intonation (Olsen, 1981), c'est-à-dire les variations de la fréquence fondamentale au cours d'un énoncé, ou encore le type de phonation (Klatt & Klatt, 1990), qui correspond à la manière de vibrer des plis vocaux³. Selon une majorité d'auteurs, les disparités acoustiques femmes-hommes s'expliqueraient en grande partie par des différences d'ordre physiologique (Fant, 1966) : plis vocaux plus longs et plus épais chez les sujets masculins, conduit vocal plus court chez les femmes. Mais alors, comment rendre compte des fortes variations inter-langues ? Il existe, par exemple, un dialecte Wu dans lequel la fréquence fondamentale moyenne serait pratiquement équivalente chez les hommes et chez les femmes (Rose, 1991). De même, en recoupant l'ensemble des études menées sur le sujet, on constate que les différences formantiques femmes-hommes semblent varier sensiblement d'une langue à l'autre (Johnson, 2005). On peut donc regretter le peu d'intérêt porté jusqu'ici à ces

¹ J'utiliserai dans ce travail le mot *genre*, traduction de l'anglais *gender*, plutôt que le mot *sexe*, afin de souligner la dimension construite des différences entre femmes et hommes. Voir section 1 du chapitre 1.

² Le terme *voix* sera ici employé dans une acceptation large, renvoyant à l'ensemble des composantes phonétiques, au niveau segmental comme suprasegmental.

³ *Plis vocaux*, de l'anglais *vocal folds*, est le terme scientifique le plus rigoureux, du point de vue de la description anatomique, pour désigner ce qui est communément appelé *cordes vocales*.

différences inter-langues, et surtout l'absence d'analyse acoustique multiparamétrique menée conjointement sur plusieurs langues.

Au plan perceptif, le rôle joué par chacun de ces paramètres acoustiques dans l'identification du genre par la voix fait l'objet de débats. Si la majorité des auteurs américains considère que la fréquence fondamentale moyenne est le principal paramètre (voir notamment Pausewang Gelfer & Mikos, 2005 ; Coleman, 1976), un auteur francophone suggère qu'il s'agirait des fréquences de résonance (Arnold, 2008, 2012). Une récente étude (Pépiot, 2010, 2011) semble indiquer que la langue maternelle des auditeurs conditionnerait leurs stratégies : il conviendrait de mettre à jour d'autres éléments permettant de valider cette hypothèse. En outre, l'influence potentielle de plusieurs paramètres acoustiques, tels que l'intonation ou encore la qualité de voix n'a pour l'heure été que peu étudiée.

Les différences acoustiques inter-genres ont également des conséquences au niveau psycholinguistique sur le traitement de la parole par l'auditeur. Sokhi et al. (2005) ont montré, grâce à l'utilisation de l'IRMf, que l'écoute de voix d'hommes et de femmes n'activait pas de la même manière certaines zones du cerveau des sujets, ce qui suggère que ces deux types de voix seraient bien traités différemment par l'auditeur. Concernant la difficulté de traitement, les résultats de deux études (Ryalls & Lieberman, 1982 ; Diehl et al., 1996) indiquent qu'au dessus de 150 Hz, la difficulté d'identification de voyelles isolées synthétisées est proportionnelle à la fréquence fondamentale des stimuli. Compte tenu de ces constatations, on pourrait penser que le temps de traitement d'une voix est corrélé au F0 de celle-ci, et donc généralement supérieur pour les voix de femmes. Une expérience de répétition de mots réalisée par Strand (2000), fait pourtant état de l'absence de différence significative de temps de réponse entre les deux types de voix. Cependant, l'étude présente certaines faiblesses méthodologiques : ces résultats nécessitent donc d'être confirmés par de nouvelles recherches.

Le présent travail se divise en quatre chapitres. Le premier d'entre eux est consacré à un état de la question. Une revue de la littérature la plus exhaustive possible sera exposée, elle porte sur tous les aspects relatifs aux voix de femmes et d'hommes : différences acoustiques, identification du genre par la voix et implications psycholinguistiques. Ces parties seront précédées d'un avant-propos permettant d'éclaircir le concept de *genre* et d'introduire les notions de bases de phonétique,

nécessaires à la bonne compréhension de ce travail de recherche. Les trois chapitres suivants sont dédiés à des études expérimentales. Pour commencer, le deuxième chapitre présente l'analyse acoustique multiparamétrique conduite sur des productions de femmes et d'hommes francophones parisiens et anglophones américains du nord-est des États-Unis : cette étude a pour but de mettre en évidence les différences acoustiques inter-genres, mais également les éventuelles variations inter-langues. Le chapitre 3 est lui consacré à une double expérience d'identification du genre par la voix, utilisant la technique du *gating*⁴ et menée conjointement sur des francophones et des anglophones américains, avec des stimuli dans la langue correspondante. Ces stimuli ne sont autres que les productions analysées dans le chapitre 2 : il sera ainsi possible de mettre en relation les réponses des auditeurs avec les données acoustiques et donc d'établir le rôle joué par les différents paramètres acoustiques dans l'identification du genre par la voix, pour les auditeurs des deux langues. Enfin, le quatrième et dernier chapitre est dédié à une expérience de détection de mots ayant pour but de mesurer les temps de traitement des voix de femmes et d'hommes. Les résultats de cette expérience apporteront notamment des éléments intéressants concernant la difficulté de traitement de ces deux types de voix chez l'auditeur.

L'intérêt d'une telle recherche est multiple. En premier lieu, ce travail permettra une meilleure compréhension des différences entre les voix de femmes et d'hommes, de leurs conséquences sur le traitement de la parole, et des processus cognitifs entrant en jeu lors de l'identification du genre par la voix. Au-delà de ces aspects théoriques, de nombreuses applications pratiques peuvent être envisagées, dans le domaine de la didactique des langues, de la phonétique clinique, du traitement automatique des langues, ainsi que pour les personnes transgenres. D'autre part, ces données pourraient être largement exploitées dans d'autres disciplines, telles que la sociologie ou la philosophie, plus spécifiquement dans le domaine des études de genre.

⁴ Paradigme de la psycholinguistique mis au point par Grosjean (1980) et consistant en un dévoilement progressif des stimuli. Voir section 1 du chapitre 3.

CHAPITRE 1

VOIX ET GENRE : UN ETAT DE LA QUESTION

RESUME

La présente étude est une revue de la littérature dans le domaine des différences entre voix de femmes et d'hommes. D'un point de vue acoustique, le F0 moyen est communément considéré comme la principale différence entre ces deux types de voix. De nombreuses études ont mis en évidence d'autres différences : formants vocaliques, fréquences de résonance des consonnes, intonation ou encore type de phonation. Selon une majorité d'auteurs, ces disparités s'expliqueraient en grande partie par des différences d'ordre physiologique (Fant, 1966). Cependant, les importantes variations inter-langues (Johnson, 2005) suggèrent qu'une partie de ces différences est le produit d'une construction sociale. Au plan perceptif, le rôle joué par chacun des paramètres acoustiques dans l'identification du genre par la voix fait l'objet de débats. Si la plupart des auteurs américains considèrent le F0 moyen comme le principal paramètre (Pausewang Gelfer & Mikos, 2005), un auteur francophone suggère qu'il s'agirait des fréquences de résonance (Arnold, 2008). Une récente étude (Pépiot, 2011) indique qu'en réalité, la langue maternelle des auditeurs conditionnerait leurs stratégies. Les différences acoustiques inter-genres ont également des conséquences sur le traitement de la parole par l'auditeur. On sait notamment que les frontières perceptives des phonèmes sont ajustées en fonction du genre du locuteur (Johnson et al., 1999) et que l'écoute de ces deux types de voix n'active pas de la même manière certaines zones du cerveau (Sokhi et al., 2005). L'ensemble des études sont présentées de manière critique et des perspectives de recherche sont esquissées.

ABSTRACT

The present study is a literature review about differences in female and male voices. Acoustically speaking, mean F0 is commonly considered as the major difference between adult male and female voices. A number of studies have brought to light other cross-gender acoustic differences: vowel formants, consonant noise, intonation, voice quality, etc. According to a majority of authors, cross-gender acoustic variations can mainly be accounted for by anatomical and physiological differences (Fant, 1966). However, cross-language variations (Johnson, 2005) suggest that cross-gender acoustic differences are partly socially constructed. The influence of these acoustic parameters in gender identification from speech is in debate. Most American authors consider that mean F0 is the main acoustical cue (Pausewang Gelfer & Mikos, 2005) but a French-speaking author (Arnold, 2008) suggests that resonant frequencies are more important. A recent study (Pépiot, 2011) indicates that the speaker's native language significantly influences his/her strategies. Cross-gender acoustical differences also have consequences on speech processing. It is known that perceptual boundaries of phonemes are adjusted as a function of speaker's gender (Johnson et al., 1999). Furthermore, female and male voices do not activate the same way some areas of the subject's brain (Sokhi et al., 2005). All studies are presented from a critical point of view and research perspectives are given.

MOTS-CLES : voix et genre, état de la question, différences acoustiques inter-genres, identification du genre par la voix, phonétique, psycholinguistique.

KEYWORDS: voice and gender, literature review, cross-gender acoustic differences, gender identification from voice, phonetics, psycholinguistics.

Dans ce chapitre 1 il sera question, dans un premier temps, de présenter la notion de genre, en évoquant les principales théories et les auteurs majeurs ayant travaillé sur le sujet. Une seconde partie sera consacrée au rappel d'éléments essentiels en phonétique, aussi bien d'un point de vue articulatoire qu'acoustique, pour comprendre les différents processus entrant en jeu lors de la production et de la perception des sons de la parole.

Une fois ces deux points préalables développés je traiterai des différences acoustiques entre les productions des locuteurs féminins et masculins, notamment sur le plan de la fréquence fondamentale moyenne, de l'intonation, des fréquences de résonance des voyelles et des consonnes, ou encore de paramètres temporels tels que le délai d'établissement du voisement⁵. Je tenterai dans cette même section d'apporter des éléments explicatifs sur l'origine de ces différences qui, comme nous allons le voir, n'est pas nécessairement que physiologique. Enfin, j'aborderai dans la dernière section de ce chapitre les problématiques psycholinguistiques liées aux différences entre les voix de femmes et d'hommes, en particulier la question de l'identification du genre à partir de la voix et les conséquences des différences acoustiques inter-genres sur le traitement de la parole chez l'auditeur.

1. SEXE ET GENRE

J'ai choisi d'utiliser dans cette étude la notion de *genre* pour renvoyer aux identités « femme » et « homme », plutôt que le terme *sexe*. Ce choix permet de mettre en évidence la dimension construite de ces identités. Ainsi, bien que n'étant pas sociologue ou philosophe, j'ai souhaité consacrer cette courte section au concept de *genre*, afin d'y présenter un bref historique, dans lequel il sera question de l'émergence de cette notion, et de son opposition au concept historique de *sexe*.

La notion de *sexe*, telle qu'elle est aujourd'hui communément utilisée dans nos sociétés occidentales, c'est-à-dire renvoyant à une catégorie binaire (« homme » / « femme ») basée uniquement sur des différences innées, semble avoir émergé dès la fin du XVIII^{ème} siècle (voir Laqueur, 1990, sur le concept de *sexe* à travers l'histoire). Au début du XX^{ème} siècle, plusieurs recherches menées par des médecins et biologistes ont

⁵ Plus communément désigné par son équivalent anglais *voice onset time*.

permis de définir de manière plus précise cette notion (voir Löwy & Rouch, 2003). Ainsi, les différences entre les deux sexes seraient non seulement d'ordre physiologique et anatomique, mais également hormonal et chromosomique. A cette époque, les comportements et traits de caractères traditionnellement associés aux hommes et aux femmes sont alors considérés comme naturels et innés, dépendant simplement du sexe de l'individu (Deaux & LaFrance, 1998).

C'est au milieu du 20^{ème} siècle que de nouvelles bases vont commencer à émerger, avec notamment la publication du *Deuxième Sexe* de Simone de Beauvoir (1949), et son célèbre « on ne naît pas femme, on le devient ». Ainsi, l'être-homme et l'être-femme ne sont plus considérés comme innés, mais comme le fruit d'une construction sociale. A partir du début des années 60, les notions de *genre* et d'*identité de genre* commencent alors progressivement à être employées, notamment dans le domaine de la psychologie et de la psychiatrie (voir Haraway, 1991), par opposition à celui de *sexe*, afin de souligner la dimension sociale et construite des identités « femme » et « homme ».

Le terme de *genre* semble à l'origine avoir été emprunté au vocabulaire de la linguistique. Unger et Crawford (1992) définissent ainsi le double usage de ce mot :

« In grammar, gender is understood as a way of classifying phenomena - a socially agreed upon system of distinctions rather than an objective description of inherent traits [...]. In a similar way, societies prescribe particular characteristics for males and females on the basis of assigned sex [...]. Thus gender must be examined as a cultural as well as a linguistic phenomenon. » (p.123)

Ce concept de *genre* sera ensuite très largement utilisé dans les mouvements féministes des années 70 (voir notamment Oakley, 1972) : les rôles et les comportements associés aux hommes et aux femmes ne relevant pas de l'inné mais d'une construction sociale, ces derniers apparaissaient donc comme *modifiables*, et la domination masculine pouvait donc être remise en cause. Par la suite, de nombreuses théories sur le genre verront le jour ; citons notamment les travaux de Wittig (1992), qui présente la norme hétérosexuelle comme instrument de domination politique et sociale et ceux, très influents, de Butler (1990, 1993), pour qui l'identité de genre n'est pas figée, mais peut évoluer tout au long de l'existence d'un individu en fonction de ses actes (théorie de la *performativité*).

Le concept de *genre* semble donc nettement plus approprié que celui de *sexe* pour rendre compte des *voix de femmes* et des *voix d'hommes*. En effet, comme nous allons le voir à travers cette étude⁶, les différences acoustiques entre ces deux types de voix ne découlent pas uniquement de différences anatomiques et physiologiques, mais également d'une part importante de construction sociale. Ainsi, la voix peut être vue comme un des éléments constitutifs de l'identité de genre.

Les explications sur le choix du terme *genre* plutôt que *sexe* ayant été avancées, la seconde section est consacrée à des notions importantes de phonétique, indispensables à la bonne compréhension de la suite de l'étude.

⁶ Voir notamment la section 3 du présent chapitre.

2. QUELQUES NOTIONS IMPORTANTES CONCERNANT LA PRODUCTION ET LA PERCEPTION DES SONS DE LA PAROLE

Dans cette section sont présentés synthétiquement quelques éléments relatifs à la production (d'un point de vue articulatoire et acoustique) et à la perception des sons de la parole. Le cadre théorique utilisé ici pour les aspects acoustiques sera celui du modèle *source-filtre*, mis au point par Gunnar Fant (1970). Ces informations permettront de clarifier les sections suivantes, consacrées aux voix de femmes et d'hommes.

2.1 Les sons voisés

Il existe dans la parole des sons dits *voisés*⁷ (voyelles, certaines consonnes) et *non-voisés* (consonnes occlusives sourdes, consonnes fricatives sourdes). Les sons voisés utilisent la vibration des *plis vocaux*, deux bandes de tissus musculaires situées au niveau du larynx. Leur longueur est de l'ordre de 10 à 20 mm (Vaissière, 2005) pour une épaisseur de 2 à 3 mm environ chez le sujet adulte⁸ (Stevens, 1999). Une représentation des plis vocaux dans le larynx est visible ci-dessous (figure 1).

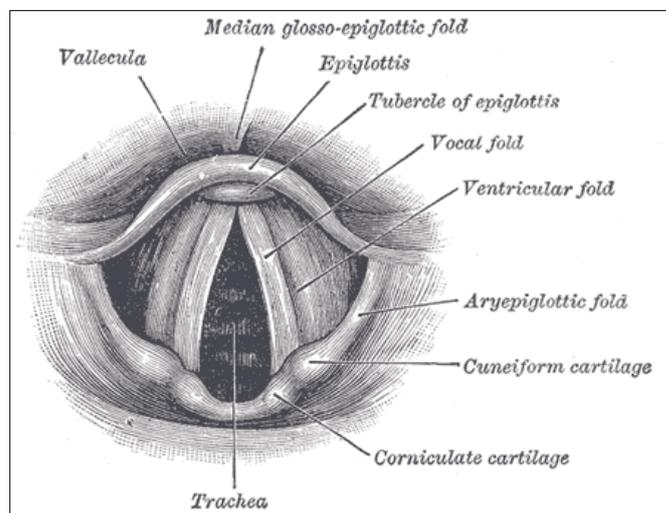


Figure 1 – Représentation de l'intérieur du larynx. Les plis vocaux (*vocal folds*) sont visibles au centre de l'image, de part et d'autre de la trachée. (Gray, 1918)

⁷ Sauf dans le cas particulier de la voix chuchotée.

⁸ Comme nous le verrons par la suite, leur taille varie significativement en fonction du genre du locuteur.

Lors de la phonation, le locuteur accole ses plis vocaux et fait augmenter légèrement la pression sub-glottique⁹ en envoyant un flux d'air à partir de ses poumons. Le nombre de vibrations des plis vocaux¹⁰ en une seconde constitue la *fréquence laryngée*, qui s'exprime en hertz (Hz). La *fréquence fondamentale* (également appelée *F0*), mesure acoustique fréquemment utilisée en phonétique, est une approximation de cette fréquence laryngée (Martin, 2008). Chez l'adulte, les plis vocaux vibrent généralement entre 80 et 300 fois par seconde lors de la phonation, avec, j'y reviendrai dans la section suivante, une importante différence femmes-hommes. D'un point de vue perceptif, la *hauteur tonale* de la voix correspond principalement à la fréquence fondamentale : plus elle est élevée, plus le son est perçu comme aigu. Les variations de la fréquence fondamentale dans le temps constituent l'*intonation*¹¹, c'est-à-dire la *mélodie* des énoncés.

Sur l'échelle tonale, pour qu'un son soit perçu deux fois plus aigu (octave supérieur), il faut *doubler* sa fréquence fondamentale (voir Martin, 2008). Ainsi, si un son a une fréquence fondamentale de 75 Hz, un son à 150 Hz sera perçu comme deux fois plus aigu que ce dernier (soit une différence de 75 Hz). Mais pour qu'un son puisse être perçu deux fois plus aigu que celui à 150 Hz, il faudra passer à une fréquence fondamentale de 300 Hz (+ 150 Hz et non + 75 Hz). Par conséquent, plus une variation brute de fréquence fondamentale (par exemple, + 50 Hz) a lieu dans les basses fréquences, plus la variation perçue sera grande (Stevens, 1999) : cet élément prendra toute son importance lorsque nous aborderons, un peu plus loin dans ce chapitre, les différences femmes-hommes sur le plan de l'intonation.

Les plis vocaux peuvent donc vibrer plus ou moins rapidement, mais ils peuvent également vibrer de manière plus ou moins *ample*, en fonction du niveau de pression sub-glottique. Ainsi, plus cette dernière est élevée, plus *l'amplitude* de la vibration sera grande, et plus le son sera alors perçu comme fort : c'est la notion d'*intensité*, que l'on mesure généralement en décibels (dB).

⁹ Sous les plis-vocaux.

¹⁰ C'est-à-dire le nombre de cycles d'ouverture-fermeture des plis vocaux.

¹¹ Le terme *intonation* sera utilisé ici dans cette définition étroite (voir Vaissière, 2005).

En dehors du type de phonation normal, dit *modal* (mise en vibration de la totalité des plis-vocaux, fermeture complète dans chaque cycle, vibration relativement régulière) les plis vocaux peuvent également vibrer de manière différente¹² : c'est notamment le cas de la voix *breathy* ou soufflée¹³, avec laquelle la fermeture de la glotte n'est jamais complète (voir Ladefoged, 1971 ; Fritzell et al., 1986), et de la voix craquée, également appelée *creaky* ou *fry*, qui présente une vibration irrégulière des plis vocaux, généralement aux alentours de 20 à 70 Hz, et surtout une ouverture incomplète et plus courte de ces derniers lors de chaque cycle (Klatt & Klatt, 1990). Les différents types de phonation (ou *qualités de voix*) sont présentés dans la figure 2 ci-dessous sous la forme d'un continuum basé sur le degré d'aperture des plis vocaux.

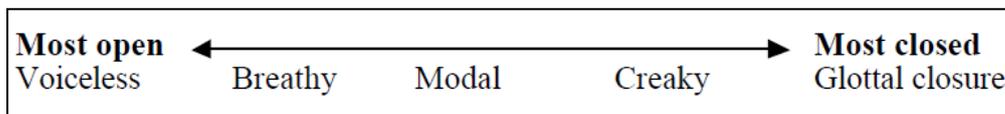


Figure 2 – Continuum des types de phonation en fonction du degré d'aperture des plis vocaux. (Gordon & Ladefoged, 2001)

La vibration des plis vocaux constitue ce que l'on nomme la *source* (Fant, 1970) des sons voisés¹⁴. La fréquence fondamentale génère automatiquement d'autres fréquences dites *harmoniques*, qui sont des multiples entiers de cette dernière : par exemple, avec F0 à 200 Hz, des harmoniques apparaissent à 400 Hz, 600 Hz, 800 Hz, 1000 Hz, 1200 Hz, etc. Leur amplitude est alors globalement décroissante à mesure que l'on monte dans les fréquences, comme le montre la figure 3 ci-après.

¹² De nombreux auteurs (e.g. Klatt & Klatt, 1990 ; Gordon, 2001 ; Podseva, 2007) font référence à ces différents types de phonation par le terme *qualité de voix* (*voice quality*). Mais ce dernier renvoie souvent à une définition plus large, c'est-à-dire à l'ensemble des propriétés suprasegmentales résultant de la configuration de l'appareil vocal (Laver, 1980 ; Coadou, 2007). J'utiliserai pour ma part indifféremment ces deux notions (*type de phonation* et *qualité de voix*) pour renvoyer aux voix *creaky*, *breathy*, et modale.

¹³ Parfois traduit par « voix aspirée » (voir par exemple Liu et al., 1996).

¹⁴ Pour les consonnes obstruentes voisées (occlusives et fricatives) cette dernière est combinée à une autre source : le bruit de friction au niveau du lieu d'articulation de la consonne.

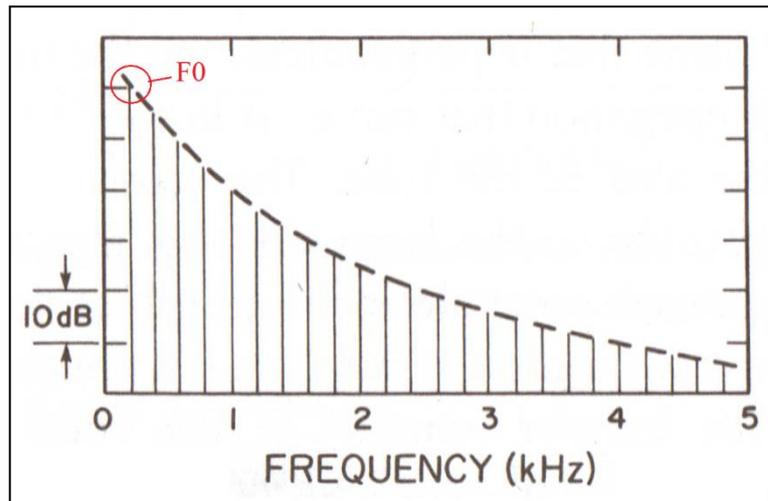


Figure 3 – Représentation schématique du spectre fréquentiel d'un son produit par une vibration des plis vocaux à 200 Hz, avant le passage dans le conduit vocal. On y voit la fréquence fondamentale (notée F_0) et les fréquences harmoniques. (D'après Stevens, 1999)

L'ensemble de ces fréquences passent ensuite dans le *conduit vocal*, c'est-à-dire l'espace situé entre le larynx et les lèvres, qui va agir comme un *filtre* : en fonction de la configuration de ce dernier (position de la langue, des lèvres, etc.), certaines zones de fréquences vont être renforcées (amplitude augmentée) et d'autres affaiblies (amplitude diminuée). Les zones de fréquences renforcées sont appelées *fréquences de résonance*. Ces dernières sont, en théorie, indépendantes de la fréquence fondamentale¹⁵ : si deux sons voisés sont produits avec une fréquence fondamentale différente (par exemple, 100 Hz et 150 Hz), mais avec une configuration du conduit vocal identique, les fréquences de résonance se situeront dans les mêmes zones.

Pour les voyelles, ces fréquences de résonance sont appelées *formants vocaliques*, que l'on nomme F_1 , F_2 , F_3 , F_4 , etc. Les formants sont ordonnés en fonction de leur fréquence. Par exemple, pour le [u] français, F_1 se situe aux alentours des 300 Hz, F_2 à environ 800 Hz et F_3 est proche des 2000 Hz (voir Calliope, 1989). La figure 4, ci-après, montre le spectre fréquentiel typique d'une voyelle centrale, produite avec une fréquence fondamentale à 200 Hz.

¹⁵ Dans le modèle mis au point par Fant (1970), la *source* et le *filtre* sont présentés comme indépendants l'un de l'autre. Néanmoins, certains auteurs ont depuis remis en question cette stricte indépendance (voir notamment Childers & Wong, 1994 ; Titze & Story, 1997 ; Barney et al., 2007).

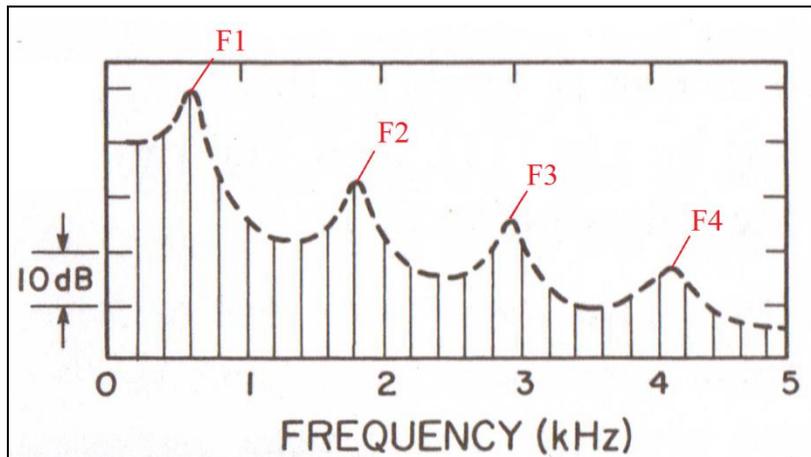


Figure 4 – Représentation schématique du spectre fréquentiel d'une voyelle centrale, produite avec une fréquence fondamentale à 200 Hz. Les quatre premiers formants (F1, F2, F3, F4) sont indiqués en rouge. (D'après Stevens, 1999)

La position de ces formants constitue le *timbre vocalique* (notion d'ordre perceptif), qui permet à l'auditeur de distinguer entre les différentes voyelles¹⁶.

2.2 Les sons non-voisés

Concernant les sons non-voisés, à savoir les occlusives ([p], [t], [k]...) et fricatives ([f], [s], [ʃ]...) sourdes, le procédé est similaire. La source n'est alors plus la vibration des plis vocaux, mais le *bruit de friction* créé par la constriction en un point du conduit vocal (consonnes fricatives), ou le relâchement brutal à la suite d'une occlusion (consonnes occlusives). Ce *bruit* se répartie dans tout le spectre fréquentiel. Il va ensuite être lui aussi renforcé ou affaibli à certaines fréquences en fonction de la configuration du conduit vocal.

Quelques concepts essentiels liés à la production et à la perception des sons de la parole ont maintenant été introduits. La section suivante est consacrée aux différences acoustiques entre les voix de femmes et les voix d'hommes.

¹⁶ Les 3 premiers formants (F1, F2 et F3) permettent à eux seuls de définir précisément une voyelle (voir Vaissière, 2006).

3. DIFFERENCES ACOUSTIQUES ENTRE LES PRODUCTIONS DES LOCUTEURS FEMININS ET MASCULINS

De très nombreuses études ont mis en évidence des différences acoustiques significatives entre les productions des locuteurs masculins et féminins. Dans cette section, je présente de manière critique les principales recherches menées sur le sujet, en prenant un à un les différents paramètres acoustiques investigués. Je détaille également les tentatives d'explication avancées par les auteurs, qu'elles soient d'ordre physiologique ou socio-phonétiques.

3.1 Fréquence fondamentale moyenne

Le F0 moyen est traditionnellement considéré comme la principale différence entre les voix d'hommes et les voix de femmes (voir notamment Coleman, 1983, et Henton, 1989a), avec des valeurs moyennes tournant généralement autour de 120 Hz pour les hommes, et de 240 Hz pour les femmes (Vaissière, 2006). Comme cela est rappelé dans l'ouvrage collectif Calliope (1989), il convient néanmoins de prendre en compte différents facteurs de variation :

« La gamme de variation moyenne de la fréquence fondamentale dépend, essentiellement, de l'âge et du sexe du locuteur. Elle se situe aux alentours de 100 à 150 Hz pour l'homme adulte, et de 140 à 240 Hz pour la femme adulte, mais peut présenter des variations considérables chez un même locuteur selon le type de phrase prononcée, son attitude et état émotif lors de l'acte de parole, etc. »
(p.133)

D'autres auteurs s'accordent également pour dire que l'âge du locuteur jouerait un rôle crucial sur le F0 moyen des hommes et des femmes. Chez les hommes, le F0 moyen continuerait de descendre légèrement même après la puberté, et ce jusqu'à environ 35 ans (Traunmüller & Eriksson, 1995), avant d'entamer une stabilisation pendant une vingtaine d'années. Vers l'âge de 55 ans la fréquence fondamentale moyenne tend alors à augmenter (Hollien & Ship, 1972 ; Kitzing, 1979 ; Pegoraro-Krook, 1988 ; Harnsberger et al., 2008 ; Torre & Barlow, 2009 ; Gugatschka et al., 2010), jusqu'à la fin de la vie. Pour les femmes, le F0 moyen reste relativement constant jusqu'à l'âge de la ménopause, puis décroît jusqu'à la fin de vie (Chevrie-Muller et al., 1971; Kitzing, 1979 ; Stoicheff,

1981; Pegoraro-Krook, 1988 ; Torre & Barlow, 2009). Cette diminution serait de l'ordre d'une quinzaine de hertz entre l'avant-ménopause et l'âge de 70 ans (Chevrie-Muller et al., 1971). Ces différentes données sont présentées de manière synthétique dans la figure ci-dessous.

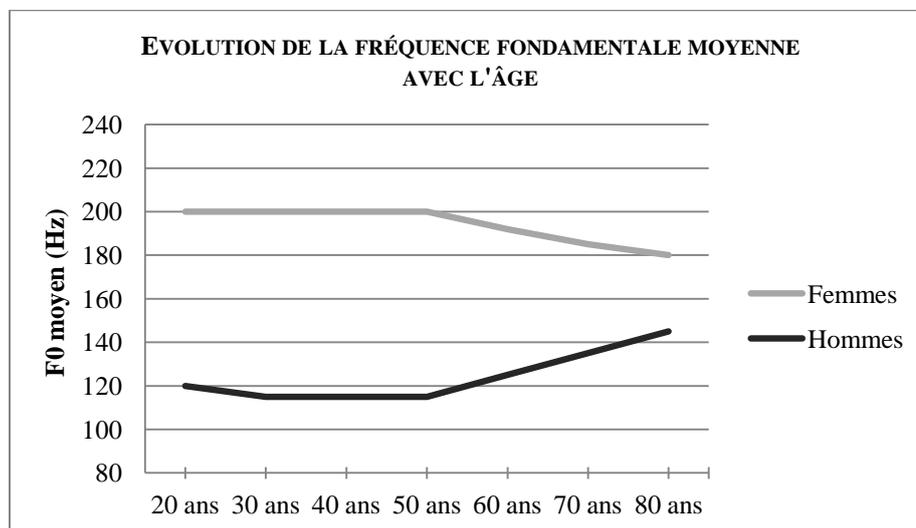


Figure 5 – Représentation schématique de l'évolution du F0 moyen des femmes (courbe grise) et des hommes (courbe noire) adultes avec l'âge. (D'après Pegoraro-Krook, 1988 et Traunmüller & Eriksson, 1995)

D'autre part, une consommation élevée de tabac entraînerait une baisse assez forte de la fréquence fondamentale moyenne. Gilbert et Weismer (1974) ont mesuré le F0 moyen de deux groupes de locutrices anglophones américaines (15 fumeuses¹⁷ et 15 non-fumeuses, âgées de 30 à 54 ans) lors de la lecture d'un texte et en discours spontané. Les résultats font état d'une différence de 19 Hz en moyenne entre femmes non-fumeuses et fumeuses en lecture (183 Hz contre 164 Hz), et de 8 Hz en discours spontané (176 Hz contre 168 Hz). Dans une étude similaire, Sorensen et Horii (1982) ont retrouvé cette même diminution du F0 moyen avec la consommation de tabac chez des locuteurs de genre masculin. Selon ces différents auteurs, ce phénomène s'explique par l'épaississement des plis vocaux chez les sujets fumeurs, observé notamment par Auerbach et al. (1970).

A ces différents facteurs, s'ajoute un autre élément déterminant : la langue parlée (Dolson, 1994). En effet, au regard des différentes études statistiques menées sur la

¹⁷ Avec une consommation au moins égale à un paquet de cigarettes par jour, depuis plus de 15 ans.

fréquence fondamentale moyenne des hommes et des femmes, il ressort de fortes différences inter-langues. 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. Autre élément intéressant, dans une recherche particulièrement originale conduite sur neuf femmes bilingues anglais / russe, Altenberg et Ferrand (2006) ont observé que ces locutrices augmentaient significativement leur F0 moyen lors de l'emploi de la langue russe (208 Hz en moyenne contre 190 Hz pour l'anglais).

Si l'on s'intéresse aux recherches menées sur des langues asiatiques, d'importantes variations sont également présentes : 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 taïwanais 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.

Cette dernière étude est particulièrement remarquable car la différence obtenue (17 Hz, qui équivalent ici à seulement 1,65 demi-ton¹⁸ sur le plan perceptif) est extrêmement faible en comparaison avec celles traditionnellement présentes dans les langues européennes et même asiatiques, et ce principalement à cause du F0 moyen exceptionnellement élevé chez les locuteurs masculins de ce dialecte. Une même tendance à un F0 moyen très élevé chez les hommes a également été observée chez les locuteurs d'un dialecte suédois parlé dans la province de Småland, située au sud de la Suède (Elert and Hammarberg, 1991). Les résultats obtenus dans ces deux dialectes font néanmoins figure d'exception.

¹⁸ Pour deux fréquences de 170 Hz et 187 Hz, l'écart en demi-tons correspond à $40 * \text{Log}10 (187/170)$, soit environ 1,65.

La physiologie des plis vocaux comme seule explication ?

Selon la grande majorité des auteurs (voir entre autres Coleman, 1983, et Titze, 1989), les différences de fréquence fondamentale moyenne entre les hommes et les femmes s'expliqueraient par la physiologie des plis vocaux. D'après Kahane (1978) et Titze (1989), les plis vocaux des hommes seraient en moyenne 60 % plus longs que ceux des femmes, comme l'illustre la figure 6.

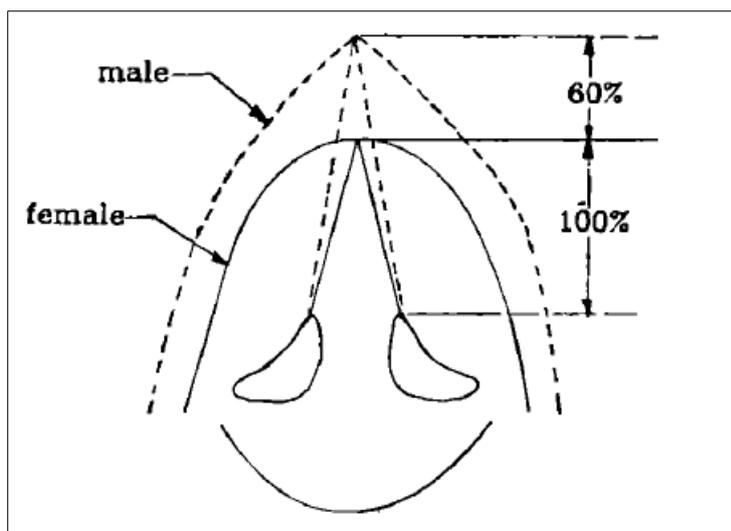


Figure 6 – Coupe horizontale du larynx d'un homme (en pointillés) et d'une femme (trait continu), les pourcentages dans la partie droite illustrent la différence de taille des plis vocaux. (Kahane, 1978)

Les plis vocaux étant plus courts et moins épais chez les femmes ils auraient donc tendance, pour un effort phonatoire équivalent, à vibrer à une fréquence plus élevée que ceux des hommes (voir Hollien, 1960, Titze, 1976 et Nishizawa et al., 1988). Selon Traunmüller et Eriksson (1995), les locuteurs cherchent à minimiser l'effort nécessaire à la mise en vibration de leurs plis vocaux. L'effort phonatoire augmentant généralement avec le F_0 produit (voir Ladefoged, 1967), ils utiliseraient donc préférentiellement la partie basse de leur plage de variation physiologique de F_0 ¹⁹, qui correspond effectivement aux fréquences fondamentales moyennes observées couramment dans la

¹⁹ Cette dernière est le plus souvent représentée sous forme de phonétogramme (voir Teston, 2004). Ce type de profil de F_0 est fréquemment utilisé en médecine et en orthophonie.

littérature : de l'ordre de 100 à 150 Hz pour les hommes, et de 170 à 220 Hz pour les femmes (voir Pabon & Plomp, 1988).

Cette explication physiologique est corroborée par des observations faites par Busby & Plant (1995) et Lee et al. (1999) sur des enfants, ne présentant donc pas encore de grandes différences anatomiques au niveau des plis vocaux. Dans ces deux études, réalisées sur des locuteurs anglophones, aucune différence significative n'a pu être mise en évidence entre le F0 moyen des filles et celui des garçons.

Les différences dans l'évolution du F0 moyen avec l'âge chez les locuteurs adultes seraient, elles aussi, liées à la physiologie des plis vocaux. Ces derniers seraient affectés par les changements hormonaux, notamment par le ratio testostérone / œstrogènes qui évolue considérablement avec l'âge, en particulier chez les locuteurs féminins (voir Traunmüller & Eriksson, 1995).

Cependant, comme le rappelle Simpson (2009), les différences physiologiques ne permettent pas d'expliquer les importantes disparités inter-langues évoquées plus haut :

« 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. [...] While some cross-linguistic differences may be attributable to differences in measurement and the material recorded, others are not. [...] 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. » (p.625)

Les différences entre la fréquence fondamentale moyenne des hommes et celle des femmes seraient donc en partie dues à des comportements sociaux, et pas uniquement à des différences physiologiques. Il est fort probable que le poids respectif de ces deux facteurs varie selon la langue parlée. Ainsi, la théorie du moindre effort physiologique, avancée par Traunmüller et Eriksson (1995), semble s'appliquer parfaitement à la majorité des langues étudiées (français standard, anglais britannique et américain, mandarin standard, etc.) mais pas aux deux exceptions mentionnées plus haut, à savoir le dialecte Wu (Rose, 1991) et le dialecte suédois de la province du Småland (Elert and Hammarberg, 1991). Pour ces deux dialectes, le F0 moyen exceptionnellement élevé des

locuteurs masculins relève très certainement d'une variation sociolinguistique, c'est-à-dire d'une conduite sociale apprise et non d'un phénomène d'origine physiologique.

3.2 Intonation

Dans cette sous-section je m'intéresse à un autre paramètre acoustique lié à la fréquence fondamentale : l'intonation, que je définirai ici de manière étroite (voir Vaissière, 2005) comme *les variations de la fréquence fondamentale dans le temps*.

Plusieurs études ont tenté de mettre en évidence l'existence de schémas intonatifs différents en fonction du genre (voir notamment Brend, 1971 ; Elyan, 1978 ; McConnell-Ginet, 1978 ; Olsen, 1981 ; Huber, 1989 ; Tranmüller & Eriksson, 1995 ; Warren & Daly, 2000 et Daly & Warren, 2001). Ainsi, le genre du locuteur conditionnerait la fréquence d'utilisation de certains contours mélodiques : Edelsky (1979), avance par exemple que le contour *montant-descendant-montant* serait plus utilisé par les femmes que par les hommes anglophones américains dans certains contextes ; Clopper & Smiljanic (2011), montrent quant à elles que ces préférences pour l'utilisation de certains contours intonatifs existent bien mais semblent dépendre fortement du dialecte parlé, tout du moins en anglais américain.

De manière plus générale et inter-langues, la principale différence observée concerne la plage de variation de F0 (*F0 range* dans la terminologie anglo-saxonne), qui serait plus grande chez les locuteurs féminins (e.g. Takefuta et al., 1972, pour l'anglais américain ; Hwa Chen, 2007, pour le mandarin). Cela signifie que les mouvements de leur courbe de F0 seraient plus amples, atteignant ainsi des cibles (hautes ou basses) plus extrêmes que chez les locuteurs masculins.

Cependant, comme le signalent à juste titre Graddol (1986), Henton (1989a) et Simpson (2009), beaucoup des études menées sur ce sujet utilisent une échelle linéaire (en hertz) pour établir leurs comparaisons, or la perception de la hauteur mélodique (ou *pitch*) par l'oreille humaine est logarithmique²⁰ : une variation de 50 Hz entre 80 et 130 Hz sera perçue comme plus importante qu'une même variation entre 200 et 250 Hz, par exemple (voir Stevens, 1999).

²⁰ Voir section 2.1.

Il serait donc normal que les femmes, ayant un F0 moyen globalement plus élevé, utilisent une plage de variation plus grande pour arriver au même rendu perceptif que leurs homologues masculins. Afin de tester cette hypothèse il convient donc, comme l'ont fait Henton (1989a), Traummüller & Eriksson (1994), ou encore Simpson (2009), de convertir les données brutes (en hertz) en une échelle logarithmique qui rende compte de la perception de la hauteur par l'oreille humaine, telle que le demi-ton, le bark²¹ ou encore l'ERB²². Une illustration de ce procédé est visible ci-après avec la figure 7.

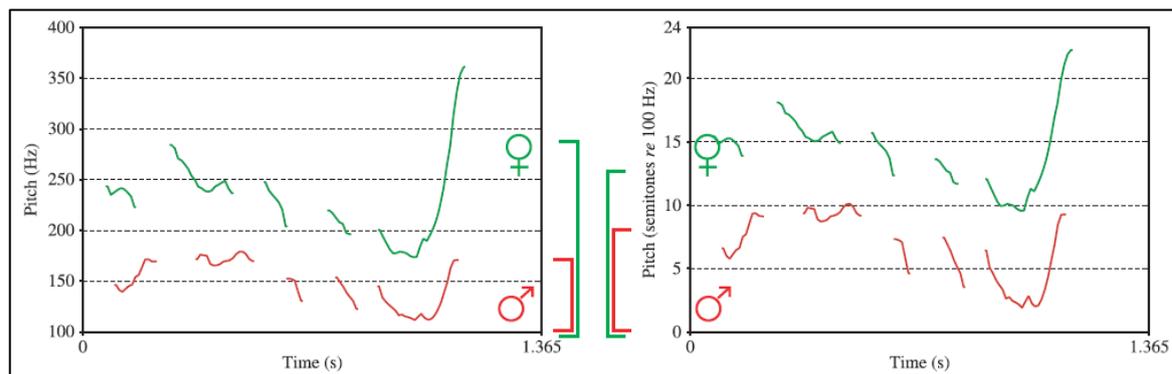


Figure 7 – Représentations de la fréquence fondamentale d'un homme (en rouge) et d'une femme (en vert) pour la phrase allemande « Riechst du nicht die frische Luft? » avec une échelle linéaire en hertz (à gauche) et en demi-tons (à droite). Les accolades verticales au centre indiquent les plages de variation pour chacune des représentations. (Simpson, 2009)

Comme on peut le voir sur cette figure, la différence entre la plage de variation de la voix d'homme (en rouge) et celle de la voix de femme (en vert), paraît très importante lorsque l'on utilise une échelle linéaire en hertz, sur la partie gauche de la figure. En revanche, cette différence diminue fortement lorsqu'on convertit les hertz en demi-tons (partie droite de l'illustration). Henton (1989a) a effectué cette conversion *hertz vers demi-tons* sur des données tirées de plusieurs articles antérieurs ainsi que sur des données recueillies par ses soins à l'occasion de cette étude. Elle en conclut qu'il n'existe aucune différence significative entre hommes et femmes concernant la plage de variation de F0 chez les locuteurs de l'anglais américain. Dans une étude antérieure similaire, menée cette fois-ci sur une centaine de locuteurs germanophones, Herbst (1964, 1969) conclut même

²¹ Voir Zwicker, 1961.

²² Voir Moore et Glasberg, 1983.

à l'existence d'une plage de variation très légèrement plus importante chez les hommes que chez les femmes, tout comme Graddol (1986).

Un point important semble pourtant avoir été négligé dans les études de Henton et de Herbst : le matériau linguistique utilisé pour ces recherches est constitué quasi-exclusivement de phrases déclaratives, impliquant un contour intonatif descendant. Pour une étude du même type portant sur le hollandais, Haan et Van Heuven (1999) ont quant à eux utilisé des phrases interrogatives, aussi bien en discours spontané qu'en lecture. Selon ces auteurs, il convenait de ne pas négliger les phrases interrogatives car :

« female speech has been shown to be more expressive, more involved, more listenerdirected [. . .] than male speech » (p.1581)

Ainsi, les phrases interrogatives seraient plus propices à l'expression de différences hommes-femmes dans les schémas intonatifs que les phrases déclaratives, considérées comme plus neutres. Si l'on peut trouver cette justification des auteurs peu convaincante car fondée sur des stéréotypes sexistes, les résultats ont néanmoins été concluants : Haan et Van Heuven sont parvenus à montrer que les locutrices du hollandais utilisaient une plage de variation significativement plus large que les hommes pour les phrases interrogatives (voir Haan & Van Heuven, 1999 et Haan, 2002).

Cela rejoint les constatations faites par Traunmüller et Eriksson (1995), pour qui la plage de variation des femmes serait plus grande que celle des hommes uniquement pour les types de discours qualifiés par ces auteurs de hautement « vivants » (*lively*). Traunmüller et Eriksson sont arrivés à cette conclusion après avoir classé un grand nombre de recherches réalisées sur le *F0 range* des hommes et de femmes en fonction du degré de *liveliness*²³ du type de production enregistrée (phrase lue, texte lu, discours spontané, conversation etc.). Les auteurs reconnaissent néanmoins que l'échelle qu'ils ont établie est largement subjective.

Un autre facteur de variation a été mis en évidence par Gilbert and Weismer (1974), dans une étude portant sur des locutrices de l'anglais américain. Ces auteurs ont montré que la consommation de tabac n'avait pas seulement un effet sur le F0 moyen, mais jouait également sur la plage de variation, cette dernière étant plus large chez les

²³ Terme que l'ont pourrait éventuellement traduire en français par « vivacité ».

sujets fumeurs. Exprimé en demi-tons, l'écart-type moyen était de 3,25 pour les locutrices fumeuses, contre seulement 2,35 pour les non-fumeuses. Cette différence s'est avérée nettement significative. Malgré la mise en évidence de ce phénomène, une partie des auteurs ayant travaillé sur les plages de variation du F0 des hommes et des femmes n'en ont pas tenu compte pour le choix de leurs participants, qui pouvaient être indifféremment fumeurs ou non-fumeurs. La proportion de fumeurs étant plus grande chez les hommes que chez les femmes, on peut donc s'attendre pour ces études à un biais favorisant une plage de variation supérieure pour les hommes.

Dernier élément important, une recherche menée par Pegoraro-Krook (1988) met en lumière une forte influence de l'âge des locuteurs sur les différences de *F0 range* entre hommes et femmes. Son étude a été réalisée sur plus de 650 locuteurs du suédois âgés de 20 à 79 ans. Si l'on regarde les résultats obtenus en fonction des classes d'âge, on s'aperçoit que chez les sujets âgés de 50 à 79 ans, il n'existe aucune différence significative entre hommes et femmes. En revanche, chez les 20-50 ans, la plage de variation des femmes est significativement plus grande que celle de leurs homologues masculins. Cela suggère que les différences hommes-femmes sur le plan du *F0 range* tendent à disparaître avec l'âge.

Une origine multifactorielle ?

Malgré l'existence de ces nombreux facteurs faisant varier les résultats d'une étude à l'autre (langue des locuteurs, type de discours enregistré, âge des locuteurs, consommation de tabac, etc.), il semble qu'il existe effectivement une tendance générale à une plage de variation plus grande chez les locutrices, et ce dans plusieurs langues. Cette différence femmes-hommes serait, selon la majorité des auteurs (e.g. Haan & Van Heuven, 1999), une variation d'ordre socio-linguistique. Cependant, certains auteurs sont moins catégoriques :

« [...] *it appears likely that the difference is due to physiological, or possibly psychological or cultural factors. It may, for example, be the case that a smooth transition into the falsetto register is more difficult to achieve for men than for women or that men, perhaps as an adaptation to such a physiological factor, consider a lower degree of liveliness to be appropriate.* » (Traunmüller & Eriksson, 1995 : 6)

Selon ces auteurs, la physiologie des plis vocaux pourrait donc jouer elle aussi un rôle dans ces différences de *F0 range*. Une fois encore, il semble que nous ayons à faire à un phénomène complexe, dont la cause ne peut se résumer à un seul facteur.

La fréquence d'utilisation de certains contours intonatifs en fonction du genre du locuteur serait quant à elle une variation purement socio-linguistique, car hautement variable en fonction de la langue et même du dialecte parlé (voir notamment Clopper & Smiljanic, 2011).

3.3 Fréquence des formants vocaliques

Le timbre vocalique, notion perceptive qui correspond acoustiquement à la position des formants, semble être également une donnée hautement variable en fonction du genre du locuteur. En effet, un très grand nombre d'études reposant sur l'analyse acoustique de voix d'hommes et de femmes ont révélé d'importantes variations dans la position des formants vocaliques²⁴ selon le genre du locuteur : voir notamment Yang (1992) sur le coréen, Trittin & De Santos y Lleo (1995) sur l'espagnol, Ericsson (2005) sur le suédois, Calliope (1989), Gendrot & Adda-Decker (2005), Pépiot (2009) et Kamiyama (2009) sur le français, Henton (1985) et Whiteside (1998a, 2001) sur l'anglais britannique, Peterson & Barney (1952), Bachorowski & Owren (1999), Clopper (2004) et Hillenbrand et al. (1995) sur l'anglais américain.

Cette dernière étude a été menée sur 45 locuteurs et 48 locutrices, originaires pour la plupart du Michigan. Les valeurs moyennes de F1 et F2 pour les hommes et les femmes obtenues par Hillenbrand et al. sont présentées sous forme de triangle vocalique ci-après (figure 8). Ce type de représentation en deux dimensions permet de rendre compte, au moins partiellement, du lieu d'articulation de la voyelle : le degré d'aperture étant proportionnel à la valeur du premier formant et l'antériorité de la langue globalement proportionnelle à la valeur du deuxième formant²⁵.

²⁴ Pour la grande majorité de ces études, il est question des deux premiers formants (F1, F2), parfois du troisième (F3) et plus rarement du quatrième (F4).

²⁵ Notons néanmoins que ces corrélations sont loin d'être parfaites. Par exemple, l'arrondissement des lèvres entraîne une diminution de la valeur de F2. D'autre part, ce type de représentation est souvent critiqué (voir Vaissière, 2006) car il néglige totalement le troisième formant, une donnée pourtant capitale pour l'analyse acoustique et l'identification des voyelles.

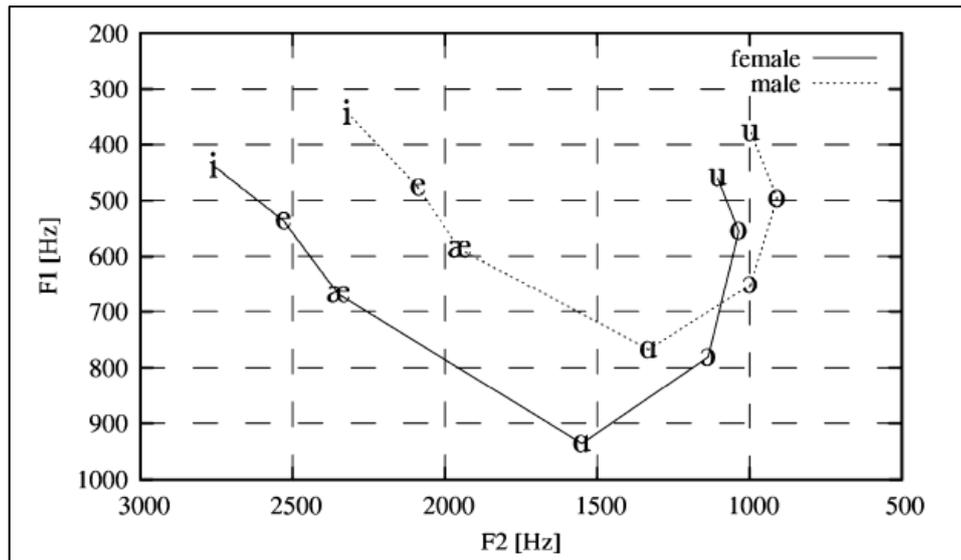


Figure 8 – Triangle vocalique moyen pour les femmes (trait continu) et les hommes (en pointillés) locuteurs de l'anglais américain. L'axe vertical indique la fréquence (en Hz) du premier formant (F1), et l'axe horizontal celle du deuxième formant (F2). (Simpson, 2009, d'après les données recueillies par Hillenbrand et al., 1995)

Dans toutes des études précédemment citées, les valeurs des formants sont globalement plus élevées chez les femmes. Selon Peterson et Barney (1952), cette différence serait en moyenne de l'ordre de 20 %. Néanmoins, comme l'illustre la figure 8, ce phénomène est souvent plus accentué pour les voyelles antérieures, en particulier sur le [i] (voir notamment Pépiot, 2009), et pour les voyelles ouvertes (voir Simpson & Ericsdotter, 2007), le tout se traduisant par un triangle vocalique plus étendu chez les locuteurs féminins (Traunmüller, 1988). Notons également que ces différences formantiques hommes / femmes semblent être accentuées en parole chuchotée, tout du moins pour les voyelles [i] et [a] (voir Vercherand, 2010). Par ailleurs, comme nous allons le voir, ces tendances pourraient ne pas être universelles.

Dans une étude pionnière, Bladon et al. (1984) se sont trouvés face à l'impossibilité d'établir un coefficient multiplicateur universel qui permettrait de passer d'une structure formantique d'homme à une structure formantique de femme (ou inversement) et ont ainsi mis en évidence l'influence du facteur langue sur les différences formantiques entre les voix d'hommes et les voix de femmes : ces différences varieraient considérablement en fonction de la langue parlée par les locuteurs.

Cette même constatation a été faite par Henton (1995), dans une étude comparative menée sur le français, le suédois, le hollandais et trois variétés de la langue anglaise. Un article de Keith Johnson, paru en 2005, va également dans ce sens. Cet auteur a collecté dans la littérature des relevés formantiques (F1, F2 et F3) de voyelles prononcées par des femmes et des hommes locuteurs de 26 langues et dialectes. Une synthèse des données recueillies par l'auteur est visible sous forme de représentation graphique dans la figure 9, présentée ci-après.

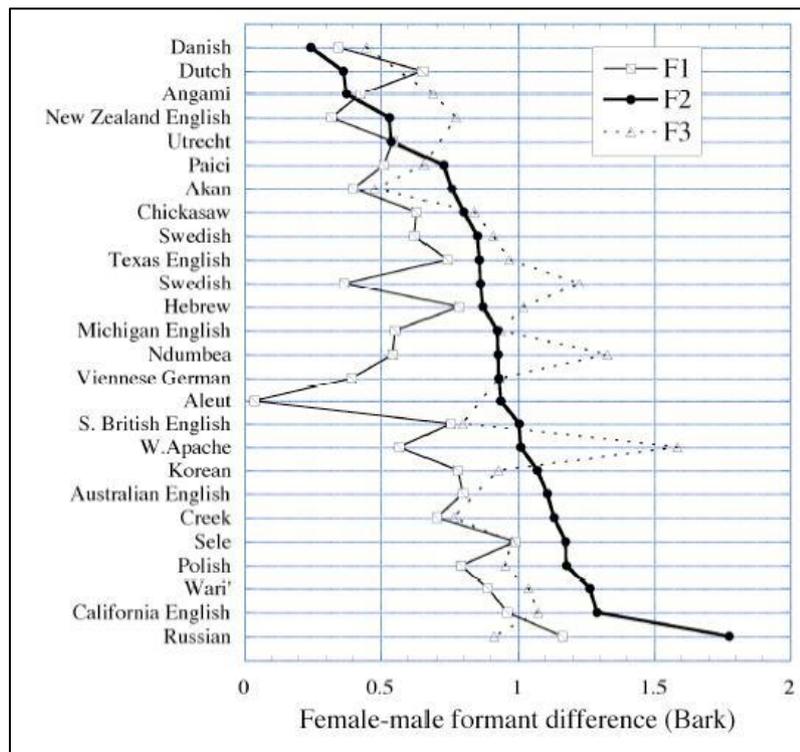


Figure 9 – Différences entre les valeurs moyennes de F1, F2 et F3 chez les femmes et chez les hommes dans 26 langues, exprimées en bark. (Johnson, 2005)

Comme l'illustre parfaitement cette figure, il en est ressorti une très grande disparité interlinguistique :

« [...] men and women have quite similar vowel formants in some languages like Danish with less than half of a Bark difference between men and women for F1, F2 and F3, while other languages show formant frequency differences that are more than double this, like Russian. » (Johnson, 2005 : 380)

Les formants vocaliques auraient donc tendance à être globalement plus élevés chez les femmes, mais d'importantes variations existent entre les langues et entre les différentes voyelles.

Des différences qui ne peuvent s'expliquer uniquement par la physiologie

Ces différences acoustiques seraient dues, pour le moins partiellement, à une différence physiologique : la longueur du conduit vocal, c'est-à-dire de la zone allant des plis vocaux jusqu'aux lèvres en passant par le pharynx, est plus importante chez les hommes que chez les femmes (voir Fant, 1966). Selon Simpson (2009), la longueur moyenne du conduit vocal des femmes adultes avoisinerait les 14,5 cm, contre environ 17,5 cm pour les hommes. Cet écart trouve sa principale origine dans l'abaissement du larynx qui survient à la puberté chez les sujets masculins, et qui entraîne un allongement du pharynx. Keith Johnson précise cependant, dans un article de 2005, que la taille et la forme du conduit vocal varie considérablement d'un sujet à l'autre, y compris du même sexe.

La taille du conduit vocal ayant une incidence directe sur les fréquences de résonance, cet élément physiologique permet d'expliquer en partie les différences de structures formantiques entre hommes et femmes. En effet, 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 locuteurs féminins (voir Turner et al., 2009 ; De Boer, 2010).

Mais cela n'explique pas pourquoi les écarts hommes-femmes varient très nettement d'une voyelle à une autre, comme nous l'avons vu précédemment, et ce 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.

De nombreux auteurs ont travaillé sur cette question et ont apporté plusieurs éléments de réponse. Fant (1966, 1975) et Nordström (1977) suggèrent que la non-uniformité de ces différences aurait elle-aussi une origine physiologique. Les différences de taille du conduit vocal entre les femmes et les hommes se concentrent principalement au niveau des cavités pharyngale et laryngale (voir notamment Chiba & Kajiyama, 1941

et Goldstein, 1980) : cela expliquerait pourquoi les différences formantiques inter-genres sont généralement plus fortes sur le F1 des voyelles ouvertes et sur le F2 des voyelles antérieures fermées non-arrondies, qui dépendent essentiellement de la longueur de la partie postérieure du conduit vocal, que sur les formants des voyelles arrières fermées, telles que le [u] et le [o], comme l'illustre la figure 10.

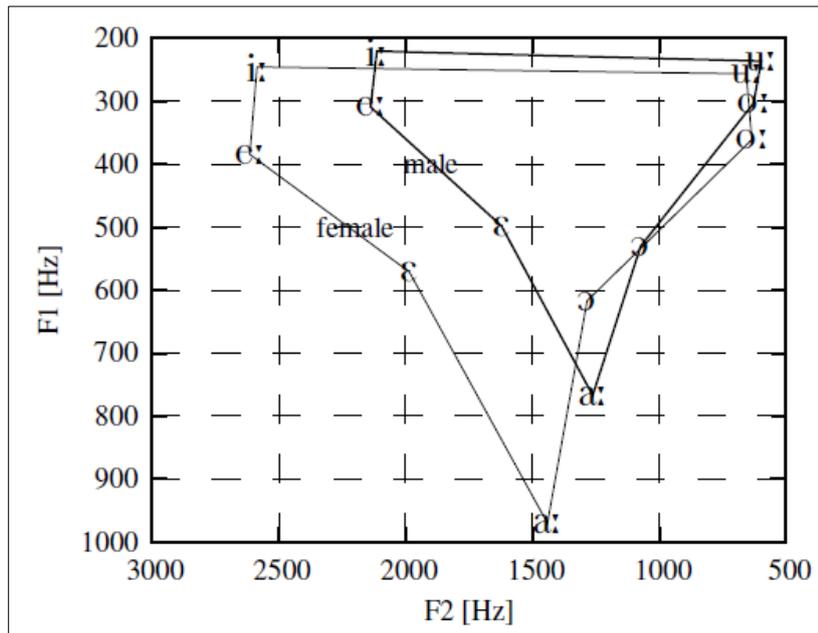


Figure 10 – Triangles vocaliques de l'allemand pour des locuteurs masculins (trait foncé) et féminins (trait clair) réalisés à partir de données recueillies sur 27 femmes et 14 hommes. (Simpson & Ericsson, 2007)

Autre élément d'explication avancé par Fant (1966, 1975), les faibles différences inter-genres au niveau des formants des voyelles [o] et [u] pourraient s'expliquer par l'utilisation de gestes compensatoires par les locutrices pour s'approcher le plus possible des valeurs formantiques des hommes : il s'agirait en particulier d'une protrusion plus importante des lèvres et d'un plus fort rapprochement entre le dos de la langue et le palais mou, permettant ainsi d'abaisser les deux premiers formants. Ces gestes compensatoires sont efficaces pour les F1 et F2 des voyelles arrières fermées, car ces derniers sont dus à une résonance de type Helmholtz, ce qui n'est pas le cas, par exemple, des formants de la voyelle ouverte [a] ou des deuxième et troisième formants du [i] (voir Fant, 1970).

Pour expliquer cette non-uniformité, Traummüller (1984) évoque quant à lui la possibilité selon laquelle les hommes continueraient à l'âge adulte à utiliser les mêmes commandes motrices qu'ils utilisaient avant leur puberté, lorsque leur larynx et leur

pharynx étaient de taille inférieure. Cette hypothèse, bien qu'intéressante, n'a pour l'instant pas été vérifiée ni reprise par d'autres auteurs.

Une autre possibilité a été explorée très largement par Simpson (2000, 2001, 2002, 2009). Comme l'illustre la figure 11, le conduit vocal des femmes étant plus petit que celui des hommes, la distance à parcourir pour passer d'un [a] à un [i] est plus faible pour les locutrices (de l'ordre de 11 %).

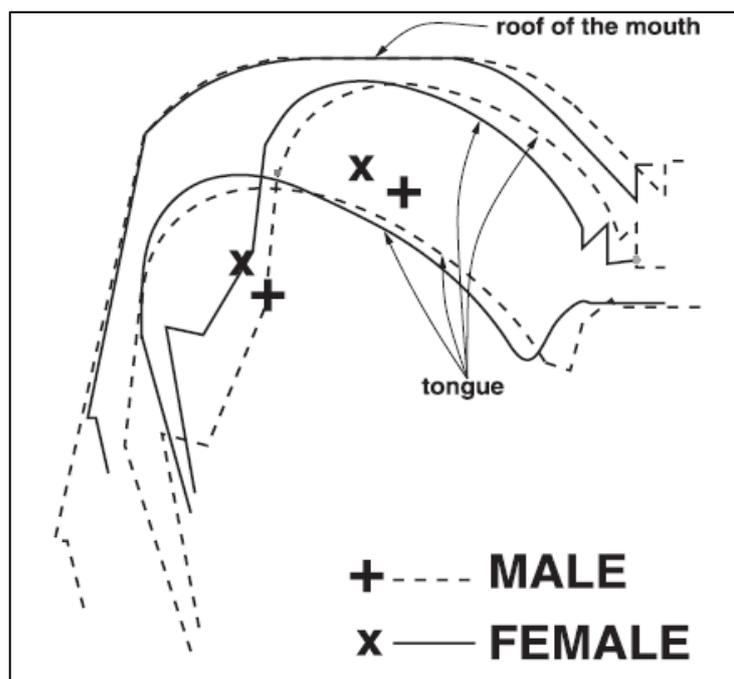


Figure 11 – Représentation schématique de la position de la langue dans le conduit vocal lors de l'articulation d'un [a] (en bas à gauche) et d'un [i] (en haut à droite) par un homme (en pointillés) et une femme (trait continu). Les symboles « + » (homme) et « x » (femme) indiquent un point central de la langue pour les deux voyelles. (Simpson, 2009)

Ainsi, selon l'auteur :

« [...] if males and females were to produce their utterances in approximately the same time frame, that is, the vowels and consonants have the same duration, a female moving her articulators at the same speed as a male should be able to reach more extreme articulatory and, of course, acoustic targets. From this, we could predict that the female acoustic vowel space might be larger because females can reach more extreme vowel positions in the same time. In particular, this would predict the relatively more open position of female open vowels that has been repeatedly observed in languages. » (Simpson 2009 : 628-629)

L'espace vocalique plus étendu chez les femmes trouverait donc lui aussi son origine dans la taille du conduit vocal. Ce dernier étant plus court et plus étroit chez les femmes, il laisserait la possibilité aux locutrices d'atteindre des cibles articulatoires plus extrêmes que les hommes dans un même laps de temps, du fait de la moindre distance à parcourir.

Et pourtant, même si cette hypothèse paraît séduisante de prime abord, l'auteur n'a pas été à même de la vérifier expérimentalement. Dans deux articles publiés en 2001 et 2002, Simpson a examiné de manière détaillée les différences hommes-femmes dans la production de la diphtongue de « *light* », et de la séquence vocalique de « *they all* », sur le plan articulatoire et acoustique. Cette étude portait sur 26 femmes et 22 hommes, locuteurs de l'anglais américain. D'un point de vue articulatoire, il en est ressorti que la langue des locuteurs masculins parcourait en moyenne une plus grande distance, avec une vitesse d'articulation supérieure aux femmes. Malgré cela, les variations acoustiques (c'est-à-dire l'évolution de la position des formants) étaient plus importantes chez les femmes. Ces résultats contredisent donc largement l'hypothèse initiale de Simpson, il convient néanmoins de rappeler que l'étude a été menée sur un nombre assez restreint de locuteurs, et avec un matériau linguistique très réduit.

Pour plusieurs auteurs (Goldstein 1980 ; Ryalls & Lieberman 1982 ; Maurer et al., 1992 ; Diehl et al. 1996 ; Simpson & Ericsson, 2007 ; Ditcheva & Dommergues, 2008), l'existence d'un espace vocalique plus important chez les femmes pourrait s'expliquer, tout du moins partiellement, par la fréquence fondamentale plus élevée des voix de femmes. Plus cette fréquence est élevée, moins il y a d'harmoniques : une voix d'homme à 100 Hz générera des harmoniques à 200 Hz, 300 Hz, 400 Hz et ainsi de suite, alors qu'une voix de femmes à 200 Hz en générera deux fois moins (400 Hz, 600 Hz etc.). Les formants qui, rappelons-le, sont des zones d'harmoniques renforcés, sont nécessairement moins bien définis lorsqu'il y a peu d'harmoniques. Selon ces auteurs, les voyelles seraient alors plus difficilement identifiables, et les locuteurs féminins tenteraient donc de compenser ce phénomène en espaçant plus fortement leurs formants, rendant ainsi leurs productions vocaliques plus clairement identifiables.

Pour tester cette hypothèse, Diehl et al. (1996) ont synthétisé des voyelles /ɪ/ et /ʊ/ avec différentes fréquences fondamentales, allant de 90 à 360 Hz, mais en gardant la structure formantique constante (voix d'homme ou voix de femme). Ces stimuli ont été

présentés à des auditeurs ayant pour tâche d'identifier la voyelle. Les résultats ont été concluants : toutes choses égales par ailleurs, à partir de 150 Hz, plus la fréquence fondamentale est élevée et plus l'identification des voyelles est difficile pour les auditeurs²⁶. Diehl et al. (1996) en concluent que :

« [...] the findings suggest that the greater between-category dispersion of female vowels may be plausibly explained as a means of offsetting the deleterious effects on vowel identifiability of (typically) higher F0s. » (p.205)

Afin de tester s'il existe réellement une corrélation entre fréquence fondamentale moyenne et taille de l'espace vocalique, Simpson et Ericsson (2007) ont examiné les productions vocaliques d'un groupe d'hommes (n=17) et d'un groupe de femmes (n=70) locuteurs germanophones, partant du postulat que si cette corrélation existe, elle devrait alors se trouver également à l'intérieur même de ces deux groupes : par exemple, les hommes ayant un F0 relativement élevé devraient eux aussi utiliser un espace vocalique plus important que ceux ayant un F0 bas. Les résultats de cette étude sont mitigés : une faible corrélation a été trouvée, et seulement pour une partie des données, la plus forte corrélation se trouvant au sein du groupe des femmes.

Cela pourrait confirmer l'idée selon laquelle la compensation du manque d'harmoniques par l'élargissement de l'espace vocalique serait nécessaire uniquement au dessus d'un certain seuil (aux alentours de 150 Hz, voir Diehl et al., 1996), et ne concernerait donc que très peu de locuteurs masculins. Néanmoins, les auteurs n'ont pas été en mesure de confirmer entièrement et formellement leur hypothèse de départ, et concluent qu'il conviendrait de renouveler ce même type d'investigation mais avec une méthodologie plus stricte et un plus grand nombre de locuteurs masculins. A une échelle plus réduite, une recherche menée par Cleveland (1976, 1977) sur huit chanteurs professionnels locuteurs du suédois semble appuyer cette hypothèse : les chanteurs basses (F0 bas) et ténors (F0 très élevé) présentaient des différences formantiques très similaires à celles existant entre hommes et femmes.

Revenons maintenant sur les disparités inter-langues évoquées précédemment (voir Johnson, 2005, 2006 ; Henton, 1995 ; Bladon et al., 1984). Plusieurs chercheurs ont

²⁶ Voir également la section 4.3.2.

tenté d'apporter des explications sur ces variations. Johnson (2005) a confronté les différences inter-langues observées dans son étude (voir figure 9) au modèle de prédiction des différences formantiques inter-genres de Traunmüller (1984), modèle basé uniquement sur les différences physiologiques standards entre les conduits vocaux des hommes et des femmes. Les résultats ont été les suivants : « *Overall only 30 % of the variance is accounted for by the vocal tract differences model* » (p.381). Dans une étude plus récente (Johnson, 2006) ce même auteur a cherché s'il existait une corrélation entre la taille moyenne des hommes et des femmes²⁷ selon leur pays d'origine²⁸ et les différences formantiques inter-langues : seule une très faible corrélation a été trouvée, pour le troisième formant. De plus, des études similaires, mais menées sur des individus du même genre (hommes ou femmes) n'ont jamais pu mettre en évidence une forte corrélation entre taille moyenne du locuteur et position des formants (voir Greisbach, 1999 et González, 2004).

Selon Johnson (2005, 2006), ces variations inter-langues constituent donc la preuve que les différences formantiques entre hommes et femmes ne peuvent pas s'expliquer exclusivement par la taille du conduit vocal : des stratégies articulatoires entrent nécessairement en jeu. Il existerait donc une forte influence culturelle sur les conduites articulatoires, c'est d'ailleurs à cette même conclusion qu'arrivent Mattingly (1966), Kahn (1974) et Simpson (2009) :

« *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 : 632)

Cette hypothèse tend à être confirmée par les études menées sur des locuteurs non pubères, garçons et filles. En effet, si les différences formantiques hommes-femmes s'expliquaient uniquement par la taille du conduit vocal, aucune différence ne devrait exister avant la puberté, moment où survient l'abaissement du larynx chez les hommes. Et pourtant, des différences formantiques ont bel et bien été observées entre des locuteurs des deux genres avant même la puberté, et malgré des conduits vocaux de taille

²⁷ La taille étant elle-même très fortement corrélée à la taille du conduit vocal (Fitch et Giedd, 1999 ; Cherng et al., 2002).

²⁸ L'auteur a utilisé les données de Tolonen et al., 2000.

équivalente²⁹. Busby et Plant (1995), ont analysé les productions langagières de 20 garçons et 20 filles locuteurs de l'anglais australien, âgés de 5, 7, 9 ou 11 ans. Les valeurs des deux premiers formants vocaliques étaient significativement plus élevées pour les locutrices. Dans une étude plus large, menée sur les productions de 436 enfants et adolescents américains, âgés de 5 à 18 ans³⁰, Lee et al. (1999) arrivent aux mêmes conclusions : les valeurs de F1 et F2 sont significativement plus élevées chez les filles que chez les garçons, dès le plus jeune âge. Cette tendance serait particulièrement forte pour le premier formant :

« F1 values for young female and male children are significantly different, far larger than any that could be explained by the physiological differences at this age. » (p.1468)

Des constatations similaires ont été faites dans un article plus récent de Perry et al. (2001). Ces auteurs ont observé des différences garçons-filles significatives au niveau de F1, F2 et F3 chez des groupes d'enfants prépubères locuteurs de l'anglais américain, âgés de 4, 8 et 12 ans. Citons également une étude plus ancienne de Sachs et al. (1973), qui présente des conclusions identiques.

Les formants vocaliques plus élevés chez les femmes semblent donc relever en grande partie d'une conduite sociale, qui pourrait être apprise avant même la puberté. Notons cependant que les trois études menées sur des enfants évoquées plus haut impliquent uniquement des locuteurs anglophones.

3.4 Zones de bruit des fricatives sourdes

Autre élément lié au domaine fréquentiel, la position des zones de bruit des consonnes fricatives sourdes a fait l'objet de quelques recherches sur le plan des différences hommes-femmes.

²⁹ L'absence de différence entre les enfants des deux sexes sur le plan de la morphologie du conduit vocal a été démontrée par Fitch et Giedd (1999), grâce à l'utilisation de l'imagerie par résonance magnétique (IRM) sur 129 sujets, dont 66 enfants âgés de 2 à 10 ans.

³⁰ Les enregistrements analysés par les auteurs sont tirés de la base de données créée par Miller et al. (1996).

Dans une étude pionnière datée de 1968, Schwartz a comparé la réalisation des fricatives sourdes /ʃ/, /s/, /f/ et /θ/ par des femmes et par des hommes, locuteurs de l'anglais américain. Peu de différences acoustiques ont été trouvées pour les fricatives inter-dentales et labio-dentales. En revanche les réalisations de /s/ et /ʃ/ ont bien varié significativement en fonction du genre du locuteur, le bruit se concentrant dans des fréquences plus hautes chez les femmes que chez les hommes. Des résultats similaires ont été obtenus par Shadle et al. (1991), Nittrouer (1995), Whiteside (1998b) et Jongman et al. (2000). Cette tendance est visible ci-dessous sur la figure 12, où l'on peut observer, pour la consonne /s/ produite par des anglophones, que la concentration d'énergie se trouve dans des fréquences globalement plus élevées chez la locutrice. Ces spectres ont été obtenus à partir de données recueillies par Pépiot (2010).

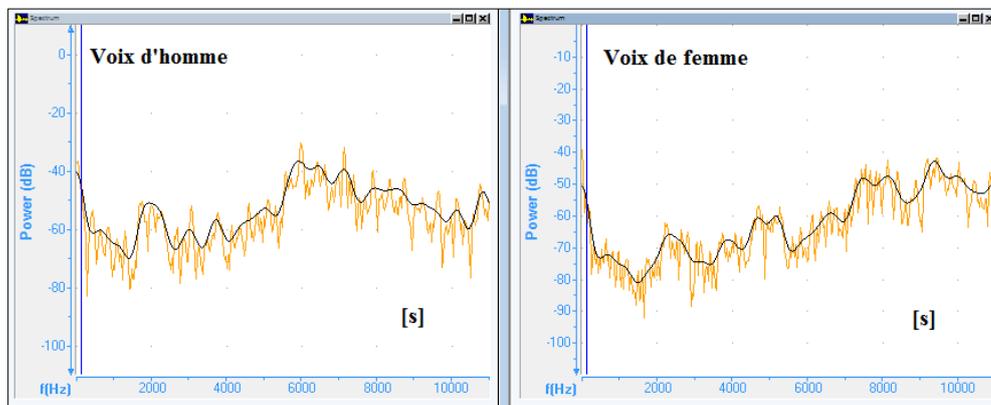


Figure 12 – Spectres de la fricative [s] du mot « Steven », prononcée par un homme (à gauche) et une femme (à droite) anglophones obtenus avec le logiciel Speech Analyzer³¹. L'axe vertical représente l'intensité, l'axe horizontal les fréquences.

Plus récemment, une importante étude a été réalisée sur le même sujet par Fox et Nissen (2005). Les auteurs ont enregistré 50 hommes et 50 femmes locuteurs de l'anglais américain, équitablement répartis en cinq classes d'âge : 6-7 ans, 8-9 ans, 10-12 ans, 13-14 ans et adultes. L'analyse acoustique détaillée des fricatives sourdes produites par les locuteurs a permis d'obtenir plusieurs résultats particulièrement intéressants. Tout d'abord, l'énergie spectrale était située en moyenne dans des fréquences significativement plus élevées chez les femmes pour les réalisations de /s/ et /ʃ/, mais aucune différence significative n'est apparue pour /f/ et /θ/. La même tendance a été trouvée pour le pic

³¹ Speech Analyzer version 3.0.1.

spectral (*spectral peak*), c'est-à-dire l'endroit où se trouve la plus forte concentration d'énergie. Ces derniers résultats sont présentés sur la figure 13 ci-dessous.

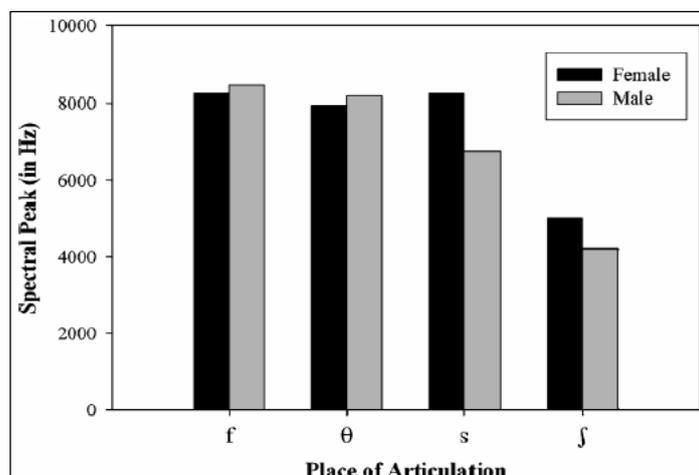


Figure 13 – Pic spectral moyen (exprimé en Hz) pour chacune des fricatives sourdes de l'anglais lorsque produite par des hommes (en gris) et des femmes (en noir). (Fox & Nissen, 2005)

Notons également que les différences femmes-hommes sont plus élevées pour le /s/ que pour le /ʃ/. D'autre part, si l'on s'intéresse aux classes d'âge des locuteurs, on s'aperçoit que ces différences sont déjà présentes de manière significative chez les locuteurs âgés de 6 à 9 ans, bien que moins importantes que chez les adolescents et adultes. Fox et Nissen sont ainsi les premiers auteurs à démontrer que les différences hommes-femmes pour les fricatives sourdes sont déjà présentes, de manière atténuée, avant même la puberté.

Pourquoi de telles différences ?

Ces différences pourraient, comme pour la structure formantique, trouver leur origine dans la taille du conduit vocal : celui des femmes étant plus court et plus étroit que celui des hommes, les fréquences de résonance devraient être globalement plus élevées, y compris avec des sons apériodiques (voir McGowan & Nittrouer, 1988). Cette explication est parfaitement concordante avec le fait que ces différences hommes-femmes soient présentes sur /s/ et /ʃ/ mais pas sur /f/ et /θ/, consonnes articulées à l'extrémité du conduit vocal et donc moins affectées par la taille de ce dernier. C'est également à cette conclusion qu'arrivent Fox et Nissen (2005). Mais selon eux, cette explication n'est pas suffisante pour expliquer à elle-seule ces différences hommes-femmes qui, comme le

montre leur étude, semblent émerger avant même la puberté, alors que les conduits vocaux des filles et des garçons ne présentent pas encore de différence morphologique³² :

«*Considering recent magnetic resonance imaging data suggesting that sexual vocal tract dimorphism for most children occurs at the peripubertal and postpubertal stages of development, it is likely that the sex-related acoustic differences found in this study cannot be entirely explained by anatomical differences alone. It may be concluded that a portion of these sex-linked differences may be associated with learned or behavioral factors.* » (pp. 764-765)

Cette hypothèse concorde avec les résultats obtenus par Heffernan (2004). Cet auteur a conduit une analyse acoustique sur des réalisations de la fricative /s/ par des locuteurs et locutrices anglophones américains et japonophones³³ : des différences inter-genres mais également inter-langues sont apparues. La différence femmes-hommes est en effet significativement plus grande chez les anglophones que chez les japonophones, ce qui permet de conclure à l'existence d'un facteur culturel.

A l'instar des paramètres acoustiques investigués dans les sous-sections précédentes, ces différences inter-genres pourraient là encore être à la fois d'origine physiologique et socio-linguistique. Notons également que toutes les études évoquées ici, à l'exception de Heffernan (2004), ont été menées sur des locuteurs anglophones : il est donc difficile de conclure à l'universalité des phénomènes observés.

3.5 Qualité de voix

Des différences existeraient également entre voix d'hommes et voix de femmes sur le plan du *type de phonation* (ou *qualité de voix*³⁴), c'est-à-dire sur la manière dont les plis vocaux vibrent lors de la production des sons voisés. De nombreuses études présentent le *souffle* (de l'anglais *breathiness*) comme un trait différenciant les voix de

³² Voir Fitch & Giedd, 1999.

³³ Notons que, comme l'indique à juste titre l'auteur, la réalisation articulatoire du phonème /s/ n'est pas identique dans les deux langues.

³⁴ Pour rappel, j'utiliserai ici indifféremment les termes *type de phonation* et *qualité de voix* pour désigner les voix *breathy*, *creaky* et modale. Voir la section 2.1 du présent chapitre.

femmes de celles des hommes. Une voix *breathy*³⁵ se caractérise par un type de phonation dans lequel la fermeture des plis vocaux est incomplète lors de chaque cycle et l'espace entre les plis vocaux lors de leur phase ouverte est plus important que dans un type de phonation dit normal (voir Fritzell et al., 1986, et Jeannon et al., 1998). Le tout entraînant ainsi le passage d'un flux d'air plus important qu'avec une qualité de voix dite normale³⁶, comme l'illustre la figure 14 ci-après.

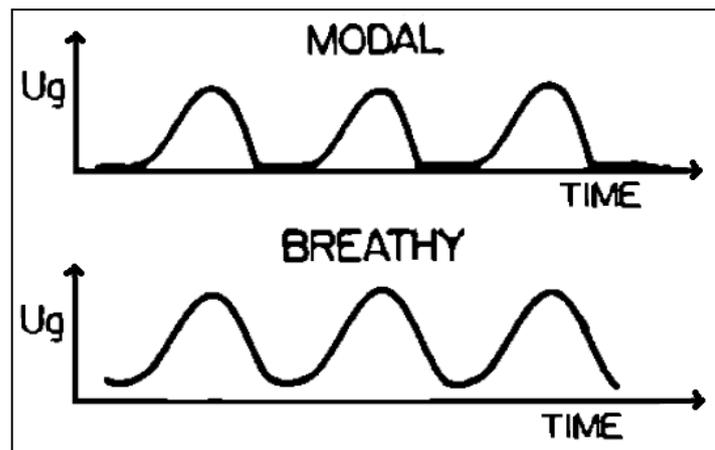


Figure 14 – Représentation des cycles d'ouverture-fermeture des plis vocaux avec une voix modale (en haut) et avec une voix soufflée (en bas). Le temps figure sur l'axe horizontal, le flux d'air glottal (U_g) sur l'axe vertical. (D'après Klatt & Klatt, 1990)

Acoustiquement, cette qualité de voix se traduit par la présence d'un bruit d'aspiration au niveau du troisième formant, et surtout par une forte amplitude du premier harmonique³⁷ par rapport au deuxième (voir notamment Klatt & Klatt, 1990 ; Shrivastav & Sapienza, 2003). Cette différence H1-H2 (exprimée en dB) est en effet fortement corrélée (positivement) au quotient d'ouverture glottique (ou *glottal open quotient*), c'est-à-dire au rapport de la durée d'ouverture des plis vocaux sur une période laryngée donnée (Henrich et al., 1999). La valeur de ce quotient est donc comprise entre 0 et 1 : plus elle est élevée, plus la voix sera perçue comme « soufflée » (voir Gordon & Ladefoged, 2001).

Les voix de femmes sont le plus souvent présentées comme plus *breathy* que les voix d'hommes. Citons par exemple Klatt (1986, 1987a) et Klatt & Klatt (1990), qui comptent parmi les premières études à avoir mis ce phénomène en évidence chez les

³⁵ Voix « soufflée » ou « aspirée » dans la terminologie française.

³⁶ Plus généralement appelée voix *modale*.

³⁷ C'est-à-dire de F0.

locuteurs de l'anglais américain, mais aussi Henton & Bladon (1985), sur l'anglais britannique, Hanson & Chuang (1999), chez des locuteurs anglophones américains, et enfin Henton (1992) qui présente le souffle comme un trait propre aux voix de femmes chez les anglophones, toutes variétés d'anglais confondues.

Ce phénomène se retrouverait également chez les locutrices d'autres langues, comme le suédois (voir Karlsson, 1981, 1982 ; Södersten & Lindestad, 1990 et Södersten et al., 1995). Il en va de même pour l'espagnol : dans un article publié en 1996, Mendoza et al. ont mené une étude acoustique comparative entre des voix d'hommes et des voix de femmes hispanophones. Selon ces auteurs :

« *The female voice showed greater levels of aspiration noise, located in the spectral regions corresponding to the third formant, which causes the female voice to have a more "breathy" quality than the male voice.* » (p.59)

Ces observations sont confirmées par Trittin & De Santos y Lleo (1995), dans une autre étude acoustique réalisée sur des productions de locuteurs et de locutrices espagnols. Cependant, selon ces derniers, la différence serait moins importante chez les locuteurs hispanophones que chez les locuteurs anglophones :

« *To the extent that relative first-harmonic strength is an indication of the aspiration noise found in the vowel spectrum, our results seem to indicate that Spanish females are slightly more breathy than Spanish males. The difference is smaller than the results given by Klatt and Klatt, but statistically significant.* » (p.367)

Certains locuteurs masculins auraient quant à eux tendance à utiliser la voix « craquée » (de l'anglais *creaky*) de manière récurrente, contrairement aux locutrices. Ce type de phonation *creaky* se caractérise sur le plan articulatoire par une vibration irrégulière des plis vocaux (Gordon & Ladefoged, 2001), généralement entre 20 et 60 fois par seconde, avec lors de chaque cycle une ouverture des plis vocaux incomplète et très brève (Klatt & Klatt, 1990). Le flux d'air transitant entre les plis vocaux étant par conséquent plus faible qu'avec une qualité de voix modale, à l'inverse de la voix *breathy*. Au niveau acoustique, l'amplitude relative du premier harmonique (i.e. F0) se trouve sensiblement affaiblie (voir Klatt & Klatt, 1990), en raison d'un quotient d'ouverture glottique peu élevé.

Plusieurs études, portant sur des locuteurs de différentes variétés de l'anglais, ont mis en évidence une tendance à l'utilisation de la voix *creaky* (ou *fry*) par les locuteurs masculins : citons notamment Henton (1989b, 1992) et Henton & Bladon (1986, 1988). Selon Henton et Bladon (1988), ce type de phonation serait principalement utilisé sur la partie finale des énoncés. Dans une étude menée sur des locuteurs et locutrices du suédois, Huber (1989) remarque également une utilisation plus fréquente de la voix *creaky* chez les hommes que chez les femmes, particulièrement en fin d'énoncé. Hammarberg et al. (1988) avancent eux aussi que les locuteurs du suédois auraient plus souvent recours à ce type de phonation que les locutrices.

Le type de phonation comme marqueur sociolinguistique

L'utilisation plus fréquente de la voix *creaky* par les locuteurs masculins ne semble pas trouver son origine, ne serait-ce que partiellement, dans la physiologie des plis vocaux. Les auteurs ayant étudié ce phénomène le considèrent quasi-unaniment comme un marqueur sociolinguistique, issu de la construction sociale (voir notamment Simpson, 2009, Henton & Bladon, 1988). Ayant étudié ce phénomène dans plusieurs langues, Caroline Henton va même plus loin :

« [...] *creaky voice may be employed as a sociophonetic marker across cultures. Indeed, its role as a marker of male speech may be enhanced in certain cultures where gender-role expectations are more pronounced.* » (Henton, 1989b : S26)

Ainsi, l'utilisation de la voix craquée serait un marqueur phonétique de l'identité de genre masculin et ce phénomène pourrait, selon Henton, être plus présent dans les cultures où les rôles hommes / femmes sont les plus différenciés. Notons cependant qu'une récente publication (Wolk et al., 2012) met en évidence l'émergence de l'utilisation de la voix *creaky* chez les jeunes locutrices de l'anglais américain, avec une étude menée sur 34 étudiantes de l'Etat de New-York.

Revenons-en maintenant à la voix *breathy*, qui serait comme nous l'avons vu précédemment un trait propre aux voix de femmes. Certains auteurs (Titze, 1989 ; Simpson, 2009) suggèrent qu'une différence physiologique au niveau des plis vocaux pourrait en être l'une des causes. Comme nous l'avons vu précédemment, lors d'un cycle de vibration dit « normal » les plis vocaux ont une phase d'ouverture, puis une phase de fermeture, durant laquelle de passage de l'air est bloqué. En revanche, pour la voix

soufflée, les plis vocaux ne se ferment pas complètement (voir figure 14) et laissent donc passer un flot d'air continu entre les plis vocaux.

Dans une étude de 1989, Titze a élaboré un modèle mathématique simulant la vibration des plis vocaux chez les hommes et les femmes, en considérant leurs différences physiologiques. Selon ce modèle, les plis vocaux des femmes, plus fins que ceux des hommes, auraient tendance à ne pas se fermer complètement durant un cycle, ou tout du moins à rester fermés moins longtemps. Par conséquent, le flux d'air étant plus important, les voix de femmes seraient par nature plus *breathy* que celles des hommes. Le modèle de Titze semble globalement attesté par les données articulatoires recueillies par Södersten & Lindestad (1990, 1992).

Néanmoins, cette explication physiologique n'est pourtant pas partagée par tous les auteurs (voir notamment Wagner & Braun, 2003). Dans leur article de 1995 évoqué plus haut, Trittin et De Santos y Lleo avancent l'hypothèse selon laquelle l'utilisation de la voix *breathy* par les femmes serait pour partie une conduite acquise, construite socialement :

“These findings seem to support the suspicion that a breathy voice quality may be a learned behaviour.” (p.360)

Pour affirmer ce point de vue, les auteurs s'appuient sur les variations observées à travers les langues : les locutrices anglophones semblent avoir globalement des voix plus *breathy* (Klatt & Klatt, 1990) que les locutrices hispanophones (Trittin & De Santos y Lleo, 1995) ou suédophones (Karlsson, 1991, 1992).

3.6 Voice onset time

Concernant les différences hommes-femmes dans le domaine temporel, plusieurs auteurs se sont penchés sur le délai d'établissement du voisement (DEV). Plus communément désigné par son équivalent anglo-saxon *voice onset time* (et son abréviation *VOT*), cette mesure acoustique correspond au 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. Sa valeur, exprimée généralement en millisecondes, peut être positive si les plis vocaux entrent en vibration après le relâchement (comme pour les occlusives sourdes en anglais et en français), ou négative si les plis vocaux débutent leurs

vibrations avant même le relâchement de la consonne (occlusives voisées du français, par exemple).

Chez les locuteurs anglophones, et bien que les résultats varient sensiblement d'une étude à l'autre, la tendance générale irait vers un temps de voisement plus long chez les femmes³⁸. A titre d'exemple, Swartz (1992) a mis en évidence des temps de voisement significativement plus courts chez les hommes que chez les femmes anglophones, sur la paire de consonnes occlusives alvéolaires [t] / [d].

Des résultats similaires ont été obtenus à plusieurs reprises sur les occlusives sourdes [p], [t] et [k] (voir Whiteside & Irving, 1997, Ryalls et al., 1997, Koenig, 2000 et Robb et al., 2005), toujours sur des locuteurs anglophones. Ce phénomène est illustré par la figure 15 ci-après, sur laquelle on peut observer le temps de voisement d'un homme et d'une femme locuteurs de l'anglais américain après le relâchement d'une occlusive sourde [t]. Les spectrogrammes ont été obtenus à partir d'enregistrements effectués par Pépiot (2010).

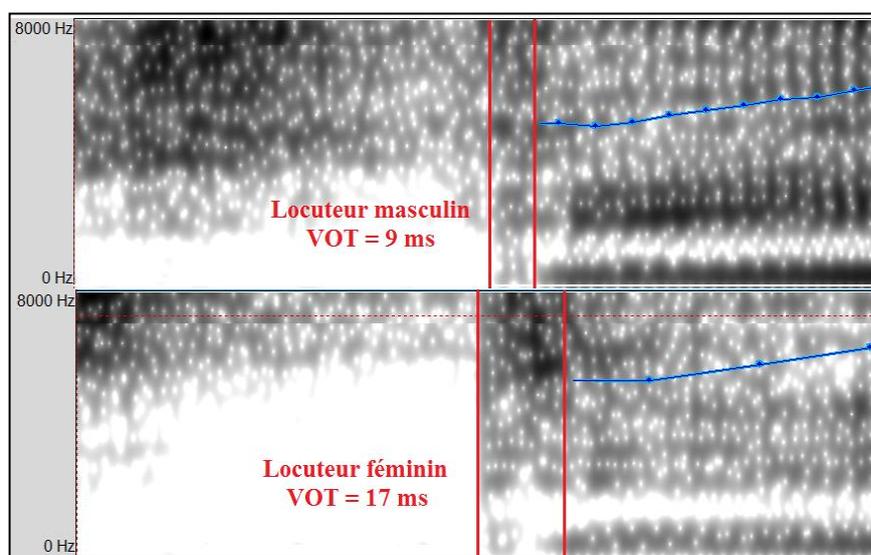


Figure 15 – Spectrogrammes de la séquence [sti:] dans le mot « Steven », prononcée par un homme (en haut) et une femme (en bas) anglophones, obtenus avec le logiciel Praat³⁹. Les deux cadres rouges représentent les VOT des locuteurs après le relâchement du [t], la courbe bleue est celle du F0.

³⁸ A l'exception notable de Sweeting & Baken (1982) et Morris et al. (2008), qui ne trouvent aucune différence significative entre hommes et femmes, et de Smith (1978) pour qui le VOT des hommes serait sensiblement plus long que celui des femmes pour [d] et [g].

³⁹ Voir Boersma & Weenink (2010)

Plusieurs études portant sur le *VOT* et le genre ont été menées sur des enfants et préadolescents. Les résultats de ces études sont particulièrement intéressants. Whiteside et Marshall, dans un article publié en 2001, ont mesuré les *VOT* de trois groupes de jeunes locuteurs de l'anglais britannique (7 ans, 9 ans et 11 ans), composés chacun de cinq garçons et cinq filles, pour les occlusives [p], [b], [t] et [d]. Les auteurs ont pu mettre en évidence une diminution significative du *VOT* avec l'âge chez les garçons pour les consonnes [p] et [b], et une tendance à l'augmentation du *VOT* avec l'âge chez les jeunes filles pour la consonne [t].

Dans une étude ultérieure de Whiteside et al. (2004b), les classes d'âge étudiées ont été étendues avec l'ajout d'un groupe composé de locuteurs et locutrices âgés de 5 ans, et d'un autre groupe d'adolescent(e)s de 13 ans. Le nombre de consonnes étudiées a lui aussi été augmenté, avec l'ajout des occlusives vélares [k] et [g]. Globalement, les *VOT* des filles étaient significativement plus élevés que ceux des jeunes garçons. De plus, ces différences sont maximales pour le groupe le plus âgé (13 ans), laissant supposer que ce phénomène émerge progressivement lors de la préadolescence. Enfin, les différences garçons / filles observées sont plus importantes sur les occlusives sourdes ([p], [t], et [k]) que sur les occlusives voisées.

Il convient néanmoins de ne surtout pas tirer de conclusions hâtives quant à l'universalité du phénomène décrit ci-dessus. En effet, l'ensemble des recherches mentionnées précédemment ont été menées sur des locuteurs et locutrices anglophones. L'analyse de données complémentaires provenant d'autres langues est donc plus que nécessaire. Un récent article d'Oh (2011) montre d'ailleurs que chez les locuteurs du coréen, la tendance semble opposée à celle observée chez les anglophones : dans cette étude réalisée sur 19 femmes et 19 hommes, les *VOT* des locutrices sont, selon le type de consonne, soit significativement plus courts que ceux des hommes, soit de durée équivalente. De même, une étude de Karlsson et al. (2004) conduite sur des locuteurs suédophones ne fait état d'aucune différence significative entre les *VOT* des hommes et des femmes adultes. A ma connaissance, ce type d'étude semble ne pas encore avoir été réalisé sur des locuteurs francophones.

L'influence du cycle menstruel sur le VOT des locutrices

Selon plusieurs auteurs, les variations hormonales durant le cycle menstruel joueraient sur le *voice onset time* des locutrices. Whiteside et al. (2004a) ont mesuré les VOT de 7 femmes anglophones à deux moments distincts de leur cycle menstruel : entre le 2^{ème} et le 5^{ème} jour, moment où le taux d'estrogènes et de progestérone (taux d'EP) est au plus bas, et entre le 18^{ème} et le 25^{ème} jour, période durant laquelle le taux d'estrogènes et de progestérone est le plus élevé. Entre ces deux relevés, les auteurs ont observé une diminution significative du VOT pour les occlusives voisées ([b], [d] et [g]) et en même temps une augmentation significative du VOT pour les occlusives sourdes ([p], [t] et [k]). En d'autres termes, le contraste entre occlusives sourdes et voisées est renforcé chez les femmes au moment du cycle où le taux d'estrogènes et de progestérone est le plus élevé. D'autre part, une comparaison a également été effectuée avec les VOT de cinq locuteurs masculins : aucune différence significative n'a été trouvée avec celui des femmes pendant la période avec faible taux d'EP, en revanche le VOT des femmes était globalement plus élevé que celui des hommes pour les occlusives sourdes durant la période avec fort taux d'EP.

Cette étude a été reproduite par Wadnerkar et al. (2006) avec un plus grand nombre de locuteurs (15 femmes et 20 hommes) : toutes les tendances observées précédemment par Whiteside et al. ont été confirmées. Ces auteurs en concluent que :

« [...] it can be said that the results from the present study support the role of activational ovarian hormones in regulating one temporal component of speech—VOT, and provide further evidence for the flexibility, and sensitivity of the speech—motor system to the hormonal environment. Hence, it is necessary to consider the effects of menstrual cyclicity while investigating VOT within or between sexes. »
(p.23)

Ainsi, la quasi-totalité des études comparatives menées sur le VOT des hommes et des femmes auraient, à tort, négligé le facteur « cycle hormonal des femmes », qui semble jouer un rôle prépondérant sur la durée de leurs VOT.

Des différences d'origine physiologique mais sujettes à l'influence culturelle

Les auteurs ayant travaillé sur les différences de *VOT* entre hommes et femmes ont avancé plusieurs éléments d'explication quant à leur raison d'être. Robb et al. (2005), en s'appuyant sur une étude médicale menée sur la physiologie des plis vocaux (Bless & Abbs, 1983), avancent l'hypothèse selon laquelle la morphologie des plis vocaux des femmes serait responsable d'une pression sub-glottale plus importante que chez les hommes lors de la phrase de fermeture des plis vocaux, entraînant ainsi un démarrage du voisement plus tardif (c'est-à-dire un *VOT* supérieur). Koenig (2000) et Whiteside et al. (2004b) proposent une explication très proche : les plis vocaux des femmes seraient intrinsèquement plus tendus que ceux des hommes. Ainsi, les plis vocaux des hommes, plus relâchés, pourraient entrer en vibration plus rapidement que ceux des femmes (*VOT* plus court).

Il est à noter cependant que pour certains auteurs (voir Robb et al., 2005) l'explication physiologique seule ne permet pas d'expliquer l'intégralité des différences inter-genres observées sur le *voice onset time* : des facteurs socio-phonétiques pourraient également intervenir. Pour Oh (2011), ce facteur social pourrait même aller jusqu'à inverser la tendance induite par les différences physiologiques, sans quoi il ne serait pas possible d'expliquer pourquoi le *VOT* des locuteurs du coréens est globalement plus élevé que celui de leurs homologues féminins.

Concernant la variation du *VOT* des femmes avec le cycle menstruel, Wadnerkar et al. (2006) supposent que ce phénomène serait en partie lié au taux d'estrogènes, qui causerait des changements physiologiques au niveau du larynx (e.g. Amir & Biron-Shental, 2004 et Abitbol et al., 1999). Ces changements hormonaux pourraient aussi, toujours selon ces auteurs, affecter le système nerveux central et ainsi les commandes motrices entrant en jeu lors de la production de la parole. Ces explications n'ont pu être testées expérimentalement pour le moment.

3.7 Durée des voyelles

Toujours dans le domaine temporel, des chercheurs se sont intéressés à la durée des voyelles, du point de vue de la comparaison femmes-hommes. Des tendances significatives ont pu être mises en évidence pour de nombreuses langues.

Dans une étude publiée en 1998, Simpson a étudié les productions vocaliques de 29 hommes et 25 femmes germanophones dans le discours spontané et lors de la lecture d'un texte. Dans les deux cas, les voyelles produites par les locutrices ont été significativement plus longues que celles produites par les locuteurs. L'écart moyen sur l'ensemble des productions est d'environ 11 %. Dans une vaste étude acoustique des voyelles de l'anglais américain menée sur une centaine de locuteurs, Hillenbrand et al. (1995) ont eux aussi observé une durée supérieure pour les voyelles produites par les locuteurs féminins⁴⁰ et ce quelle que soit la voyelle. Une tendance similaire a été constatée par Ericsson et Ericsson (2001) sur des locuteurs suédophones (cinq hommes et cinq femmes). Enfin, en français québécois (Martin, 1995, 1998a, 1998b, 2001), des différences significatives ont été trouvées pour certaines voyelles antérieures, qui affichent une durée là encore supérieure chez les femmes.

Cependant, Martin n'a pu observer aucune différence significative sur les voyelles postérieures, et le [ø] semble au contraire être plus long lorsque produit par des hommes. Notons également que Jacewicz et al. (2007), dans une étude portant sur la durée des voyelles dans trois dialectes de l'anglais américain n'ont pas trouvé de différences significatives entre locuteurs féminins et masculins. Seule une *tendance* a été observée pour l'anglais américain de Caroline du Nord, dans lequel les voyelles produites par les femmes étaient très légèrement plus longues que celles produites par les hommes.

Un autre phénomène a pu être mis en évidence dans certaines langues, dont le système vocalique présente des oppositions entre voyelles longues et voyelles brèves : le contraste de durée entre ces deux types de voyelles serait plus important chez les locutrices que chez les locuteurs. Citons par exemple Wassink (1999), avec une étude menée sur 20 locuteurs du créole jamaïcain, ou encore Johnson et Martin (2001) et leur recherche portant sur les productions vocaliques de locuteurs du Creek (quatre hommes et

⁴⁰ Les voyelles produites par les locutrices étaient en moyennes 20 % plus longues que celles produites par les hommes.

quatre femmes), langue amérindienne. Ces auteurs sont arrivés aux mêmes conclusions : la différence de durée entre les voyelles brèves et les voyelles longues était significativement plus grande dans les productions des femmes que dans celles des hommes.

Cette tendance à l'amplification des contrastes de durée entre les voyelles a également été trouvée chez les locutrices du suédois, mais cette fois-ci sur le plan de l'opposition entre voyelles accentuées et voyelles non-accentuées :

« *The Swedish data indicate that women use greater vowel duration contrasts than men do, the women producing shorter or similar vowel durations in non- stressed positions, and longer vowel durations in stressed positions.* » (Ericsson & Ericsson, 2001 : 36)

Selon ces auteurs, les résultats obtenus pour le suédois pourraient ne pas être en contradiction avec ceux obtenus sur d'autres langues, qui montrent en grande majorité que les voyelles des femmes seraient *globalement* plus longues que celles des hommes : ces divergences ne seraient en réalité qu'une question de méthodologie. En effet, Ericsson et Ericsson font remarquer que les principales études publiées jusqu'alors sur les différences hommes-femmes dans la durée des voyelles portaient soit exclusivement sur des voyelles en position accentuée (e.g. Hillenbrand et al., 1995, sur l'anglais américain), soit sur des moyennes qui mélangent voyelles accentuées et non-accentuées (e.g. Simpson, 1998, sur l'allemand). Notons tout de même que la recherche menée par Ericsson & Ericsson a été effectuée sur un nombre très restreint de locuteurs (cinq hommes, cinq femmes) ; il semble donc raisonnable de ne pas en tirer de conclusions hâtives.

La durée des voyelles est étroitement liée à la *réduction* des voyelles, c'est-à-dire à la centralisation de voyelles non accentuées, généralement accompagnée d'un affaiblissement de la durée. Dans une étude portant sur les productions de locuteurs de l'anglais américain, Byrd (1992, 1994) a découvert que les hommes réduisaient plus fréquemment leurs voyelles en [ə] que les femmes, et ce de manière significative. Cette même tendance a été observée par Whiteside (1996), sur une autre variété d'anglais (anglais nord-britannique).

Une explication socio-phonétique ?

Les différences évoquées dans cette section ont probablement une origine socio-phonétique. En effet, l'hypothèse selon laquelle ces variations seraient liées à la différence de taille du conduit vocal a été formellement contredite expérimentalement (voir Simpson, 2001, 2003). Dans un article ultérieur, Simpson en conclut donc que :

« As with vowel space size differences, a sociophonetic reason is the most likely. [...] the longer duration of stressed vowels, the greater durational distinction between stressed and unstressed vowels and less reduction of vowel qualities to [ə] can all be treated as phonetic correlates of speaking clearly. Alternatively, opposite patterns in male speakers can be seen as a tendency to speak less clearly. » (Simpson, 2009 : 636)

Les femmes auraient donc tendance à parler avec plus de *clarté* que les hommes. Cela rejoint les résultats de l'étude de Kramer (1977), portant sur les préjugés des auditeurs à propos des voix de femmes et d'hommes. Pour expliquer ce comportement, Simpson reprend à son compte une hypothèse avancée par Labov (1990) : les femmes adopteraient un type d'énonciation plus clair car elles seraient plus fréquemment que les hommes la principale source d'*input* langagier pour les nourrissons et jeunes enfants. Notons que cette hypothèse est critiquable à double titre : elle est invérifiable expérimentalement et se fonde sur des stéréotypes sexistes, voire machistes.

3.8 Durée des énoncés

Autre élément dépendant de la vitesse d'élocution, la durée des énoncés a elle aussi fait l'objet d'études comparatives entre les deux genres. Mais contrairement au point précédemment évoqué, à savoir la durée des voyelles, aucune tendance claire ne semble se dégager à l'échelle de l'énoncé.

Dans son étude publiée en 1994, portant sur plus de 600 locuteurs de l'anglais américain, Byrd trouve une différence significative entre la durée moyenne des mêmes énoncés selon qu'ils sont prononcés par des hommes ou par des femmes. Comme l'illustre la figure 16 ci-après, les énoncés produits par des hommes présentent une durée moyenne de 6,2 % inférieure à ceux produits par des femmes.

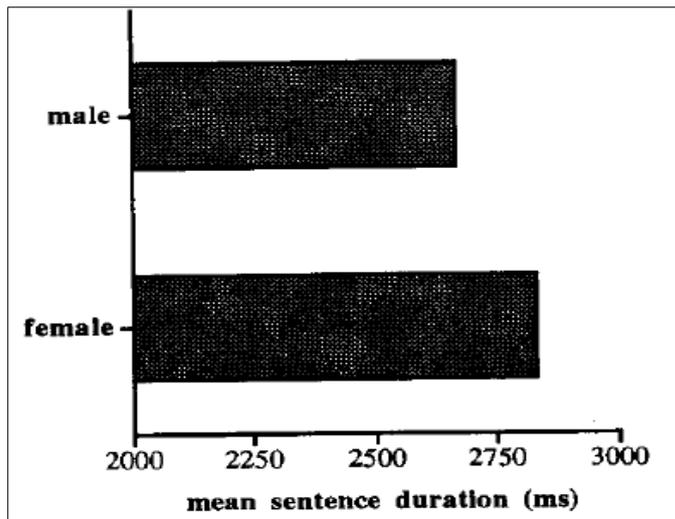


Figure 16 – Durée moyenne (en ms) des phrases produites par les hommes (en haut) et les femmes (en bas) locuteurs de l’anglais américain. (Byrd, 1994)

Une tendance analogue a été mise en évidence par Whiteside (1995, 1996), avec des locuteurs de l’anglais nord britannique, et par Fitzsimons et al. (2001) avec des locuteurs irlandais (débit environ 10 % plus rapide chez les hommes). De façon similaire, Simpson (1998) observe une durée globalement supérieure des énoncés produits par les femmes par rapport à ceux produits par les hommes, dans une recherche menée sur des locuteurs germanophones.

En revanche, aucune différence significative de durée à l’échelle de la phrase n’a été trouvée par Ray & Zahn (1990) et Jacewicz et al. (2009) sur l’anglais américain. Simpson et Ericsson (2003) ont abouti aux mêmes conclusions dans une étude portant sur des phrases produites par des locuteurs anglophones américains (26 femmes, 22 hommes) et suédoises (5 hommes et 5 femmes). Les auteurs ont simplement constaté une *distribution temporelle* différente : les consonnes tendaient à être sensiblement plus longues chez les hommes (ce qui rejoint les résultats obtenus pour les fricatives sourdes par Fox et Nissen, 2005), à l’inverse des voyelles accentuées, d’une durée globalement plus élevée chez les locutrices⁴¹. Le tout donnant au final une durée totale de l’énoncé relativement équivalente chez les hommes et les femmes, et ce dans les deux langues investiguées.

⁴¹ Ces observations n’ont pas fait l’objet de test statistique.

Il est donc pour le moment délicat de tirer des conclusions sur les différences inter-genres au niveau de la durée des énoncés : même si, comme nous l'avons vu, un grand nombre d'auteurs ont pu observer des énoncés significativement plus longs lorsque produits par des femmes, certains n'ont pas trouvé de différence significative. De nouvelles études sont donc nécessaires, en particulier dans d'autres langues telles que le français, l'espagnol et les langues asiatiques, où aucune donnée sur ce point ne semble être encore disponible. Dans tous les cas, les différences inter-genres sur ce paramètre acoustique ne peuvent avoir qu'une origine socio-linguistique : selon Whiteside (1995), les locuteurs masculins tendraient à parler avec un débit rapide et en réalisant peu de pauses afin de dominer les conversations, en réduisant ainsi les possibilités d'interruption et de prises de parole par les autres interlocuteurs.

3.9 Intensité

Contrairement à la structure formantique, la fréquence fondamentale ou encore la qualité de voix, l'intensité est un paramètre acoustique qui n'a été que très peu étudié du point de vue des différences femmes-hommes. Pausewang Gelfer et Ryan Young (1997) affirment que ce faible nombre de recherches s'explique par les difficultés méthodologiques que pose la mesure de l'intensité de la voix :

« Part of the problem in developing normative data for vocal loudness may stem from the difficulty encountered in measuring its physical correlate, vocal intensity. Intensity measures vary with the type of instrumentation used, as well as the signal processing parameters (e.g., integration time, elimination of pauses), and mouth-to-microphone distance. Further, it is very difficult to measure intensity from tape-recorded voice samples, unless careful calibration procedures have been followed. Finally, [...] vocal intensity can be expected to vary as a function of the speaking situation. » (p.178)

Malgré ces difficultés méthodologiques, ces auteurs ont entrepris une étude du niveau d'intensité des voix de 20 hommes et 20 femmes, anglophones américains âgés de 20 à 30 ans, en s'imposant une procédure expérimentale extrêmement rigoureuse. L'intensité moyenne des voix de femmes en conversation était de 68.15 dB, contre 70.42 pour les hommes (enregistrement effectué avec une distance locuteur-microphone de 30 cm). Cette faible différence (environ 2 dB) est néanmoins apparue comme étant

significative. Les différences concernant la plage de variation de l'intensité (*Available Intensity Range*, voir Schmidt et al., 1990) se sont quant à elles avérées non-significatives.

Les résultats obtenus par Pauswang Gelfer et Ryan Young, contredisent partiellement ceux de Brown et al. (1993) et Morris et al. (1995). Brown et al. ont analysé les productions de femmes chanteuses professionnelles et de femmes non-chanteuses, tandis que Morris et al. ont mené une étude similaire et sur des hommes chanteurs professionnels et non-chanteurs⁴². Si l'on rapproche ces deux études, les différences entre hommes et femmes apparaissent bien plus importantes que chez Pauswang Gelfer et Ryan Young (1997) : de l'ordre de 10 dB pour les non-chanteurs (63.1 dB pour les femmes contre 72.8 pour les hommes⁴³). Notons cependant que la tranche d'âge des locutrices enregistrées par Brown et al. (1993) paraît bien trop large (de 20 à 55 ans) et n'est pas la même que pour les locuteurs de Morris et al. (20 à 35 ans). Il est donc peu opportun de tirer des conclusions en rapprochant ces deux études. De plus, dans l'une comme dans l'autre, les différents réglages des équipements utilisés pour l'enregistrement ne sont pas détaillés, rendant ainsi très difficile la reproduction de ces expériences.

La recherche menée par Pauswang Gelfer et Ryan Young (1997) semble donc être la plus fiable qu'il existe à ce jour sur le sujet, et les résultats obtenus par ces auteurs ne font état que d'une très faible différence hommes-femmes. L'intensité n'apparaît donc pas, contrairement, par exemple, à la fréquence fondamentale ou à la structure formantique, comme un paramètre acoustique hautement variable en fonction du genre du locuteur.

Ce dernier paramètre acoustique fait donc figure d'exception, comparé aux autres évoqués dans cette section. En effet, comme nous l'avons vu, d'importantes différences acoustiques entre voix de femmes et voix d'hommes ont été établies, aussi bien dans le domaine fréquentiel, comme pour le F0 moyen et la position des formants vocaliques, que dans le domaine temporel (*VOT*, durée des voyelles, etc.). Lorsque des données existaient dans plusieurs langues, ce qui malheureusement n'était pas toujours le cas, une certaine

⁴² Les locuteurs et locutrices ayant participé à ces deux études sont tous anglophones américains.

⁴³ L'enregistrement a été ici effectué avec une distance locuteur-microphone d'environ 25 cm (10 pouces).

disparité inter-langues est souvent apparue. Ainsi, la plupart de ces différences inter-genres semblent ne pas pouvoir s'expliquer *uniquement* par les différences physiologiques (taille du conduit vocal, des plis vocaux) : une dimension culturelle entre également en jeu.

La section suivante traite des questions psycholinguistiques liées à ces différences acoustiques entre voix d'hommes et voix de femmes.

4. IMPLICATIONS PSYCHOLINGUISTIQUES DES DIFFERENCES ENTRE VOIX D'HOMMES ET VOIX DE FEMMES

Dans cette section, j'aborde les problématiques psycholinguistiques que soulèvent les différences acoustiques inter-genres. Il est question du traitement différencié des ces deux types de voix par le cerveau, de l'identification du genre par la voix, et des conséquences des différences acoustiques entre voix de femmes et voix d'hommes sur le traitement de la chaîne parlée.

4.1 Traitement différencié des voix de femmes et d'hommes par le cerveau humain

Sokhi et al. sont parvenus, dans une étude publiée en 2005, à mettre en évidence la réalité neurologique du traitement différencié des voix d'hommes et des voix de femmes. Les auteurs ont procédé, dans un premier temps, à l'enregistrement de 24 locuteurs anglophones britanniques (12 hommes et 12 femmes) lors de la lecture de plusieurs phrases. Ces stimuli ont ensuite été présentés à 12 auditeurs (des hommes anglophones britanniques). Grâce à l'IRMf⁴⁴, Sokhi et al. ont alors analysé les zones du cerveau qui étaient activées lors de l'écoute de ces stimuli acoustiques, étant soit des voix d'hommes, soit des voix de femmes. En outre, afin de s'assurer que les éventuelles différences ne seraient pas uniquement dues à la hauteur du F0, des stimuli de voix d'hommes et de voix de femmes dits ambigus⁴⁵, obtenus par resynthèse, ont également été présentés aux participants.

Dans tous les cas (stimuli ambigus ou non), les voix d'hommes et les voix de femmes n'ont pas activé de la même manière certaines régions du cerveau. Selon ces auteurs :

« The perception of female voices, when compared with male voices, affected greater activation of the right anterior superior temporal gyrus, near the superior temporal sulcus. Similarly, male voice perception activated the mesio-parietal precuneus area. These different gender associations could not be explained by

⁴⁴ Imagerie par Résonance Magnétique fonctionnelle.

⁴⁵ F0 moyen resynthétisé à 156,83 Hz pour les voix de femmes, et à 159,13 Hz pour les voix d'hommes.

either simple pitch perception or behavioural response because the activations that we observed were conjointly activated by both 'gender-apparent' and 'gender-ambiguous' voices. » (Sokhi et al., 2005 : 572)

La figure 17 indique la localisation de la zone du cerveau activée plus fortement par les voix de femmes (située dans le lobe temporal droit) et de celle activée plus fortement par les voix d'hommes (située dans le lobe occipital).

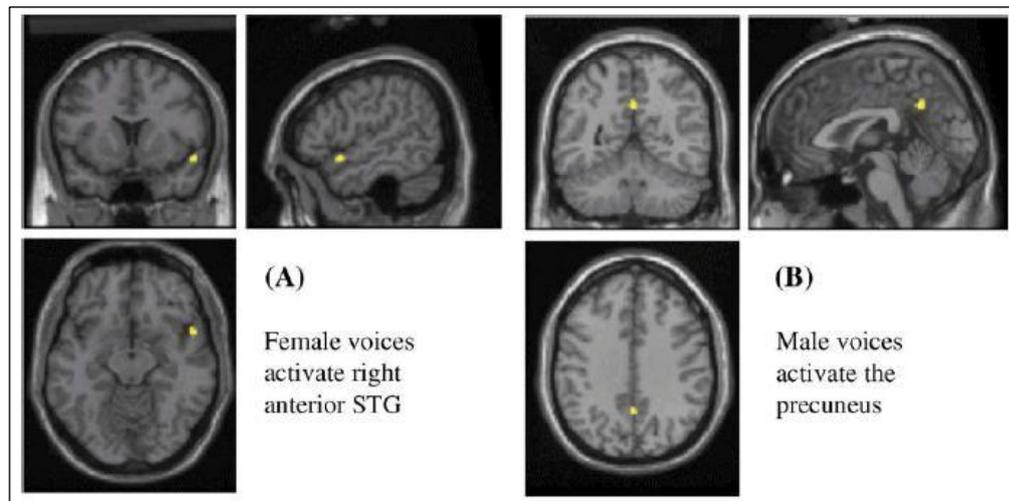


Figure 17 – IRMf montrant sous trois angles différents la zone du cerveau (en jaune) activée plus fortement par les voix de femmes (A) et celle activée plus fortement par les voix d'hommes (B). (Sokhi et al., 2005)

Globalement, le cortex auditif des auditeurs a été activé plus fortement par les voix de femmes. Selon Sokhi et al., cela s'accorde parfaitement avec l'hypothèse selon laquelle les voix de femmes seraient plus complexes, plus difficiles à traiter que les voix d'hommes⁴⁶. Cette hypothèse a déjà été étayée dans le domaine du traitement automatique de la parole, que ce soit en reconnaissance (e.g. Noyes & Frankish, 1989) ou en synthèse (voir notamment Klatt, 1987b).

Une étude très similaire à celle de Sokhi et al. a été publiée la même année par Lattner et al. (2005). Ces auteurs ont eux aussi présenté des phrases produites soit par des hommes, soit par des femmes, à des auditeurs, et ont observé à l'aide de l'IRMf les zones du cerveau activées par ces stimuli. A la différence de l'étude précédemment citée, les

⁴⁶ Voir section 4.3.2 du présent chapitre.

auditeurs étaient ici aussi bien des hommes (n=8) que des femmes (n=8), ayant pour langue maternelle l'allemand. Les résultats suivants ont été observés :

« [...] the contrast of the natural voice conditions M vs. F revealed stronger responses to the female voice than to the male voice in the right hemisphere and was comprised of three main centres of gravity: (1) in the supratemporal plane (STP) anteriorly to Heschl's gyrus [...] (2) in the posterior part of the superior temporal gyrus (STG) [...] and (3) in the superior part of the first long insular gyrus in the right hemisphere. [...] A second level analysis revealed no significant differences between the activation patterns of female and male listeners; both groups showed the right hemisphere dominance in response to the female voices.»
(Lattner et al., 2005 : 14-15)

Les voix de femmes ont donc activé plus fortement que les voix d'hommes l'hémisphère droit du cerveau des auditeurs. L'une des trois zones particulièrement concernée est la circonvolution temporale supérieure (STG) droite : c'est cette même zone qui a été plus fortement activée par les voix de femmes dans Sokhi et al. (2005). Les résultats de ces deux études sont donc très largement concordants. De plus, les résultats obtenus par Lattner et al. permettent désormais d'affirmer que :

- (1) Ces différences se retrouvent aussi bien chez les auditeurs masculins que féminins.
- (2) Ces différences ne semblent pas être propres à la langue (résultats très similaires chez des anglophones et des germanophones).

Une autre étude plus ancienne, menée par Dehaene-Lambertz (2000) sur des nourrissons, fournit d'autres éléments attestant le traitement différencié des voix de femmes et d'hommes. Cette chercheuse a étudié les réponses du cerveau des nourrissons à des stimuli linguistiques (syllabes ouvertes /ba/ et /ga/ produites par des voix d'hommes et de femmes) et non-linguistiques présentés auditivement. L'objet principal de cette expérience de neurolinguistique était d'établir le rôle respectif joué par les deux hémisphères du cerveau dans le traitement de stimuli auditifs, selon qu'ils sont de nature linguistique ou non. Cependant, l'auteur a également évoqué les réponses différentes aux stimuli linguistiques produits par un homme et ceux produits par une femme : les zones du cerveau activées n'ont pas été les mêmes selon le genre du locuteur.

Nous venons donc de voir que le cerveau humain traite de manière différente les voix de femmes et les voix d'hommes. La section suivante est dédiée au processus d'identification du genre par la voix.

4.2 Identification du genre par la voix

Si les voix d'hommes et de femmes sont traitées différemment par le cerveau humain, cela signifie que lors de l'écoute d'un stimulus linguistique, l'auditeur doit tenter d'identifier, de manière probablement inconsciente et automatique, quel est le genre du locuteur. Comme l'expliquent Le Breton (2011) et Simpson (2009), ce phénomène s'inscrit dans un processus plus large d'acquisition d'informations extralinguistiques :

« 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 : 621)

Pour Mullennix et al. (1995), cette identification est réalisée à partir de représentations auditives des voix de femmes et d'hommes, présentes dans le cerveau des auditeurs. Quels sont alors les indices acoustiques sur lesquels nous fondons notre jugement ? Quelle est leur importance respective ? Ces questions font l'objet des sous-sections suivantes.

4.2.1 Fréquence fondamentale moyenne et position des formants vocaliques : les deux indices majeurs ?

En premier lieu, le rôle de la position des formants vocaliques dans l'identification du genre par la voix a été mis en évidence dans une étude pionnière menée par Schwartz & Rine (1968) sur des locuteurs anglophones américains. Ces chercheurs ont testé la capacité des auditeurs à reconnaître le genre du locuteur lorsqu'ils sont privés de l'information véhiculée par la fréquence fondamentale. Pour cela, ils ont soumis à huit auditeurs des voyelles isolées chuchotées [i] et [a] (d'une durée de trois secondes) prononcées par cinq hommes et cinq femmes : l'identification du genre du locuteur a été

correcte dans plus de 97 % des cas. L'utilisation de voyelles isolées chuchotées ayant permis de neutraliser la quasi-totalité des indices acoustiques à l'exception du timbre, ce dernier serait donc à lui seul un indice acoustique suffisant pour permettre l'identification du genre par la voix :

« *The results of this experiment confirm the hypothesis that listeners are able to identify speaker sex from isolated productions of whispered vowels; the primary acoustic cue that underlies the distinction appears to be the upward frequency displacement of the resonance peaks of the female vowels.* » (Schwartz & Rine, 1968 : 1737)

Ces observations ont été confirmées ultérieurement par Bennett & Montero-Diaz (1982) et par Coleman (1971) avec des expériences du même type. Concrètement, plus les formants se trouvent dans des fréquences élevées, plus la voix sera perçue comme étant celle d'une femme (Coleman, 1976).

Dans une étude plus récente, Bédard et Belin (2004) ont également démontré le rôle joué par le timbre lors de l'identification du genre. Ces auteurs ont travaillé sur ce qu'ils appellent un « effet de la voix inversée » (« *voice inversion effect* »), afin de trouver d'éventuels indices qui suggéreraient l'existence d'un module de traitement spécifique à la voix. L'une de leurs expériences était une tâche d'identification du genre. Des syllabes lues par des hommes, des femmes et des enfants des deux sexes, étaient présentées à des sujets dans quatre conditions : sans inversion, inversion temporelle⁴⁷, inversion des fréquences⁴⁸ centrée autour de 4000 Hz, et inversion des fréquences centrée autour de 2500 Hz.

Il s'est avéré que l'inversion temporelle (timbre préservé), ne faisait que peu baisser le pourcentage de bonnes réponses, ce dernier passant de 78 % à 72 %. En revanche l'inversion des fréquences, qui modifie la position des formants et donc le timbre de la voix, a fait chuter le pourcentage de bonnes réponses à 65 % lorsqu'elle était centrée autour des 4000 Hz, et à 60 % lorsque centrée autour de 2500 Hz. Même si cela n'était pas l'objet premier de cette étude, ces résultats permettent de conclure, à l'instar de

⁴⁷ Lecture inverse du signal (sens droite-gauche).

⁴⁸ Inversion de la répartition des fréquences sur l'axe fréquentiel (rotation à 180° du spectrogramme).

Schwartz & Rine (1968), que les formants semblent jouer un rôle important dans l'identification du genre par la voix.

Le rôle de la fréquence fondamentale moyenne a quant à lui été maintes fois établi. Citons par exemple Coleman (1976), Pausewang Gelfer & Mikos (2005), deux études sur lesquelles je reviens plus en détails ultérieurement, Biemans (2000), Barlow & Wagner (2000), Fu et al. (2004), Bellandese (2009), ou encore Pépiot (2010). Dans cette dernière étude, menée à la fois sur des locuteurs anglophones américains et sur des francophones, l'auteur a montré que le fait de resynthétiser le F0 moyen de phrases produites par des hommes à une valeur caractéristique des voix de femmes (209 Hz) et vice versa (voix de femmes resynthétisées à 129 Hz) entraînait une diminution significative du pourcentage d'identifications correctes, ce dernier passant de 100 % (pour les voix naturelles) à 84,5 %. Une resynthèse du F0 moyen à un niveau dit intermédiaire (169 Hz) n'a en revanche pas entraîné une baisse significative du pourcentage d'identifications correctes, ce qui rejoint les résultats obtenus par Fauth (2012). Cette resynthèse à 169 Hz a néanmoins fait baisser significativement le degré de certitude des auditeurs.

Une récente étude de Nagle et al. (2012), tend à confirmer également l'importance de la fréquence fondamentale moyenne. Ces auteurs ont enregistré des patients américains équipés d'électro-larynx à différentes fréquences fondamentales (75, 130 et 175 Hz) : les auditeurs n'ont pas été à même d'identifier correctement le genre des locutrices sur les stimuli à 75 Hz. Mentionnons également une recherche intéressante réalisée par Andrews et Schmidt (1997) sur des hommes travestis, tous locuteurs anglophones américains. Les auteurs ont montré que ces derniers avaient pratiquement tous tendance à augmenter leur fréquence fondamentale lorsqu'ils imitaient une voix de femme. Des auditeurs ayant eu pour tâche de juger de la féminité ou de la masculinité de leurs voix se seraient principalement basés sur la fréquence fondamentale moyenne pour établir leur jugement, ce qui rejoint des constations déjà faites par Aronovitch en 1976, sur les stéréotypes associés aux voix d'hommes et de femmes. En outre, il est établi que la fréquence fondamentale moyenne est un paramètre décisif pour la construction de l'identité de genre chez les personnes transgenres (voir notamment Spencer, 1988 et Wolfe et al., 1990).

Enfin, Whiteside (1998a) a montré à travers une expérience menée sur des anglophones britanniques, que les auditeurs étaient capables d'identifier le genre des

locuteurs à partir de voyelles isolées de très courte durée (50 ou 100 ms) extraites de phrases lues, à près de 98,9 %⁴⁹. Ces résultats prouvent, selon l'auteur, que la position des formants vocaliques associée au F0 moyen sont deux indices acoustiques suffisamment puissants pour permettre l'identification du genre à partir d'un stimulus de quelques millisecondes⁵⁰.

Quelle hiérarchie entre ces deux paramètres ?

Si le F0 moyen et le timbre vocalique sont présentés quasi-unaniment comme les deux indices acoustiques les plus importants pour l'identification du genre, la hiérarchie entre ces deux indices est sujette à débat.

Le premier auteur à s'être penché sur cette question est Coleman (1976), dans une recherche menée sur des anglophones américains, et comportant deux expériences. Pour la première d'entre elles, 20 hommes et 20 femmes ont été enregistrés lors de la lecture d'un texte. Un extrait de cinq secondes tiré de chacun de ces enregistrements a ensuite été présenté⁵¹ à 17 auditeurs ayant pour tâche d'identifier le genre du locuteur et d'exprimer un degré de féminité / masculinité pour chaque voix. Après avoir analysé les caractéristiques acoustiques de chacune des voix soumises à jugement, Coleman a alors cherché à établir une corrélation entre le degré de féminité / masculinité moyen associé à chaque voix et le F0 moyen de cette dernière ; le même travail a été également effectué pour la position des formants. Les résultats obtenus sont très clairs :

« It can readily be seen that the listeners were basing their judgments of the degree of maleness or femaleness in the voice on the frequency of the laryngeal fundamental. The correlation coefficient of .94 represents an almost perfect one-to-one correspondence between the rankings of how male or female sounding a person's voice was judged to be and the frequency of his laryngeal fundamental. »
(Coleman, 1976 : 174)

⁴⁹ Ces résultats sont très proches de ceux obtenus en reconnaissance automatique du genre à partir de voyelles isolées (voir Zhu & Cahill, 1992 et Sigmund, 2008).

⁵⁰ Notons néanmoins que d'autres indices mineurs, tels que la qualité de voix (voir section 4.2.4) ont pu également jouer un rôle dans cette expérience.

⁵¹ Ces extraits ont été présentés à l'envers afin de limiter, selon l'auteur, l'influence de paramètres acoustiques autres que les formants et le F0 moyen, comme par exemple le rythme.

A l'inverse, seule une assez faible corrélation (0.59) est apparue entre fréquences de résonance des voyelles et degré de masculinité / féminité. Les auditeurs auraient donc davantage fondé leur jugement sur le F0 moyen que sur la position des formants.

Une deuxième expérience, bien moins convaincante, a été réalisée dans cette même étude. Cette fois-ci, l'auteur a tenté de faire varier le paramètre « F0 moyen » pour présenter aux auditeurs des voix avec F0 et timbre contradictoires (l'un étant typique d'une voix de femme, l'autre typique d'une voix d'homme), afin de voir lequel de ces deux paramètres allait s'imposer. Pour obtenir ses stimuli, Coleman a utilisé un vibreur laryngé. Placé au niveau du larynx du locuteur, ce dernier génère une fréquence fondamentale de 120 Hz ou de 240 Hz pendant que le sujet articulait, de façon muette, une voyelle. De l'aveu même de l'auteur, cette méthodologie, utilisée à l'époque car la resynthèse assistée par ordinateur n'était pas encore disponible, est très critiquable (voir notamment Fant et al., 1972). Les résultats de cette expérience sont d'ailleurs peu lisibles, voire contradictoires : les voix avec F0 féminin et timbre masculin, ainsi que celles avec F0 masculin et timbre féminin ont presque toutes été majoritairement identifiées comme des voix d'hommes. Il convient donc, pour cette étude de Coleman (1976), de ne retenir que les résultats obtenus lors de la première expérience.

Dans un article plus récent de Whiteside (1998c), l'importance respective du F0 moyen et du timbre pour l'identification du genre par la voix a également été testée. Pour cette étude, des voyelles synthétisées⁵² ont été présentées à 50 auditeurs anglophones britanniques dans quatre conditions : F0 et timbre féminin, F0 et timbre masculin, F0 féminin et timbre masculin, F0 masculin et timbre féminin. Au final, c'est le F0 moyen qui a prédominé dans la très grande majorité des cas pour l'identification du genre par les auditeurs :

« The results of this brief study suggest that the mean level of fundamental frequency appears to be the most salient cue that listeners use when judging whether synthesized vowels represent a man's or a woman's voice, with low values contributing the perception of maleness and high values contributing to the perception of femaleness. » (Whiteside, 1998c : 599)

⁵² Les valeurs de F0 et des formants attribuées lors de la synthèse sont celles ayant été mesurées sur de vraies productions de locuteurs britanniques enregistrés pour cette étude.

Ces résultats ont été confirmés quelques années plus tard par Pausewang Gelfer et Mikos (2005). Ces chercheuses ont elles aussi réalisé une expérience d'identification du genre basée sur des voyelles isolées. La méthodologie de cette expérience est très proche de celle utilisée par Whiteside (1998c). Des voyelles isolées obtenues par resynthèse ont été présentées à des auditeurs anglophones américains avec : soit le timbre en adéquation avec le F0 moyen (e.g. formants d'homme avec F0 moyen à 120 Hz), soit le timbre et le F0 moyen contradictoires (e.g. formants d'homme avec F0 moyen à 240 Hz). Les résultats de cette étude sont explicites :

« Results of this study showed unequivocally that fundamental frequency cues were more salient to listeners than were formant frequency cues. For example, when the formants of a man were paired with a female-appropriate fundamental frequency, listeners perceived a male speaker only 19.3 % of the time. When the formants of a woman were paired with a male-appropriate fundamental frequency, listeners perceived a female speaker only 20.2 % of the time. These results suggest that formant frequency cues do not contribute strongly to gender identification, at least in isolated vowels. » (Pausewang Gelfer & Mikos, 2005 : 553)

Il est cependant regrettable de n'avoir utilisé que deux positions extrêmes de F0 pour cette étude (120 et 240 Hz) : l'ajout d'une troisième condition, dans laquelle le F0 moyen se serait trouvé à un niveau « intermédiaire » et donc ambigu (aux alentours des 170 Hz) aurait pu permettre d'obtenir de précieuses données complémentaires sur le rôle effectivement joué par le F0 moyen dans l'identification du genre.

Une étude plus ancienne de Lass et al. (1976) donne des résultats concordant avec ceux obtenus par Whiteside (1998c) et Pausewang Gelfer & Mikos (2005). Cette recherche portait également sur l'identification du genre à partir de voyelles isolées chez des auditeurs anglophones américains, mais la nature des stimuli utilisés diffère. Les auteurs ont utilisé trois conditions : voyelles normales (timbre et F0 préservés), voyelles chuchotées (timbre préservé), et voyelles filtrées au dessus de 255 Hz (F0 préservé). Près de 1800 identifications ont alors été réalisées pour chaque condition, les résultats sont visibles sur la figure 18 ci-après.

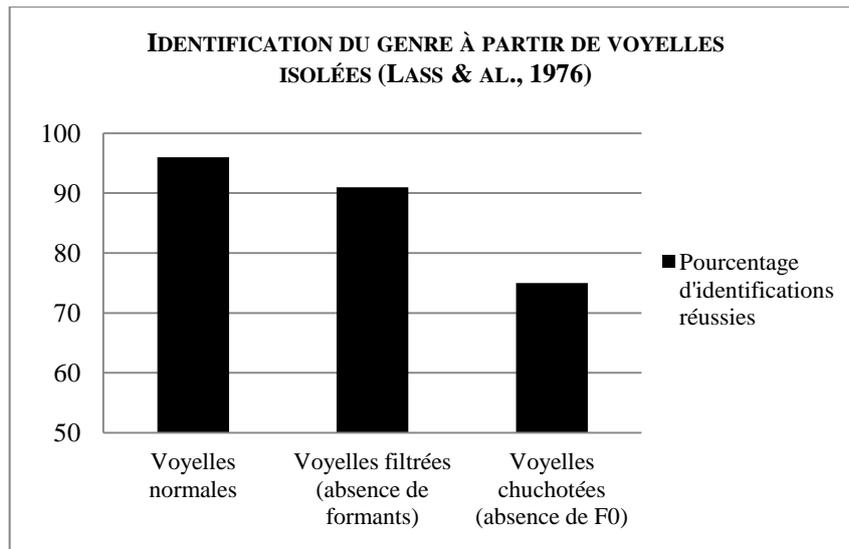


Figure 18 – Pourcentage d'identifications du genre réussies par des auditeurs anglophones américains à partir de voyelles normales, filtrées au dessus de 255 Hz, et chuchotées. (D'après Lass et al., 1976)

Alors que 96 % d'identifications correctes ont été obtenues pour les voyelles normales et 91 % pour les voyelles filtrées au dessus de 255 Hz (dépourvues de formants), ce chiffre chute à 75 % pour les voyelles chuchotées, qui ne possèdent donc pas de fréquence fondamentale. Le F0 moyen semble donc bien constituer un indice acoustique plus important que les formants vocaliques pour l'identification du genre, tout du moins sur des voyelles isolées et pour des locuteurs anglophones.

A l'inverse, une étude plus récente d'Arnold (2008) sur la contribution relative du F0 moyen et des fréquences de résonance dans l'identification du genre tend à montrer que le timbre serait le paramètre acoustique le plus déterminant. Des stimuli acoustiques (phrases isolées), présentant toutes choses égales par ailleurs des fréquences fondamentales moyennes et des fréquences de résonance⁵³ différentes, ont été soumis à des auditeurs francophones ayant pour tâche d'identifier le genre de la voix et de noter le degré de féminité ou de masculinité de celle-ci. Des corrélations significatives ont été trouvées entre variations des fréquences de résonance et catégorisation femme / homme (ainsi qu'avec le degré de féminité / masculinité). A l'inverse, les corrélations entre variations de fréquence fondamentale et catégorisation femme / homme furent non significatives. L'auteur en conclut que :

⁵³ Formants vocaliques mais également fréquences de résonance des consonnes.

« [...] le timbre est un paramètre prosodique plus important que la fréquence fondamentale dans la perception du genre ainsi que dans la perception du degré de "féminité" / "masculinité" des voix de "femmes". » (Arnold, 2008 : 103)

Les résultats obtenus dans cette étude d'Arnold apparaissent donc contradictoires avec ceux de Pausewang Gelfer & Mikos (2005) et de Lass et al. (1976). En revanche, ils semblent en adéquation avec les observations faites par Childers & Wu (1991) sur l'anglais américain, qui montrent qu'en traitement automatique de la parole, l'indice le plus fiable pour reconnaître le genre d'un locuteur sur des voyelles isolées serait statistiquement la position du deuxième formant (devant le F0 moyen). Remarquons par ailleurs que contrairement à Pausewang Gelfer & Mikos (2005) et Lass et al. (1976), qui utilisaient des stimuli de type voyelles isolées, Arnold (2008) a travaillé avec des phrases resynthétisées, en jouant non seulement sur la position des formants vocaliques mais également sur les fréquences de résonance des consonnes. Ceci a pu avoir une forte incidence sur les résultats observés.

Un autre facteur déterminant, et pourtant bien souvent négligé, a également varié entre ces différentes études : la langue maternelle des participants. Pausewang Gelfer & Mikos (2005) et Lass et al. (1976) ayant travaillé avec des auditeurs (et locuteurs) anglophones américains, alors que la recherche d'Arnold (2008) a été menée sur des francophones. Ainsi les résultats observés pourraient, si l'on en croit Pépiot (2010, 2011), ne pas être contradictoires. En effet, cet auteur a pu montrer *via* une expérience d'identification du genre par la voix menée conjointement sur des francophones et sur des anglophones américains, que ces derniers étaient plus sensibles à la fréquence fondamentale moyenne que leurs homologues francophones. Il est donc tout à fait possible que pour l'identification du genre, le F0 moyen constitue un indice acoustique plus important que les formants vocaliques chez les auditeurs anglophones, et que l'inverse soit vrai chez les francophones.

Autre élément important : la question de la hiérarchie entre ces paramètres est peut-être tout simplement mal posée. Est-il réellement possible d'établir quel est le paramètre acoustique intrinsèquement le plus important ? Selon Coleman (1971), il existerait une interaction entre ces deux indices acoustiques : l'importance que l'auditeur accorde à chacun de ces indices dépendrait ainsi de l'état de l'autre indice (typique d'une voix de femme ? D'une voix d'homme ? Ambigu ?). C'est également ce que suggère

Pépiot (2010, 2011). Pour cet auteur, 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. On peut dès lors supposer que ce phénomène se produit également en l'absence totale d'un paramètre acoustique habituellement utilisé (comme par exemple la fréquence fondamentale dans les stimuli de type « voix chuchotée »). Cela rejoint des observations faites concernant l'identification du locuteur : l'importance des différents indices acoustiques pour reconnaître une voix serait également relative (Van Dommelen, 2000).

Outre la position des formants vocaliques et la fréquence fondamentale moyenne, d'autres paramètres acoustiques semblent jouer un rôle dans l'identification du genre par la voix. Ces autres paramètres font l'objet des sous-sections suivantes.

4.2.2 Zones de bruit des fricatives sourdes

Plusieurs études ont été menées sur l'identification du genre par la voix à partir de fricatives sourdes isolées. Cependant, comme nous allons le voir, ces dernières ont uniquement été conduites sur des anglophones.

Schwartz (1968) a enregistré neuf femmes et neuf hommes anglophones américains lors de la production des fricatives /s/, /ʃ/, /f/ et /θ/ de manière isolée. Ces stimuli (d'une durée d'une seconde) ont alors été présentés à 10 auditeurs ayant pour tâche de reconnaître le genre du locuteur. Les résultats ont été les suivants :

« The results indicated that the listeners could identify the sex of the speakers from the isolated productions of /s/ and /ʃ/, but could not from the /f/ and /θ/ productions. » (Schwartz, 1968 : 1178)

Les auditeurs ont ainsi reconnu le genre dans 93 % des cas à partir de la fricative /s/, et dans 90 % des cas avec /ʃ/. Selon l'auteur, les réalisations de ces fricatives /s/ et /ʃ/ contiendraient donc des informations sur le genre du locuteur (concentration de l'énergie dans des fréquences plus élevées chez les femmes), qui sont effectivement exploitables par l'auditeur, tout du moins lorsque ces segments sont présentés isolément.

Une étude d'Ingemann (1968) publiée quelques mois plus tard fournit des éléments concordants. Cet auteur a lui aussi réalisé une expérience d'identification du genre à partir de fricatives sourdes auprès d'auditeurs anglophones américains. Les

consonnes ont été prononcées isolément (avec une durée variable) par neuf locuteurs et six locutrices. La gamme des fricatives sourdes testées a quant à elle été étendue à neuf items : /ɸ/, /f/, /θ/, /s/, /ʃ/, /ç/, /x/, /χ/ et /h/. Fort logiquement, le pourcentage d'identifications réussies a été globalement corrélé à la postériorité de la consonne : ainsi, les consonnes /ɸ/, /f/ et /θ/ ont engendré le plus faible taux d'identification. Cela s'explique aisément : plus la consonne est articulée à l'arrière du conduit vocal, plus la position des zones de bruit est affectée par les différences physiologiques hommes-femmes, rendant ainsi le genre du locuteur plus facilement identifiable.

Partant du constat que les expériences de Schwartz (1968) et Ingemann (1968) avaient été réalisées avec des fricatives sourdes produites isolément, présentant donc une durée bien supérieure à celles qu'on trouve dans le discours lu ou spontané, Whiteside (1998b) a entrepris une étude du même type, mais avec des stimuli extraits de phrases. Pour cette recherche menée avec des locuteurs anglophones britanniques, trois hommes et trois femmes ont été invités à lire quatre phrases contenant des fricatives sourdes. Dix segments (un [ʃ], quatre [s], trois [f] et deux [θ], avec des durées allant de 30 à 100 ms) ont ensuite été extraits de ces phrases pour chaque locuteur, soit un total de 60 stimuli. Ces derniers ont été présentés à une femme et deux hommes chargés d'identifier le genre du locuteur, chose qu'ils ont globalement réussie :

« A correct score of 64.4 % was attained for the identification of a speaker's sex from the fricative segments. [...] the observed values were significantly different from chance ($p < .01$). This indicates that there were sufficient cues in the brief fricative stimuli to make accurate judgements of the speaker's sex. » (Whiteside, 1998b : 590)

Malgré l'extrême brièveté des stimuli (30 à 100 ms, soit une durée plus de dix fois inférieure à celle des stimuli utilisés par Schwartz), les auditeurs ont réussi dans près de deux tiers des cas à identifier correctement le genre du locuteur. Cela permet donc de supposer que les fricatives sourdes puissent jouer un rôle effectif dans l'identification du genre par la voix, y compris à l'intérieur de phrases. Il est néanmoins regrettable que les résultats ne soient pas ici détaillés pour les quatre types de fricatives testés.

Notons par ailleurs que les conclusions des trois études détaillées précédemment concordent avec celles de Wu & Childers (1991). Ces auteurs ont montré qu'il était possible, en traitement automatisé de la parole, de reconnaître de manière relativement

fiable (à plus de 75 %) le genre des locuteurs à partir de leurs réalisations de fricatives sourdes, tout du moins pour l'anglais américain.

4.2.3 Intonation

A l'heure actuelle, le rôle de l'intonation⁵⁴ dans l'identification du genre par la voix n'a été que peu investigué. Une récente étude de Pépiot (2010, 2011), apporte néanmoins des éléments de réponse. Pour cette recherche, des phrases identiques ont été présentées à des auditeurs dans quatre conditions : voix naturelles, F0 moyen resynthétisé à 169 Hz avec contour intonatif préservé, F0 plat (contour intonatif neutralisé) à 169 Hz et F0 moyen des hommes et des femmes permutés (à 129 Hz et 209 Hz). Entre la deuxième et la troisième condition, seul le paramètre « intonation » a donc été neutralisé. Les deux groupes d'auditeurs (des anglophones américains et des francophones) ont alors eu pour tâche d'identifier le genre du locuteur ; les résultats pour les conditions 2 et 3 sont visibles sur la figure 19 ci-après.

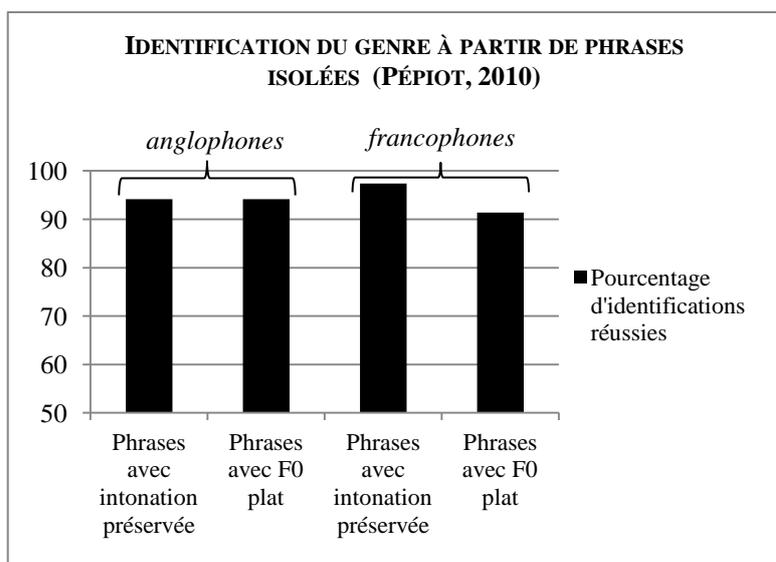


Figure 19 – Pourcentage d'identifications du genre réussies par des auditeurs anglophones américains et francophones à partir de phrases ayant soit leur contour intonatif préservé, soit un F0 plat. (D'après Pépiot, 2010)

Comme le montre la figure 19, la neutralisation du paramètre « intonation » n'a pas eu d'incidence sur les réponses émises par les auditeurs anglophones. En revanche,

⁵⁴ Pour rappel, le terme « intonation » désigne ici les variations de la fréquence fondamentale dans le temps.

chez les auteurs francophones, le pourcentage d'identifications réussies est passé de 97,4 % pour les phrases avec intonation préservée à 91,41 % lorsque le F0 a été aplati. Cette différence s'est avérée fortement significative. Notons que les jugements erronés dans cette condition expérimentale (F0 plat) sont principalement des voix de femmes identifiées comme des voix d'hommes. Ces résultats suggèrent donc que l'intonation jouerait bien un rôle dans l'identification du genre par la voix chez les auditeurs francophones, mais pas chez les anglophones.

4.2.4 Qualité de voix

Un récent article de Van Borsel et al. (2009), a mis en évidence la réalité perceptive de cet indice. Sept locutrices du hollandais, toutes non-fumeuses, ont été enregistrées lors de la production d'une voyelle isolée /a/ de manière normale et de manière soufflée⁵⁵ : 14 stimuli ont ainsi été obtenus. Ces derniers ont ensuite été présentés à 40 auditeurs (20 hommes et 20 femmes) ayant pour tâche d'évaluer le degré de féminité des voix qui leur étaient présentées sur une échelle de 1 à 5. Les résultats sont présentés dans la figure 20 ci-après.

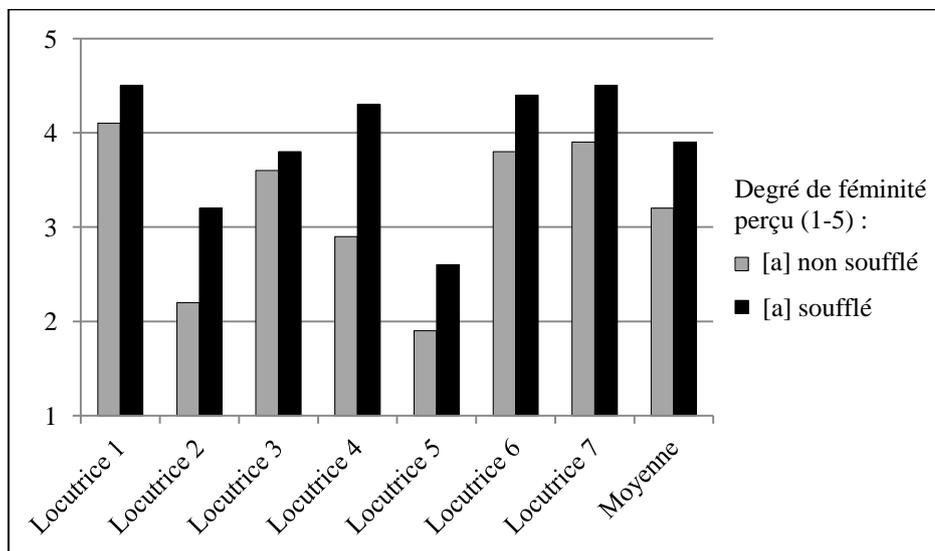


Figure 20 – Degré de féminité perçu (échelle de 1 à 5) pour des voyelles [a] produites par sept locutrices du hollandais en mode non soufflé et soufflé. Les jugements ont été émis par 40 auditeurs. (D'après Van Borsel et al., 2009)

⁵⁵ Les locutrices étaient soit orthophonistes, soit étudiantes en dernière année d'orthophonie : elles étaient donc parfaitement capables de produire des voyelles « soufflées » (ou « *breathy* »).

Quelle que soit la locutrice, les voix soufflées ont été perçues comme plus féminines que les voix non soufflées. Cette différence très significative suggère donc que le souffle (*breathiness*) contribue fortement à la perception d'une voix comme « féminine », ce qui rejoint des constatations faites par Johnson et al. (1999).

D'autre part, Andrews et Schmidt (1997) ont montré, à travers une étude menée sur 11 hommes travestis locuteurs de l'anglais américain, que ces derniers utilisaient presque tous une voix plus *breathy* lorsqu'on leur demandait de parler « comme une femme ». Quatre-vingt-huit auditeurs ayant eu pour tâche de caractériser la voix de ces personnes lorsqu'elle est en mode « femme » ou « homme » ont bien perçu les voix en mode « femme » comme étant plus soufflées et plus féminines.

On peut donc supposer, au regard de ces différentes études, que la qualité de voix puisse effectivement jouer un rôle dans la perception du genre. Soulignons cependant que le seul élément prouvé formellement est son influence sur le degré de féminité / masculinité perçu, pas sur la catégorisation homme / femme.

Outre les nombreux indices acoustiques évoqués jusqu'ici, d'autres paramètres peuvent jouer un rôle dans le processus d'identification du genre à partir de la voix.

4.2.5 Influence du contexte phonique

Schweinberger et al., dans une étude de 2008, ont mis en évidence l'importance de l'accommodation à un contexte phonique lors de tâches d'identification du genre par la voix. Quatre expériences ont été réalisées, avec pour chacune d'entre elles 12 à 24 participants (locuteurs germanophones, âgés de 19 à 30 ans, avec répartition équivalente femmes / hommes). La tâche des participants était la même pour toutes les expériences : identifier le genre de plusieurs locuteurs à partir d'une séquence VCV isolée (e.g. [aba], [igi], etc.). Ces stimuli ont été obtenus par resynthèse *via* un processus de *morphing acoustique* permettant de mélanger, dans les proportions souhaitées, les caractéristiques d'une voix d'homme et d'une voix de femme. Les items présentés étaient ainsi à différents niveaux « intermédiaires », sur le plan du F0 et des formants, entre une voix d'homme et une voix de femme : 80/20 %, 60/40 %, 50/50 %, 40/60 % ou 20/80 %.

Juste avant la présentation de chacun de ces items soumis à jugement, d'autres stimuli étaient présentés aux participants. La nature de ces stimuli d'adaptation (*adaptor stimuli*) variait d'une expérience à l'autre :

- **Expérience 1** (stimuli auditifs) : quatre séquences VCV prononcées soit uniquement par des hommes (condition 1), soit uniquement par des femmes (condition 2), soit avec voix ambiguë⁵⁶ (condition 3).
- **Expérience 2** (stimuli visuels) : cinq prénoms prototypiques masculins (condition 1) ou féminins (condition 2)⁵⁷.
- **Expérience 3** (stimuli visuels) : quatre extraits vidéo muets d'un visage articulant une séquence VCV. Les visages étaient soit uniquement masculins (condition 1) soit uniquement féminins (condition 2).
- **Expérience 4** (stimuli auditifs) : quatre sons (courbes sinusoïdales) correspondants au F0 d'une séquence VCV produite par un homme (condition 1) ou par une femme (condition 2).

Le but de ces différentes expériences était de tester si ces différents contextes, phoniques ou visuels, allaient influencer le jugement des auditeurs lors de leur tâche d'identification du genre par la voix. Les résultats de la première expérience ont été très concluants :

« Prior adaptation to male voices causes a voice to be perceived as more female (and vice versa), and these auditory aftereffects were measurable even minutes after adaptation. » (Schweinberger et al., 2008 : 684)

En d'autres termes, lorsque des voix de femmes ont été présentées comme stimuli d'adaptation, les items testés ont été plus souvent perçus comme des voix d'hommes, et inversement. Notons que ces différences ont été très largement significatives ($p < 0,001$). Logiquement, ce phénomène a été plus fort et plus durable sur les jugements émis pour les voix très ambiguës (50/50 %) que sur ceux portant sur les voix moins ambiguës

⁵⁶ F0 et formants vocaliques à un niveau intermédiaire homme / femme.

⁵⁷ Anna, Kathrin, Kristin, Lisa, et Sarah pour les prénoms féminins ; Martin, Patrick, Phillip, Stefan, et Dennis pour les prénoms masculins.

(80/20 %, 20/80 %). Comme le rappellent à juste titre les auteurs, cette découverte rejoint d'autres phénomènes similaires largement décrits par le passé en psychologie expérimentale : par exemple, la vision répétée d'un objet en mouvement descendant, suivie par la vision d'un objet statique entraîne la perception de ce dernier comme étant en mouvement ascendant (voir par exemple Antal et al., 2004).

Ces données doivent nécessairement être prises en compte lors de la conception d'expériences d'identification du genre par la voix : la présentation de plusieurs stimuli de même nature (voix de femmes ou voix d'hommes) est susceptible d'induire un biais dans les jugements portés sur les items suivants. Il conviendrait donc idéalement d'alterner la nature des stimuli, en ne présentant par exemple pas plus de deux voix de femmes ou d'hommes de manière consécutive, afin d'éviter tout effet d'accommodation.

Contrairement à la première expérience, les stimuli d'adaptation des expériences 2 et 3 n'ont pas joué sur les jugements des auditeurs : le fait de voir préalablement des prénoms d'hommes ou de voir préalablement des visages d'hommes articulant les sons présentés n'a pas entraîné une augmentation des jugements « voix de femmes » sur les stimuli testés, et inversement. Cela permet d'affirmer que pour ce type d'expérience il n'existe pas d'effet d'accommodation *visuel* : ce dernier semble être spécifiquement *auditif*.

Enfin, lors de l'expérience 4, le fait d'entendre préalablement des sons correspondants au F0 d'une séquence VCV produite par un homme (ou par une femme) n'a pas non plus eu d'effet sur les jugements des auditeurs. Voici les conclusions qu'en tirent les auteurs :

« Thus, voice perception was unaffected by whether participants previously adapted to low-pitched or high-pitched tones, corresponding in pitch to typical “male” or “female” fundamental frequencies. This shows that aftereffects on voice gender perception in experiment 1 clearly do not reflect simple adaptation to pitch. » (Schweinberger et al., 2008 : 686)

Une question se pose néanmoins : les auditeurs ont-ils pu percevoir un son pur (courbe sinusoïdale) comme un stimulus linguistique ? Cela semble très peu probable, même si la fréquence de ces sons correspondait au F0 moyen typique d'un homme ou d'une femme. Par conséquent, afin de savoir si le phénomène observé dans l'expérience 1

était uniquement lié au F0 des voix utilisées comme stimuli d'adaptation, interrogation par ailleurs tout à fait légitime, il aurait été préférable que les auteurs utilisent dans l'expérience 4 des voix d'hommes (condition 1) ou de femmes (condition 2) avec un même F0 intermédiaire⁵⁸ comme stimuli d'adaptation. Si l'effet observé dans l'expérience 1 était de nouveau présent, il aurait alors été possible d'arriver à la conclusion citée plus haut, de manière plus fiable.

Nous avons donc vu dans cette section les différents indices acoustiques et les autres facteurs pouvant jouer un rôle dans ce processus complexe qu'est l'identification du genre par la voix. Intéressons-nous maintenant aux conséquences qu'ont les différences entre les voix de femmes et d'hommes sur le traitement la chaîne parlée.

4.3 Conséquences des différences entre voix de femmes et voix d'hommes sur le traitement de la chaîne parlée

Les différences acoustiques entre les voix de femmes et d'hommes semblent avoir plusieurs conséquences sur le traitement de la chaîne parlée. Ces aspects sont développés dans cette sous-section.

4.3.1 Frontières perceptives et genre du locuteur

Dans une intéressante étude de 1999 conduite sur des anglophones américains, Johnson et al. ont utilisé des stimuli visuels (visages d'hommes ou de femmes) associés à des stimuli linguistiques acoustiques (mots monosyllabiques resynthétisés). Ces mots constituaient un continuum entre [hud] et [had], avec pour seule variable les valeurs du premier formant de la voyelle, comprises entre 450 et 750 Hz (7 degrés). De plus, chaque élément de ce continuum était produit soit par une voix d'homme prototypique⁵⁹, une voix d'homme non-prototypique, une voix de femme prototypique ou une voix de femmes non-prototypique, pour un total de 28 stimuli auditifs.

Ces items ont été présentés aux auditeurs en même temps que des stimuli visuels représentant soit un visage d'homme, soit un visage de femme : la tâche des 20

⁵⁸ Aux alentours de 160 Hz.

⁵⁹ Les voix ont préalablement été évaluées par des auditeurs.

participants (5 hommes et 15 femmes) était d'identifier le mot prononcé, avec un choix forcé *hood* / *hud*. On peut se demander dans un premier temps, sans tenir compte du facteur « genre du visage présenté », si la frontière phonémique entre /u/ et /ʌ/ est la même lorsque selon que la voix traitée soit une voix d'homme ou une voix de femme. Comme le montre la figure 21, ci-dessous, ce n'est pas le cas.

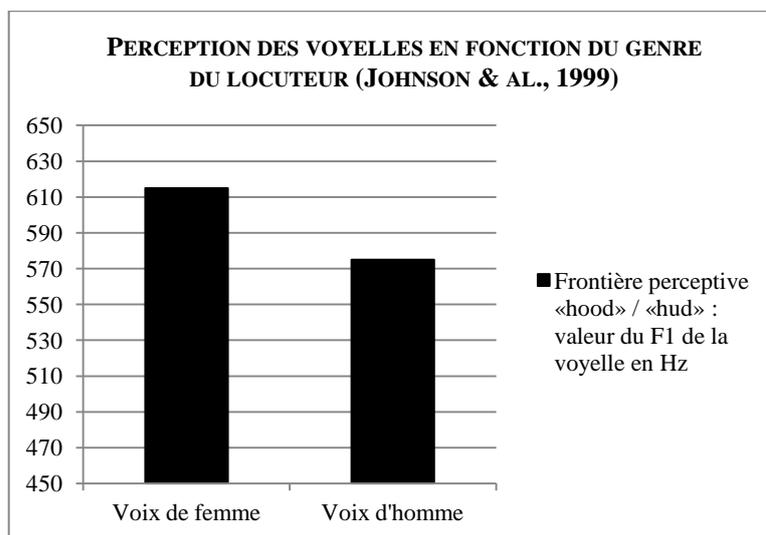


Figure 21 – Frontière perceptive entre hood et hud (valeur du F1 de la voyelle -en hertz-) en fonction du genre du locuteur, chez des auditeurs anglophones américains. (D'après Johnson et al., 1999)

La frontière perceptive entre /u/ et /ʌ/ s'est située, sur le plan du F1, à 615 Hz pour les voix de femmes contre 575 Hz pour les voix d'hommes (différence largement significative). Une voyelle avec un même timbre peut donc être perçue différemment selon qu'elle est produite par une femme ou par un homme. D'autre part, le genre du visage présenté a également influé de manière significative sur la position de la frontière phonémique (quel que soit le genre de la voix présentée), créant ainsi une sorte d'*effet McGurk*⁶⁰ : un visage d'homme faisant globalement descendre cette frontière et inversement pour le visage de femme.

⁶⁰ Voir McGurk & MacDonald (1976) et MacDonald & McGurk (1978). La présentation d'un stimulus acoustique (ex : syllabe [ba]) en même temps qu'un stimulus visuel « contradictoire » (ex : visage articulant la syllabe [ga]) entraîne généralement la perception d'un stimulus dit « intermédiaire » (ex : syllabe [da]) : l'information véhiculée par un stimulus visuel peut donc interférer sur la perception d'un stimulus acoustique.

Strand & Johnson (1996) ont réalisé une expérience similaire, en s'intéressant cette fois-ci à l'identification des fricatives sourdes /s/ et /ʃ/, à l'intérieur d'un mot monosyllabique (*sod* ou *shod*). A l'instar de Johnson et al. (1999), les auditeurs (20 femmes et 8 hommes anglophones américains) ont vu leur frontière phonémique varier en fonction du genre des locuteurs :

« [...] *the boundary between [ʃ] and [s] is generally at a lower frequency for men since lower formants⁶¹ are expected in their [s] compared with the same fricative as produced by women.* » (Strand & Johnson, 1996 : 20)

Les auditeurs s'attendaient donc à des zones de bruit plus basses pour les fricatives produites par des hommes que pour celles produites par des femmes. Des constatations identiques découlent des résultats obtenus plus récemment par Munson et al. (2006). Par ailleurs, dans l'étude de Strand & Johnson, la présentation d'un visage d'homme ou d'un visage de femme en même temps que les stimuli auditifs a également créé un *effet McGurk* similaire à celui observé dans Johnson et al. (1999).

Ces trois études suggèrent donc que pour l'identification des phonèmes, les frontières (en terme de position des formants ou des zones de bruit) ne sont pas les mêmes selon que l'auditeur identifie une voix comme féminine ou comme masculine. Ce dernier semble donc connaître et prendre en considération les différences acoustiques qu'il existe entre ces deux types de voix lorsqu'il tente d'identifier un phonème. Cela rejoint l'idée générale développée notamment par Nguyen (2007), Miller (1994) ou encore Repp & Libermann (1987), selon laquelle l'auditeur ajusterait les frontières qu'il établit entre les phonèmes en fonction du contexte (phénomène de « compensation perceptive »).

4.3.2 Les voix de femmes plus difficiles à traiter que celles des hommes ?

Les locutrices ont très souvent été négligées par les phonéticiens, en particulier dans les études nécessitant des relevés formantiques (voir Ferragne & Pellegrino, 2010). Dans un article de 1982, Johansson et al. dressaient d'ores et déjà le constat suivant :

« *Comparatively little is known about the characteristics of the female voice as compared with the male voice. The background is the high fundamental frequency*

⁶¹ Les auteurs désignent ici par *formants* les zones de bruit des fricatives sourdes.

range of the female voice which makes formant frequency estimates uncertain and, hence, information on the voice source unsafe. » (Johansson et al., 1982 : 117)

Ainsi, le peu d'intérêt porté à l'analyse des voix de femmes s'explique en partie⁶² par le fait que leurs formants vocaliques sont généralement plus durs à localiser que ceux de leurs homologues masculins sur les spectrogrammes et les spectres (voir également Pépiot, 2009). En effet, comme nous l'avons vu précédemment, les locuteurs féminins ont un F0 moyen globalement plus élevé que celui des hommes : leurs voix présentent donc moins d'harmoniques, entraînant ainsi des formants⁶³ moins bien définis. Qu'en est-il alors de l'auditeur ? Les voix présentant un F0 élevé sont-elles plus difficiles à traiter par le cerveau ? C'est ce que suggèrent un certain nombre d'études.

Un premier article important a été publié en 1982 par Ryalls & Lieberman. Dans cette étude menée sur des locuteurs anglophones américains, des voyelles isolées synthétisées ont été présentées à des auditeurs ayant pour tâche de les identifier. Le F0 des voyelles était soit de 100 Hz, de 135 Hz, ou de 250 Hz, et demeurait stable sur chaque voyelle. Cette expérience a été menée dans un premier temps avec des voyelles présentant une structure formantique de voix d'homme, puis répétée avec une structure formantique de voix de femme, en utilisant les valeurs de Peterson & Barney (1952). Dans tous les cas, les voyelles avec un F0 à 100 Hz et à 135 Hz ont été significativement mieux identifiées par les auditeurs que celles avec un F0 à 250 Hz. Selon les auteurs :

« These results suggest that the human formant frequency extractor is aided by the denser spectral sampling of the transfer function by a lower fundamental. »
(Ryalls & Lieberman, 1982 : 1631)

Autrement dit, une voix présentant beaucoup d'harmoniques (i.e. un F0 bas) facilite la localisation des formants par l'auditeur et donc l'identification des voyelles.

Ces résultats rejoignent ceux obtenus plus récemment par Diehl et al. (1996), dans une étude très similaire. Les auteurs ont synthétisé des voyelles isolées /i/ et /u/, avec des fréquences fondamentales statiques allant de 90 à 390 Hz par palier de 60 Hz, et une

⁶² Il n'est pas exclu que la pensée sexiste, qui tend à considérer comme secondaire le genre féminin, soit également à l'origine d'une partie de ce désintérêt.

⁶³ Les formants étant des zones d'harmoniques renforcés.

structure formantique d'homme ou de femme⁶⁴. Les 24 stimuli ainsi obtenus ont été présentés à 11 auditeurs anglophones américains, chargés d'identifier chacune des voyelles. Le pourcentage d'identifications réussies en fonction de la fréquence fondamentale (tous stimuli confondus) est visible sur la figure 22.

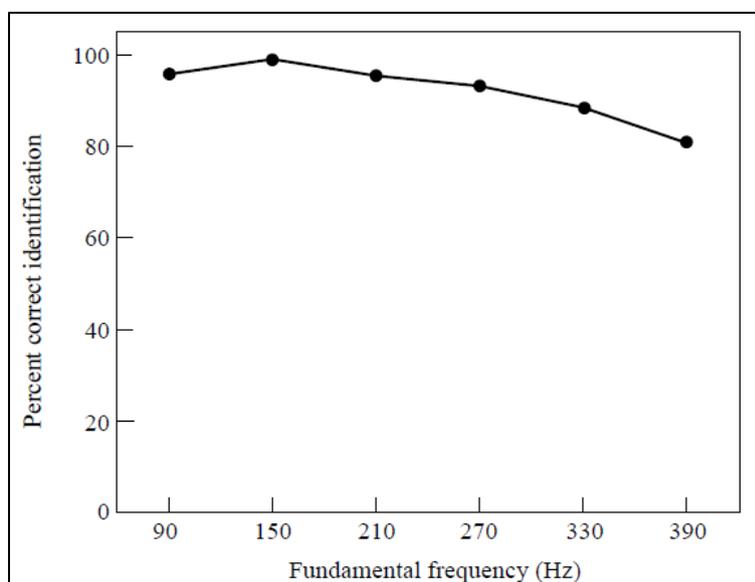


Figure 22 – Pourcentage d'identifications correctes de voyelles isolées /ɪ/ et /ʊ/ en fonction de leur F0, chez des locuteurs anglophones américains. (Diehl et al., 1996)

On constate que le pic d'identifications correctes a été atteint à 150 Hz⁶⁵ : au-dessus de cette fréquence fondamentale, le pourcentage chute de manière significative et quasi-constante.

Les études de Ryalls & Lieberman (1982) et Diehl & al (1996), réalisées avec des stimuli de synthèse, suggèrent donc que les voyelles produites avec un F0 élevé seraient plus difficiles à identifier par les auditeurs. Des constatations analogues ont été faites concernant les voyelles chantées (voir Howie & Delattre, 1965 ; Scotto Di Carlo & Germain, 1985 ; Gottfried & Chew, 1986). Plus intéressant, on remarque que l'utilisation d'un triangle vocalique plus large (structure formantique de voix de femmes) ne permet pas de compenser entièrement ce phénomène.

⁶⁴ Tout comme Ryalls & Lieberman (1986), les auteurs ont utilisé les valeurs de Peterson & Barney (1952).

⁶⁵ Selon les auteurs, le fait que le pourcentage d'identifications correctes soit plus faible à 90 Hz qu'à 150 Hz s'explique en grande partie par des jugements erronés émis par un seul auditeur.

Il n'est pas possible d'affirmer que les voix de femmes sont *systématiquement* plus difficiles à traiter que les voix d'hommes, car dans certains cas exceptionnels, un homme peut disposer d'une fréquence fondamentale plus élevée qu'une femme. Néanmoins, nous pouvons conclure que *globalement*, cette affirmation semble plausible, tout du moins sur le plan de l'identification des voyelles isolées.

Qu'en est-il alors du temps de traitement ? Compte tenu des constatations précédentes, il serait logique que ce dernier soit proportionnel au F0 de la voix traitée. Dans une étude menée sur des anglophones américains, Strand (2000) a présenté des mots isolés produits par un homme ou une femme à des participants (n=24) ayant pour tâche de répéter le mot le plus rapidement possible. Les temps de réponse ont été mesurés et comparés dans quatre conditions correspondant aux quatre stimuli testés : voix d'homme stéréotypique, voix de femme stéréotypique, voix d'homme non-stéréotypique (i.e. ambiguë) et voix de femme non-stéréotypique⁶⁶. Les résultats sont présentés dans la figure 23, ci-après.

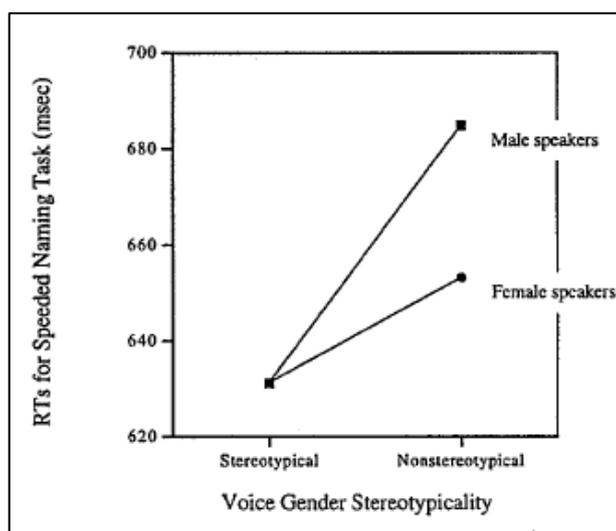


Figure 23 – Temps de réaction moyen (en ms) de sujets anglophones américains ayant pour tâche de répéter un mot, selon que ce dernier est produit par une voix de femme (stéréotypique ou non) ou une voix d'homme (stéréotypique ou non). (Strand, 2000)

⁶⁶ Ces différentes voix ont été choisies préalablement sur la base de jugements émis par 24 auditeurs.

Aucune différence significative n'est apparue entre la voix d'homme et la voix de femme stéréotypiques (temps de réaction moyens respectivement à 631,26 ms et 631,30 ms). Les voix ambiguës (femme et homme) ont quant à elles entraîné un temps de réaction significativement plus long, ce qui était prévisible. Toujours dans cette même étude (Strand, 2000), une autre expérience dérivée de celle-ci mais avec l'ajout de stimuli visuels a montré que la présentation d'un visage ne correspondant pas au genre du locuteur (exemple : visage d'homme et voix de femme) faisait augmenter significativement le temps de réaction par rapport à la condition dans laquelle visage et voix correspondaient (exemple : visage de femme et voix de femme).

Cette étude suggère donc que les voix de femmes ne seraient pas plus longues à traiter par le cerveau que celles des hommes. Cependant, seules une voix d'homme et une voix de femme (stéréotypiques) ont été utilisées, ce qui ne permet pas de tirer des conclusions définitives. De plus, plutôt que le paradigme de répétition de mots, celui de la détection de mots (ou *word spotting*) aurait sans doute été plus révélateur car il implique uniquement un travail de perception. Par ailleurs, le fait que la vue d'un visage de l'autre genre entraîne un temps de traitement plus long pourrait laisser supposer que l'écoute préalable de voix du genre opposé (par exemple des voix d'hommes avant une voix de femme) puisse également modifier le temps de traitement. Cette hypothèse pourrait elle aussi être testée dans le cadre d'une tâche de détection de mots.

5. DISCUSSION - CONCLUSION PARTIELLE

Ce chapitre a donc permis d'aborder quelques notions théoriques essentielles autour de la question du genre et de la phonétique, puis de passer en revue les principales études réalisées dans le domaine des différences acoustiques entre les productions des locuteurs féminins et masculins. Enfin, les questions d'ordre psycholinguistique relatives à ces différences femmes-hommes ont été évoquées, au travers, là encore, des principales publications sur le sujet.

Au plan acoustique, la *fréquence fondamentale moyenne* (voir Takefuta et al., 1972 ; Boë et al., 1975 ; Kitzing, 1979) et la *position des formants vocaliques* (Hillenbrand et al., 1995 ; Ericsson, 2005 ; Pépiot, 2009) apparaissent comme les principales différences entre voix de femmes et d'hommes : le F0 moyen étant globalement plus élevé chez les locuteurs féminins (190 à 210 Hz contre 110 à 130 Hz pour les hommes, selon les auteurs et les langues étudiées), de même que les fréquences de résonance des voyelles. Ces phénomènes ont pu être observés dans plusieurs langues, y compris différentes variétés d'anglais et le français standard. Néanmoins, les différences de valeurs formantiques semblent être une donnée hautement variable en fonction de la langue (Johnson, 2005).

En dehors de ces deux grands paramètres, d'autres différences ont été mises en évidence. Tout comme pour les formants vocaliques, les *zones de bruit des fricatives sourdes* (à l'exception des fricatives interdentes et labiodentes) se trouveraient dans des fréquences plus élevées chez les femmes, comme l'ont montré, par exemple, Schwartz (1968) et Whiteside (1998b). Pour ce qui est de l'*intonation*, plusieurs études font état d'une tendance à l'utilisation d'une plage de variation plus large chez les femmes (e.g. Takefuta et al., 1972, pour l'anglais américain ; Hwa Chen, 2007, pour le mandarin, même si ce constat n'est pas partagé par tous les auteurs (voir notamment Herbst, 1964, 1969).

D'autres différences femmes-hommes ont également été observées, telles que l'utilisation des types de phonation *creaky* et *breathy*. Dans certaines langues comme l'anglais américain, le premier mode serait utilisé plus fréquemment par les locuteurs masculins, et le second par leurs homologues féminins (Klatt & Klatt, 1990 et Henton, 1989b, 1992). Enfin, certaines tendances semblent exister sur le plan du *voice onset time*,

généralement plus court chez les hommes (voir Swartz, 1992), ainsi que pour la *durée des voyelles et des énoncés*, globalement supérieure pour les femmes (voir Wassink, 1999 et Byrd, 1994). Notons que ces derniers points font encore l'objet de débats (voir notamment Oh, 2011, pour le *VOT*) et n'ont été étudiés que sur un nombre très restreint de langues.

L'ensemble des différences inter-genres ainsi observées semblent avoir deux origines distinctes : les *différences physiologiques et anatomiques* au niveau des plis vocaux et du conduit vocal d'une part (Fant, 1966 ; Kahane, 1978 ; Titze, 1989), et le *facteur culturel* ou *socio-linguistique* d'autre part (voir Eckert, 1989 ; Haan & Van Heuven, 1999 ; Johnson, 2005, 2006 ; Simpson, 2009). Ainsi, comme le mentionne Simpson :

« *A large body of sociolinguistic research has repeatedly shown that gender is one of the most important factors that must be considered when trying to account for phonetic variation found within a speech community [...]. Most importantly for our discussion here is that, at the time we take a snapshot of the sounds being produced by the men and women in the same speech community, we must expect part of the male-female differences we observe to be a direct result of men and women talking differently.* » (Simpson, 2009 : 13)

L'influence respective de ces deux facteurs (*physiologique / culturel*) dépend du paramètre acoustique. Le rôle du socio-linguistique étant nécessairement prépondérant lorsque les différences hommes-femmes observées varient fortement d'une langue à l'autre, comme pour les valeurs formantiques, le type de phonation (ou qualité de voix), la durée des énoncés ou encore le délai d'établissement du voisement.

Concernant ces différences acoustiques inter-genres, il sera intéressant de mener une analyse acoustique *multiparamétrique* de productions d'hommes et de femmes locuteurs de différentes langues *au sein d'une même étude* : le chapitre 2 sera donc dédié à une étude de ce type, conduite sur les productions de locuteurs anglophones américains et francophones parisiens. Cela permettra de connaître avec plus de certitude les éventuelles variations inter-langues des différences femmes-hommes (mêmes conditions d'enregistrement et d'analyse acoustique).

La dernière grande partie de ce chapitre était consacrée aux questions psycholinguistiques que soulèvent les différences acoustiques entre voix d'hommes et voix de femmes. En premier lieu, le développement relativement récent des techniques de neuro-imagerie a permis de mettre en évidence le traitement différencié de ces deux types de voix par le cerveau humain. Ainsi, les voix de femmes activeraient plus fortement que celles des hommes certaines zones situées dans l'hémisphère droit du cerveau (voir Sokhi et al., 2005, sur des auditeurs anglophones et Lattner et al., 2005, chez des germanophones). Une question se pose alors : comment identifions-nous les voix de femmes et les voix d'hommes ?

Le processus d'identification du genre par la voix apparaît au final comme extrêmement complexe. Si la *position des formants vocaliques* (Schwartz & Rine, 1968 ; Coleman, 1971 ; Arnold, 2008, 2012) et le *F0 moyen* (Coleman, 1976 ; Whiteside, 1998c ; Pausewang Gelfer & Mikos, 2005) sont quasi-unanimement considérés comme les deux indices acoustiques principaux, leur importance respective reste à définir. Il existe très probablement une interaction et donc des phénomènes de compensation entre ces différents indices (Pépiot, 2010). L'influence d'autres paramètres acoustiques, comme la *qualité de voix* (Van Borsel et al., 2009) ou encore l'*intonation* (Pépiot, 2010, 2011), est également avérée. La *position des zones de bruit des fricatives sourdes* (Schwartz, 1968 ; Whiteside, 1998b) joue elle aussi un rôle, mais ce dernier n'est pas encore formellement établi pour une échelle supérieure au segment isolé. D'autre part, l'importance des indices acoustiques semble être hautement variable selon les langues (Pépiot, 2010, 2011). En revanche, aucune étude n'a pu mettre en évidence des différences entre auditeurs féminins et masculins locuteurs d'une même langue : il est donc fortement probable qu'ils utilisent des stratégies similaires.

Sur ce processus d'identification du genre par la voix, et en dépit du nombre assez conséquent de recherches réalisées, plusieurs questions restent actuellement en suspens. Tout d'abord, les différentes expériences ont été conduites avec plusieurs types d'unités linguistiques : fricatives sourdes isolées, voyelles isolées, syllabes isolées, mots, ou phrases. On sait donc que, pour des stimuli « naturels⁶⁷ », les locuteurs obtiennent un taux d'identification correcte de l'ordre de 63 à 90 % à partir d'une fricative sourde (selon les

⁶⁷ N'ayant pas subi d'opération de resynthèse.

auteurs et la longueur des stimuli, voir Schwartz, 1968 et Whiteside, 1998c), ce taux progresse légèrement sur des voyelles isolées (voir notamment Lass et al., 1976 ; Pausewang Gelfer & Mikos, 2005), pour atteindre un score proche des 100 % avec un énoncé long d'une douzaine de syllabes (Pépiot, 2010, 2011). Mais ces recherches ont été menées de manière séparée, par des auteurs différents, à plusieurs années voire décennies d'intervalle, presque systématiquement sur une seule langue, avec parfois des segments isolés (voyelles, fricatives sourdes) d'une durée bien supérieure aux occurrences réellement présentes dans la chaîne parlée (e.g. Schwartz, 1968 ; Ingemann, 1968 et Schwartz & Rine, 1968). De plus, dans l'ensemble des études évoquées plus haut, il n'est pratiquement jamais question de *degré de certitude* des auditeurs, une donnée pourtant cruciale.

Ainsi, il n'existe à ce jour aucune analyse *fine* et *inter-langues* du processus d'identification du genre par la voix. Pour réaliser une telle analyse, l'utilisation du paradigme de *gating*⁶⁸ adapté à l'identification du genre s'avèrera particulièrement pertinente, *a fortiori* si l'expérience est menée conjointement sur plusieurs langues. Cela permettra de répondre de manière détaillée aux questions suivantes : à partir de quelle quantité d'information acoustique l'auditeur reconnaît-il, de manière certaine, le genre d'un locuteur ? Quelle est l'influence du facteur langue ? Certaines voyelles ou certaines consonnes sont-elles plus propices que d'autres à l'identification du genre ? De plus, cette expérience devra porter sur des stimuli ayant préalablement fait l'objet d'une analyse acoustique détaillée (chapitre 2), afin d'avoir la possibilité de connaître précisément les indices exploités par l'auditeur dans cette expérience de perception. Le chapitre 3 sera donc consacré à cette étude.

Pour finir, nous avons vu que les voyelles produites avec un F0 élevé (supérieur à 150 Hz, ce qui est le cas de la très grande majorité des voix de femmes) donnaient lieu à un taux d'erreur d'identification plus important que les voyelles produites avec un F0 bas (Ryalls & Lieberman, 1982 et Diehl & al, 1996). En revanche, il n'a pas pu être montré que les voix de femmes (ayant donc un F0 globalement plus élevé que celles des hommes) étaient plus longues à traiter par le cerveau : selon Strand (2000), les voix d'hommes et de femmes dites « stéréotypiques » entraînent un temps de réponse

⁶⁸ Paradigme de la psycholinguistique mis au point par Grosjean (1980) et consistant en un dévoilement progressif des stimuli.

équivalent lors d'une tâche de répétition de mots. Il est donc fort probable que les auditeurs soient capables, tout du moins à l'échelle du mot, de traiter aussi rapidement les voix de femmes et d'hommes.

Néanmoins, l'étude de Strand ne permet pas d'arriver formellement à cette conclusion. Outre le fait d'avoir testé uniquement une voix d'homme et une voix de femme, présentées toutes deux comme « stéréotypiques », le paradigme utilisé par l'auteur nécessite à la fois une tâche de perception et une tâche de production : dès lors, il est impossible de connaître la répartition du temps de réponse entre ces deux processus. Sur ce point, une expérience de détection de mots (ou *word detection*), paradigme qui appelle exclusivement une tâche de perception / compréhension (voir notamment Marslen-Wilson & Tyler, 1980), permettra d'apporter des éléments de réponse, de même que sur un possible effet d'*accommodation de l'auditeur au genre du locuteur*, c'est-à-dire de modification du temps de traitement lors de l'écoute de plusieurs stimuli consécutifs du même genre (homme ou femme). Cette expérience sera présentée dans le chapitre 4.

HYPOTHESES GENERALES

En me basant sur les différents éléments présentés dans ce premier chapitre, j'ai établi les hypothèses suivantes :

Hypothèse 1 : Les voix de femmes et d'hommes présentent des différences acoustiques significatives au niveau de la fréquence fondamentale et des fréquences de résonance.

Hypothèse 2 : Il existe également des différences inter-genres sur d'autres composantes acoustiques de la parole, à savoir la plage de variation de F0, la qualité de voix, la durée des énoncés et le *VOT* des consonnes occlusives.

Hypothèse 3 : Les différences acoustiques entre voix d'hommes et voix de femmes varient en fonction de la langue.

Hypothèse 4 : Il est possible pour un auditeur d'identifier le genre par la voix à partir de stimuli de très faible durée.

Hypothèse 5 : Le processus d'identification du genre par la voix diffère en fonction de la langue maternelle de l'auditeur.

Hypothèse 6 : Les voix d'hommes et de femmes sont traitées par le cerveau de l'auditeur comme deux entités distinctes.

Hypothèse 7 : Toutes choses égales par ailleurs et malgré les différences acoustiques qu'elles présentent, les voix de femmes et d'hommes sont globalement traitées à la même vitesse par l'auditeur.

CHAPITRE 2

**ANALYSE ACOUSTIQUE DE PRODUCTIONS DE
FEMMES ET D'HOMMES ANGLOPHONES
AMERICAINS ET FRANCOPHONES PARISIENS**

RESUME

Un grand nombre d'études ont été menées sur les différences acoustiques entre les productions de locuteurs féminins et masculins. Cependant, elles sont presque systématiquement réalisées sur des locuteurs d'une même langue et portent le plus souvent sur un seul paramètre acoustique. La présente étude est une analyse acoustique de mots et pseudo-mots dissyllabiques produits par quatre locuteurs anglophones du nord-est des Etats-Unis et quatre francophones parisiens. Les fréquences de résonance, le F0 moyen, la plage de variation de F0, le VOT, la différence d'intensité H1-H2 ainsi que la durée des énoncés ont été mesurés. Des différences inter-genres significatives ont été observées dans les deux langues sur chacun des paramètres testés. D'autre part, d'importantes variations inter-langues ont été constatées, sur le plan de la plage de variation de F0, des formants vocaliques et de la différence H1-H2. Ces résultats suggèrent que les différences acoustiques hommes-femmes sont en partie construites socialement et dépendantes de la langue.

ABSTRACT

Many studies were conducted on acoustic differences between female and male speech. However, they were generally led on a group of speakers of a same language, and focused on a single acoustic parameter. The present study is an acoustic analysis of dissyllabic words or pseudo-words produced by four Northeastern American English speakers and four Parisian French speakers. Resonant frequencies, mean F0, F0 range, VOT, H1-H2 intensity differences and words' durations were measured. Significant cross-gender differences were obtained for each tested parameter. Moreover, cross-language variations were observed for F0 range, vocalic formants and H1-H2 differences. These results suggest that cross-gender acoustic differences are partly socially constructed and language dependent.

MOTS-CLES : voix et genre, parole et genre, différences acoustiques inter-genres, variations inter-langues, français parisien, anglais américain.

KEYWORDS: voice and gender, speech and gender, cross-gender acoustic differences, cross-language variations, Parisian French, American English.

Ce deuxième chapitre est dévolu à une analyse acoustique contrastive de productions d'hommes et de femmes anglophones américains et francophones parisiens⁶⁹. Une première section est dédiée à la méthode, la seconde à l'analyse des données. Enfin, la dernière section est consacrée aux résultats : ces derniers seront dans un premier temps présentés de manière brute, puis à travers leur analyse statistique.

1. METHODE

Dans cette section sont décrits le choix du matériau linguistique et des locuteurs, ainsi que la procédure d'enregistrement.

1.1 Matériau linguistique

Afin de mener à bien cette étude, du matériau linguistique anglais et français était nécessaire. Les mots isolés m'ont paru être la solution la plus appropriée, car ces derniers ont pour vocation d'être utilisés, par la suite, comme stimuli dans l'expérience de *gating* appliqué à l'identification du genre par la voix⁷⁰. Plus précisément, j'ai opté pour des *mots et pseudo-mots*⁷¹ *dissyllabiques* afin de pouvoir tester aisément un nombre important de combinaisons de phonèmes.

Le choix des pseudo-mots dissyllabiques s'est fait sur plusieurs critères : faire apparaître le plus de combinaisons de consonnes et de voyelles *pertinentes*⁷² possibles tout en limitant raisonnablement le nombre d'items et obtenir une correspondance la plus fidèle possible entre les items anglais et français. Sur la base de ces critères il est apparu indispensable de tester :

- Des **fricatives sourdes et voisées** en initiale de mot : /s/, /z/, /ʃ/ et /ʒ/ pour le français comme pour l'anglais. Les fricatives interdentes et labiodentes (/θ/, /ð/, /f/ et /v/) ont été sciemment écartées du fait du peu de différences acoustiques

⁶⁹ Une partie des résultats de cette étude sont également présentés dans Pépiot (2013).

⁷⁰ Voir chapitre 3.

⁷¹ Items comportant des combinaisons de phonèmes permises par les contraintes phonotactiques d'une langue, mais ne figurant pas dans le lexique de cette dernière.

⁷² Au regard des différentes études menées jusqu'ici sur les différences acoustiques entre voix d'hommes et voix de femmes et sur l'identification du genre par la voix. Voir chapitre 1.

entre les réalisations des hommes et des femmes (voir Ingemann, 1968, et Whiteside, 1998b). Notons que pour l'anglais, la fricative voisée /z/ en initiale de mot n'existe que dans des mots d'emprunt. Cependant, on la retrouve dans un certain nombre de mots anglais dans d'autres positions, comme dans /mezə/ ou /lezə/.

- Des **occlusives sourdes et voisées** en initiale de mot : /t/, /d/, /k/ et /g/ pour l'anglais comme pour le français. Les occlusives bilabiales ont été écartées pour la même raison que les fricatives interdentes et labiodentes.
- Des **voyelles** en initiale de mot et après une attaque consonantique : /i/, /a/ et /u/ pour le français et /i:/, /æ/ et /u:/ pour l'anglais. Elles représentent les trois grands types de variation observables au niveau des différences acoustiques hommes-femmes sur les voyelles du français et de l'anglais (F2 nettement plus élevé chez les femmes sur les voyelles antérieures fermées, telles que le /i/ et le /i:/, F2 et F1 plus élevés chez les femmes sur les voyelles ouvertes, telles que /a/ et /æ/, et valeurs formantiques proches sur les voyelles postérieures fermées, telles que /u/ et /u:/, voir notamment Hillenbrand et al., 1995 et Pépiot, 2009). Pour l'anglais, le phonème /æ/ a été préféré au /a/, car il est articulatoirement (voyelle ouverte antérieure) plus proche du /a/ français.

Outre ces éléments, j'ai décidé de choisir une combinaison de deux phonèmes (de type CV), apparaissant de manière constante en fin de mot. C'est finalement la suite /pi/ qui a été retenue, du fait que cette dernière peut apparaître sur la totalité des fins de mots sans violer les contraintes phonotactiques du français et de l'anglais. Par ailleurs, les mots retenus pour l'anglais portent tous l'accent lexical sur la première syllabe. Les 27 mots et pseudo-mots suivants ont ainsi été retenus :

- **Combinaisons de type /C occlusive – V – p – i / :** /tipi/, /tapi/, /tupi/, /dipi/, /dapi/, /dupi/, /kipi/, /kapi/, /kupi/, /gipi/, /gapi/, /gupi/ pour le français, /'ti:pi/, /'tæpi/, /'tu:pi/, /'di:pi/, /'dæpi/, /'du:pi/, /'ki:pi/, /'kæpi/, /'ku:pi/, /'gi:pi/, /'gæpi/, /'gu:pi/ pour l'anglais.

- **Combinaisons de type /C fricative – V – p – i / :** /sipi/, /sapi/, /supi/, /zipi/, /zapi/, /zupi/, /ʃipi/, /ʃapi/, /ʃupi/, /ʒipi/, /ʒapi/, /ʒupi/ pour le français, /^ʔsi:pi/, /^ʔsæpi/, /^ʔsu:pi/, /^ʔzi:pi/, /^ʔzæpi/, /^ʔzu:pi/, /^ʔʃi:pi/, /^ʔʃæpi/, /^ʔʃu:pi/, /^ʔzi:pi/, /^ʔzæpi/, /^ʔzu:pi/ pour l'anglais.
- **Combinaisons de type /V – p – i / :** /ipi/, /api/, /upi/ pour le français, /^ʔi:pi/, /^ʔæpi/, /^ʔu:pi/ pour l'anglais.

1.2 Choix des locuteurs pour les enregistrements

Pour ces enregistrements, j'ai fait appel à huit participants âgés de 23 à 40 ans. Une moitié étant francophone (locuteurs du français dit *parisien*⁷³), et l'autre moitié anglophone (plus précisément constituée de locuteurs de l'anglais américain). Chacun de ces deux groupes est constitué de deux hommes et deux femmes, dont les enregistrements, une fois segmentés et étiquetés, seront analysés et comparés. L'ensemble des locuteurs sont monolingues, non-fumeurs et ne présentent aucun trouble de la parole.

Une attention particulière a été portée sur le choix des quatre locuteurs américains, afin d'éviter d'obtenir des différences acoustiques qui pourraient être causées uniquement par l'origine géographique des locuteurs (voir Clopper, 2004 ; Clopper et al., 2005). Pour ce faire, les quatre locuteurs choisis sont tous originaires⁷⁴ d'une même zone du nord-est des Etats-Unis couvrant les Etats de Pennsylvanie, du Massachussets, de New-York (hors ville de New-York), et le sud du Vermont, c'est-à-dire une zone assez peu étendue et au sein de laquelle il n'existe pas ou peu de variations diatopiques, en particulier pour les phonèmes testés dans cette étude (Labov et al., 2006 ; Clopper, 2004 ; Clopper et al., 2005). De plus, tous ces locuteurs ont obtenu au moins un diplôme universitaire et évoluent dans des milieux socioculturels relativement proches, ce qui a pour effet de limiter là encore les possibles différences dialectales (Labov et al., 2006).

Voici les principales caractéristiques pertinentes des locuteurs francophones et anglophones :

⁷³ Les individus désignés ici comme locuteurs du français *parisien* vivent en Ile-de-France depuis plus de 10 ans et ne présentent pas d'accent régionalement marqué.

⁷⁴ Et y ont également résidé la majorité de leur vie.

- **Locutrice francophone numéro 1 (F1FR)** : 27 ans, francophone native, étudiante en sciences du langage niveau M2, F0 moyen⁷⁵ : 205 Hz, taille⁷⁶ : 162 cm.
- **Locutrice francophone numéro 2 (F2FR)** : 23 ans, francophone native, étudiante en information-communication niveau L3, F0 moyen : 197 Hz., taille : 175 cm.
- **Locuteur francophone numéro 1 (H1FR)** : 23 ans, francophone natif, étudiant en information-communication niveau L2, F0 moyen : 115 Hz, taille : 183 cm.
- **Locuteur francophone numéro 2 (H2FR)** : 24 ans, francophone natif, étudiant en information-communication niveau L3, F0 moyen : 125 Hz, taille : 182 cm.
- **Locutrice anglophone numéro 1 (F1AN)** : 40 ans, anglophone native (anglais américain), originaire⁷⁷ de Northampton (Massachusetts), enseignante en anglais, F0 moyen : 195 Hz, taille : 174 cm.
- **Locutrice anglophone numéro 2 (F2AN)** : 23 ans, anglophone native (anglais américain), originaire de l'Etat de du Vermont, étudiante en psychologie niveau Master, F0 moyen : 216 Hz, taille : 169 cm.
- **Locuteur anglophone numéro 1 (H1AN)** : 39 ans, anglophone natif (anglais américain), originaire de Philadelphie (Pennsylvanie), pasteur en formation, F0 moyen : 116 Hz, taille : 178 cm.
- **Locuteur anglophone numéro 2 (H2AN)** : 26 ans, anglophone natif (anglais américain), originaire de Binghamton (Etat de New-York), enseignant en anglais, F0 moyen : 97 Hz, taille : 185 cm.

⁷⁵ Les F0 moyens des locuteurs ont tous été mesurés avec le logiciel *Praat*, sur l'ensemble des 54 phrases enregistrées pour cette étude (« il a dit *MOT* deux fois » * 27 mots * 2 lectures : voir section 1.3).

⁷⁶ Cette mesure donne une indication de la longueur relative du conduit vocal des différents locuteurs du même genre. En effet, au sein d'une population du même genre (femmes ou hommes), la taille du conduit vocal est très fortement corrélée à la taille globale de l'individu, comme nous l'avons vu dans la section 3.3 du chapitre 1 (voir notamment Fitch & Giedd, 1999 ; Cherng et al., 2002).

⁷⁷ Les locuteurs anglophones ont été enregistrés à l'occasion d'un séjour d'une durée plus ou moins longue à Paris.

1.3 Procédure d'enregistrement

Afin d'homogénéiser les paramètres prosodiques sur les mots et pseudo-mots⁷⁸ devant servir pour l'expérience, chacun d'entre eux a été placé dans le contexte suivant pour les enregistrements : « *Il a dit MOT deux fois* » pour les mots⁷⁹ français, et « *He said MOT twice* » pour les mots anglais. L'utilisation de ce type de contexte permet notamment de neutraliser l'« effet de liste » (F0 montant sur tous les items d'une liste de mots jusqu'à l'avant-dernier, puis descendant sur le dernier item).

D'autre part, tous les mots et pseudo-mots apparaissent en transcription orthographique sur la feuille destinée aux locuteurs. Les transcriptions orthographiques utilisées pour les items en français sont visibles dans le tableau 1 ci-dessous.

Tableau 1 – Transcription orthographique utilisée pour les mots et pseudo-mots présentés aux locuteurs francophones.

Transcription phonologique	Transcription orthographique	Transcription phonologique	Transcription orthographique
/tipi/	tipi	/supi/	soupi
/tapi/	tapi	/zipi/	zipi
/tupi/	toupi	/zapi/	zapi
/dipi/	dipi	/zupi/	zoupi
/dapi/	dapi	/ʃipi/	chipi
/dupi/	doupi	/ʃapi/	chapi
/kipi/	kipi	/ʃupi/	choupi
/kapi/	kapi	/ʒipi/	jipi
/kupi/	koupi	/ʒapi/	japi
/gipi/	guipi	/ʒupi/	joupi
/gapi/	gapi	/ipi/	hipi
/gupi/	goupi	/api/	hapi
/sipi/	sipi	/upi/	houpi
/sapi/	sapi		

Les transcriptions orthographiques apparaissant sur la feuille destinée aux anglophones sont quant à elles présentées dans le tableau 2, ci-après.

⁷⁸ La liste des mots est visible dans la section 1.1.

⁷⁹ Par commodité, l'emploi du terme « *mots* » au sens large, c'est-à-dire incluant les *mots* et *pseudo-mots*, sera privilégié dans ce chapitre.

Tableau 2 – Transcription orthographique utilisée pour les mots et pseudo-mots présentés aux locuteurs anglophones.

Transcription phonologique	Transcription orthographique	Transcription phonologique	Transcription orthographique
/ˈti:pi/	teepy	/ˈsu:pi/	soupy
/ˈtæpi/	tappy	/ˈzi:pi/	zeepy
/ˈtu:pi/	toupy	/ˈzæpi/	zappy
/ˈdi:pi/	deepy	/ˈzu:pi/	zoupy
/ˈdæpi/	dappy	/ˈʃi:pi/	sheepy
/ˈdu:pi/	doupy	/ˈʃæpi/	shappy
/ˈki:pi/	keepy	/ˈʃu:pi/	shoupy
/ˈkæpi/	kappy	/ˈʒi:pi/	jeepy
/ˈku:pi/	koupy	/ˈʒæpi/	jappy
/ˈgi:pi/	geepy	/ˈʒu:pi/	joupy
/ˈgæpi/	gappy	/ˈi:pi/	eapy
/ˈgu:pi/	goupy	/ˈæpi/	appy
/ˈsi:pi/	seepy	/ˈu:pi/	oupy
/ˈsæpi/	sappy		

Les locuteurs anglophones ont été informés préalablement que les trois items « *jeepy* », « *jappy* » et « *joupy* » se prononçaient avec la consonne /ʒ/ en initiale, et non avec l’affriquée /dʒ/. Aucune autre consigne sur la prononciation des mots n’a été donnée de manière systématique aux locuteurs ; ils étaient simplement informés en cas de besoin (demande explicite de leur part ou prononciation ne correspondant pas à ce qui était attendu pour un item particulier).

Le matériel utilisé est un enregistreur numérique portatif *Edirol R-09HR* de marque Roland, comportant un microphone doté d’une bande passante allant de 20 Hz à 40 kHz. Les enregistrements ont été effectués au format Wave⁸⁰, avec la sensibilité des microphones réglée en mode élevé, et sans utiliser les différentes options disponibles, telles que l’ « Auto Gain Control⁸¹ » et le « Low Cut⁸² », qui auraient pu engendrer des distorsions du signal. Pour les précautions d’enregistrement, j’ai très largement tenu compte des conseils dispensés par Philippe Martin (2008) et Jean-Michel Tarrier (2003).

Pour chaque enregistrement, une phase de réglage avait lieu dans un premier temps. Durant cette phase, j’ajustais le volume d’enregistrement de manière appropriée

⁸⁰ En stéréo (pas d’enregistrement direct en mono possible avec l’*Edirol R-09HR*), avec encodage à 44,1 kHz. Ces fichiers ont ensuite été convertis en mono, en préservant la fréquence d’échantillonnage.

⁸¹ Fonctionnalité permettant l’ajustement automatique du volume d’enregistrement.

⁸² Suppression des basses fréquences (utilisé notamment pour limiter les bruits de vent lors d’enregistrements en extérieur).

(intensité élevée, mais sans saturation) pendant que le participant parlait avec le microphone positionné à une trentaine de centimètres de sa bouche, mais sans être enregistré. Par la suite, les locuteurs étaient invités à lire les phrases à une vitesse « normale » et comme si ces dernières étaient des « *phrases isolées* ». Chacune des phrases devait être lue deux fois, afin de me laisser la possibilité ultérieure de choisir l'occurrence la plus satisfaisante acoustiquement. La distance entre leur bouche et le microphone était maintenue constante, toujours aux environs de trente centimètres.

Afin de limiter au maximum le bruit de fond et la réverbération, les enregistrements ont eu lieu dans une pièce isolée phoniquement (chambre anéchoïque), à l'écart des bruits de circulation, sans appareil électrique bruyant en fonctionnement (ordinateurs, ventilateurs, néons...). L'enregistreur était posé sur un socle, lui-même posé sur un tapis en mousse afin de l'isoler au mieux de toutes vibrations. De plus, je me suis assuré que les participants ne génèrent pas de bruits parasites (coups sur la table, le sol, ou même froissement de la feuille sur laquelle était écrite la phrase à lire).

La procédure d'enregistrement a été décrite. Voyons maintenant en détails l'opération d'extraction des mots de leur contexte.

1.4 Extraction des mots

Afin d'obtenir les mots isolés et de procéder à leur analyse acoustique, chaque mot a dû être extrait des phrases enregistrées.

L'extraction des mots de leur contexte « *Il a dit MOT deux fois* » ou « *He said MOT three times* » a été effectuée à l'aide du logiciel *Praat*⁸³. Cette opération a été réalisée manuellement : pour ce faire, je me suis appuyé à la fois sur le spectrogramme, l'onde sonore et la courbe de F0 afin de déterminer avec précision les frontières de mots. Une fois ces frontières identifiées, une première sélection du mot a été effectuée. Une illustration de cette opération est visible sur la figure 24 ci-après.

⁸³ *Praat 5.1.44*. Voir Boersma & Weenink (2010).

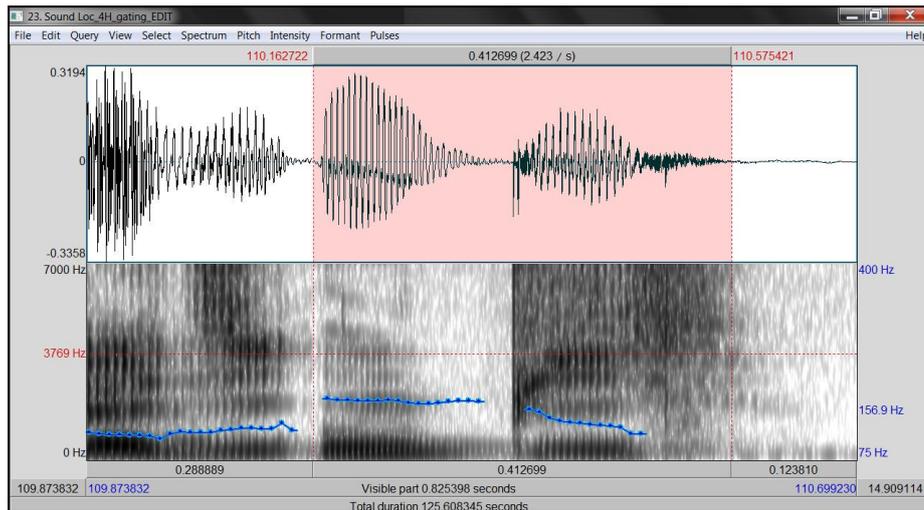


Figure 24 – Extraction du mot *api* de la phrase *il a dit « api » deux fois prononcée* par le locuteur H2FR à l'aide du logiciel Praat. La partie supérieure représente l'onde sonore ; la partie inférieure le spectrogramme et la courbe de F0 (en bleu). La zone sélectionnée figure en rose sur l'onde sonore.

Lorsque le mot débutait par une occlusive non-voisée, étant donné l'absence d'indice permettant d'identifier formellement le début de la consonne, j'ai choisi d'inclure la totalité du silence pré-relâchement, en plaçant la borne gauche à la fin du mot précédent. Dans un deuxième temps, les bornes gauche et droite de la sélection ont été décalées au passage par zéro le plus proche (opération automatisée par la création d'un script) afin d'éviter les *clics* parasites. Chacun des mots a ensuite été exporté dans un fichier au format Wave, en maintenant la fréquence d'échantillonnage à 44 kHz.

L'ensemble de la méthodologie a maintenant été exposée. La section suivante présente l'analyse des données.

2. ANALYSE DES DONNEES

Je présente dans cette section les différentes étapes de l'analyse des données, à savoir la segmentation et l'étiquetage, la mesure du F0 moyen, de la plage de variation et de la durée des mots, l'analyse acoustique multidimensionnelle des voyelles et celle des consonnes : l'ensemble de ces opérations ont été effectuées à l'aide du logiciel d'analyse acoustique *Praat*⁸⁴. Tous les fichiers sons analysés dans cette étude sont librement téléchargeables⁸⁵.

2.1 Segmentation et étiquetage

Le préalable à l'analyse acoustique détaillée a été la segmentation et l'étiquetage de tous les mots.

Cette opération a également été réalisée avec le logiciel *Praat*. Pour cela, j'ai créé pour chacun des fichiers un objet *TextGrid* avec trois paliers⁸⁶ : « phones », « C/V » (dans lequel est indiquée la position correspondant au phone : C1, V1, C2⁸⁷, V2) et « mot ». Les différents mots ont ensuite été segmentés et étiquetés. Pour des raisons pratiques, c'est l'alphabet X-SAMPA⁸⁸ qui a été utilisé pour l'étiquetage des phones plutôt que l'Alphabet Phonétique International. D'autre part, plusieurs règles strictes ont été appliquées pour la segmentation :

- Tous les segments sont contigus. Le début du segment 2 correspond à la fin du segment 1 et ainsi de suite.
- Chaque borne se trouve sur un passage par zéro (utilisation de la fonction *Move cursor to nearest zero crossing*).
- Lorsqu'un mot débute par une consonne occlusive non-voisée, la totalité du silence précédant le relâchement (voir section précédente) est incluse dans la C1.

⁸⁴ *Praat* version 5.1.43

⁸⁵ Voir annexe 1.

⁸⁶ De l'anglais « tiers ».

⁸⁷ Pour les mots sans consonne initiale (« *api* », « *ipi* », etc.), cette consonne est toujours notée C2.

⁸⁸ Alphabet phonétique utilisant uniquement des caractères présents sur les claviers informatiques standards. Pour plus de détails, voir Wells (2000).

- Le début et la fin de la voyelle 1 sont identifiés à l'aide de l'onde sonore. La borne gauche est placée à l'endroit où le signal devient relativement stationnaire et de forte amplitude, la borne droite au moment où le signal perd cet aspect.
- La borne signalant la fin de la consonne 2, à savoir le son [p], et le début de la voyelle 2, correspondant au son [i], est elle aussi identifiée en fonction de l'aspect de l'onde sonore. La borne est placée lorsque cette dernière devient relativement stationnaire et ample.
- Pour les locuteurs francophones, lorsque le [i] final présente un appendice consonantique de type [ç], ce dernier est codé sur le palier des phones : son début correspond à la fin de l'onde sonore régulière et ample. Néanmoins, sur le palier C/V, cet élément est bien compris dans la V2.

Un exemple de mot segmenté et étiqueté est présenté dans la figure 25 ci-dessous, il s'agit dans ce cas du mot français *chipi*.

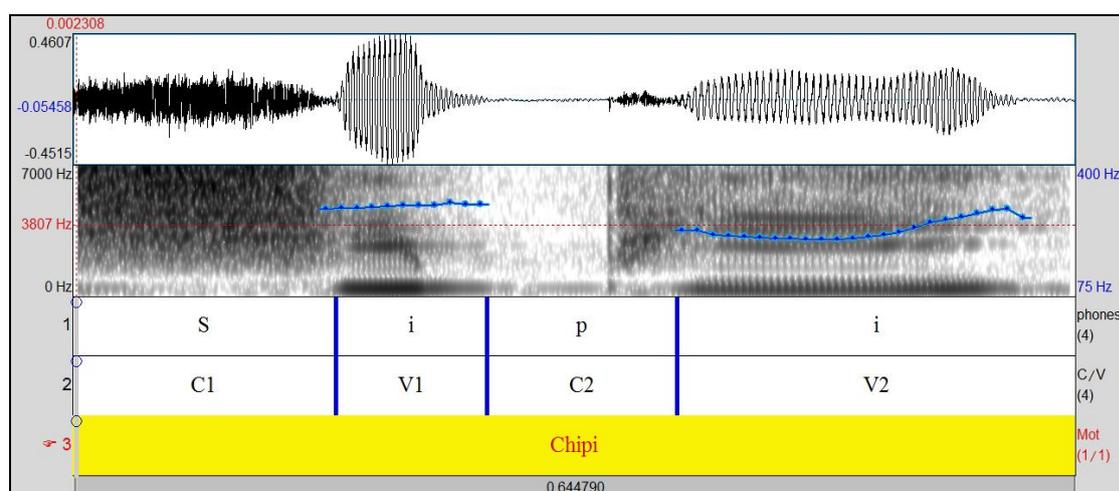


Figure 25 – Représentation du mot *chipi* produit par la locutrice FIFR, segmenté et étiqueté dans le logiciel Praat. La partie supérieure de la figure comprend l'onde sonore, la partie médiane le spectrogramme et la courbe de F0, et la partie inférieure les trois paliers d'étiquetage (phones, C/V, mot).

La procédure a donc été répétée pour les 216 mots de l'étude (27*8 locuteurs). Une fois ces données sauvegardées, j'ai pu aisément créer des fichiers sons individuels pour chacun des phones, ce qui permet de procéder de manière plus aisée à l'analyse acoustique, notamment pour les opérations automatisées par des scripts.

2.2 Mesure du F0 et de la durée de chaque mot

Les premières mesures réalisées concernent le *F0 moyen* et la *durée* des mots entiers. Ces données ont été recueillies en créant un fichier *pitch* pour chaque mot, puis en utilisant les fonctions *Get total duration* et *Get mean* à partir de ce fichier. Ces opérations ont été automatisées à l'aide d'un script *Praat* programmé à cet effet⁸⁹. Une vérification était réalisée *a posteriori* sur les données obtenues pour le F0 moyen : les items pour lesquels les résultats paraissaient non-cohérents ont fait l'objet d'une nouvelle mesure manuelle.

Dans un deuxième temps, *la plage de variation de F0* a été mesurée sur chaque mot. Pour cela, c'est à nouveau les fichiers *pitch* qui ont été utilisés. Une fois la courbe de F0 affichée via la commande *Edit*, il convenait d'ouvrir la fenêtre *pitch info* qui contient, entre autres, les informations relatives à la plage de variation (voir figure 26 ci-dessous).

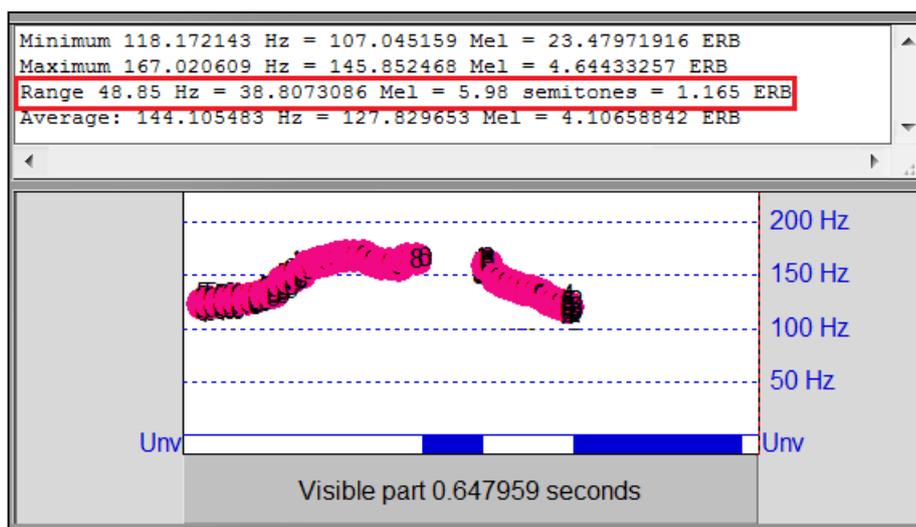


Figure 26 – Courbe de F0 du mot *zipi* prononcé par le locuteur H2FR, obtenue dans Praat via un fichier Pitch. Dans la partie supérieure figure des données calculées automatiquement par Praat, dont la plage de variation, encadrée ici en rouge, et exprimée en différentes unités de mesure.

Les valeurs ont été recueillies en *hertz*, mais également en *demi-tons*, qui est ici une échelle plus pertinente car elle respecte la perception des fréquences par l'oreille humaine, comme nous l'avons vu dans la section 3.2 du premier chapitre.

⁸⁹ Le script est visible en annexe 5, et téléchargeable en ligne.

2.3 Analyse acoustique des consonnes

Pour l'ensemble des consonnes initiales⁹⁰, *la durée* (également mesurée sur C2), le *pic spectral*, qui correspond à la fréquence pour laquelle la concentration d'énergie est la plus forte⁹¹, et le *centre de gravité*⁹² ont été analysés. Le *F0 moyen* a également été mesuré pour les consonnes voisées. Il est important d'évoquer ici le cas particulier des consonnes occlusives de l'anglais : lorsqu'elles se trouvent en position initiale de mot, devant une voyelle et sous l'accent lexical, les consonnes /t/ et /k/ se réalisent au niveau phonétique [t^h] et [k^h] (occlusives sourdes aspirées), tandis que les phonèmes /d/ et /g/ sont généralement produits comme des occlusives dévoisées non-aspirées [d̥] et [g̥], le voisement débutant uniquement au début de la voyelle suivante (voir Caramazza et al., 1973 ; Lin & Wong, 2011). Les F0 moyens n'ont donc pas pu être mesurés sur les occlusives initiales dites « voisées » des mots anglais.

Pour obtenir les durées et le F0 moyens, les commandes *Get total duration* et *Get mean* ont à nouveau été utilisées. Le pic spectral a quant à lui été détecté grâce à la fonction *Get frequency of maximum* appliquée sur l'objet *LTAS*⁹³ associé au fichier son. Enfin, le centre de gravité a pu être mesuré *via* la commande *Get centre of gravity* appliquée sur l'objet *Spectrum* correspondant au fichier son analysé. Toutes ces opérations ont été automatisées avec un script programmé pour l'occasion⁹⁴. Une vérification était réalisée *a posteriori* sur les mesures obtenues avec le script pour le pic spectral et le F0 moyen : les items n'ayant pas pu être mesurés ou ceux dont les résultats me paraissaient non-cohérents ont été à nouveau mesurés manuellement.

Une autre mesure a été effectuée uniquement sur les consonnes occlusives : le *voice onset time*, ou *délai d'établissement du voisement*. Ce dernier a été calculé manuellement pour chaque consonne, à partir du spectrogramme, comme l'illustre la figure 27 ci-après.

⁹⁰ Pour rappel, ces consonnes en position C1 sont les suivantes : [s], [z], [ʃ], [ʒ], [t], [d], [k] et [g].

⁹¹ Pour éviter que cette mesure ne corresponde au F0 ou à ses premiers harmoniques sur les consonnes voisées, le pic spectral est ici recherché à partir de 700 Hz et dans les fréquences au-delà.

⁹² Point d'équilibre de la distribution de l'énergie dans le spectre fréquentiel. Cette mesure donne donc une indication sur la position moyenne des fréquences dans le spectre.

⁹³ *Long Term Average Spectrum*.

⁹⁴ Le script est visible en annexe 5 et téléchargeable en ligne.

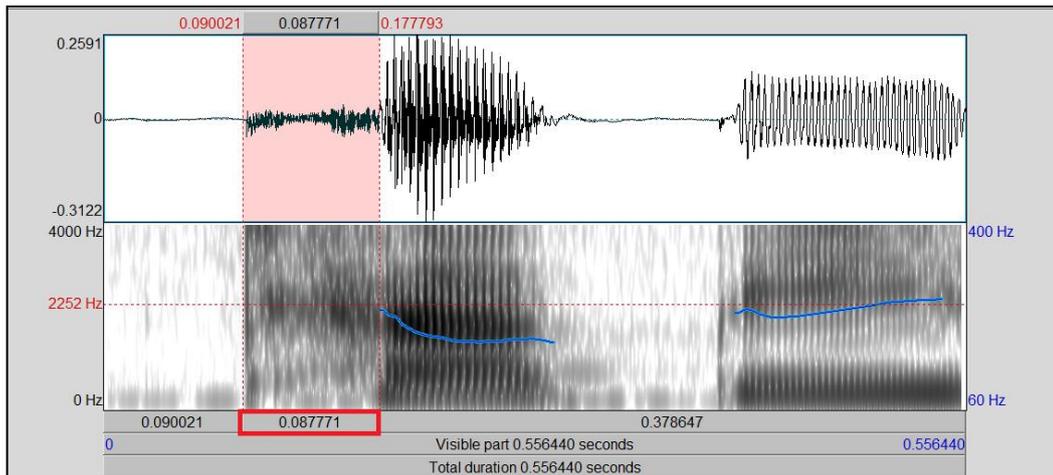


Figure 27 – Représentation du mot *tappy*, produit par la locutrice *FIAN*, dans le logiciel Praat. La partie supérieure de la figure comprend l'onde sonore, la partie inférieure le spectrogramme et la courbe de *F0* (en bleu). La sélection, visible en rose sur l'onde sonore, débute au relâchement de la consonne occlusive et se termine au démarrage du voisement. Sa durée (exprimée en s), qui constitue le *VOT*, est encadrée en rouge dans la partie inférieure gauche de la figure (ici 87 ms).

Le moment du relâchement de l'occlusive ainsi que le début du voisement ont donc dû être localisés sur les spectrogrammes, le *VOT* correspondant à l'espacement temporel entre le premier repère et le second (si le voisement débute après le relâchement de la consonne, le *VOT* est positif, et inversement).

2.4 Analyse acoustique des voyelles

Les voyelles présentes en position *V1*⁹⁵ ont fait l'objet d'une analyse acoustique multidimensionnelle. Tout d'abord, le *F0 moyen* et la *durée* ont été recueillis de la même manière que pour les mots : création d'un fichier *pitch* associé pour chaque fichier son contenant la voyelle, puis utilisation des fonctions *Get total duration* et *Get mean* à partir de ce fichier. Ces opérations ont été automatisées à l'aide du même script que celui utilisé pour les mots entiers.

Ensuite, les *valeurs des trois premiers formants vocaliques* (*F1*, *F2*, *F3*) ont été mesurées manuellement dans le logiciel *Praat*, en s'appuyant conjointement sur les

⁹⁵ C'est-à-dire [i], [a], [u] pour les mots français, [i:], [æ], [u:] pour les mots anglais. Seule la durée a également été mesurée sur la voyelle *V2* [i], afin d'établir la répartition temporelle C/V à l'intérieur des mots dissyllabiques. Cette voyelle étant non-accentuée dans les mots anglais et redondante avec la voyelle [i] déjà présente en *V1* pour les mots français, son analyse approfondie n'aurait donc pas apporté d'informations pertinentes complémentaires.

spectrogrammes, la détection automatique des formants et les spectres. Les relevés ont tous été réalisés sur une portion centrale et stable de la voyelle, afin de limiter l'influence de la coarticulation. Cette opération est visible dans la figure 28 ci-dessous.

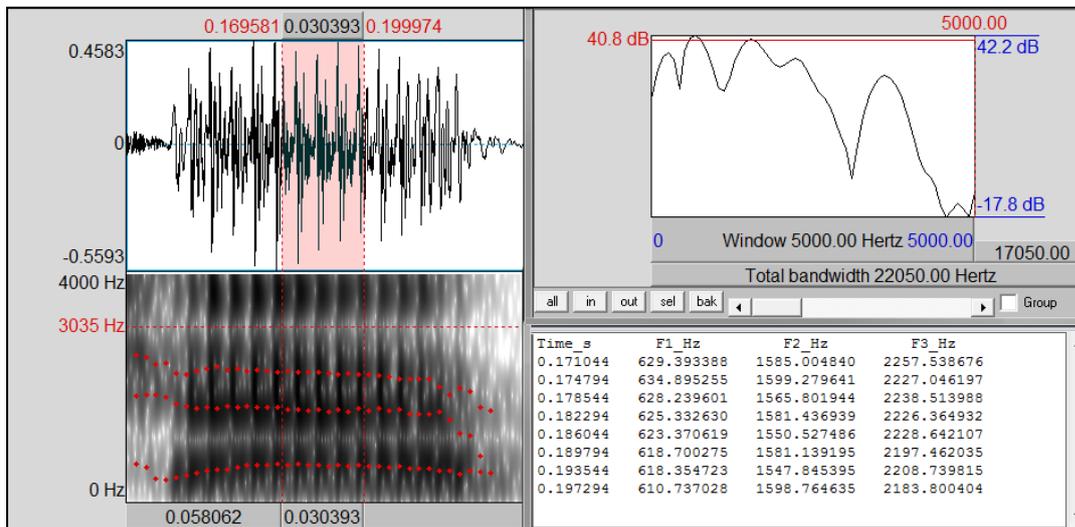


Figure 28 – Représentation de la voyelle [æ] du mot shappy produit par le locuteur HIAN, dans le logiciel Praat. La partie supérieure gauche de la figure comprend l'onde sonore, la partie inférieure gauche le spectrogramme avec en pointillés rouges les formants détectés par Praat. Le spectre correspondant à la partie sélectionnée (en rose sur l'onde sonore) est visible en haut à droite et les valeurs de F1, F2, F3 détectées par Praat pour cette même partie de la voyelle figurent dans la fenêtre se trouvant en bas à droite.

J'ai choisi de ne pas automatiser cette procédure car des erreurs d'estimation sont souvent commises par le logiciel *Praat*, en particulier pour les voyelles focales telles que le [u], pour lequel les deux premiers formants sont extrêmement proches.

Enfin, j'ai souhaité réaliser une analyse de la *qualité de voix*⁹⁶ des locuteurs. La mesure acoustique actuellement considérée comme la plus fiable pour ce type d'étude (voir notamment Klatt & Klatt, 1990 ; Shrivastav & Sapienza, 2003) est l'intensité relative du premier harmonique (H1) par rapport au deuxième (H2). Selon Klatt & Klatt (1990) et Gordon & Ladefoged (2001), la force relative du premier harmonique est corrélée positivement au *quotient d'ouverture glottique* (*glottal open quotient*). Concrètement, plus le premier harmonique est puissant, plus la voix est soufflée (ou *breathy*) ; plus ce dernier est faible (en l'occurrence, plus faible que le deuxième harmonique), plus la voix peut-être considérée comme craquée (ou *creaky*).

⁹⁶ Terme que j'emploie ici comme synonyme de *type de phonation*. Voir sections 2.1 et 3.5 du chapitre 1.

Selon Simpson (2012), il convient d'éviter de procéder à ce type de mesure sur des voyelles précédées ou suivies d'une consonne nasale, ou sur des voyelles prononcées isolément : dans ces contextes, la présence possible de traces de nasalité biaiserait les résultats en renforçant le premier harmonique des locutrices et le second harmonique des locuteurs. Le corpus est ici composé exclusivement de mots dissyllabiques ne contenant aucune consonne nasale, ce qui permet d'écarter ce biais. D'autre part, cette mesure *H1-H2* doit impérativement être réalisée sur une voyelle ouverte (voir Klatt & Klaat, 1990), sans quoi le premier formant viendrait biaiser fortement les résultats⁹⁷.

Compte tenu de ces éléments, l'analyse a porté uniquement sur les neuf occurrences de la voyelle [a] pour les locuteurs francophones et sur les neuf occurrences de la voyelle [æ] pour les anglophones. Pour ce faire, une sélection d'une durée de cinq périodes⁹⁸ était réalisée sur une partie centrale de chaque voyelle, afin d'afficher le spectre correspondant (voir figure 29, ci-dessous).

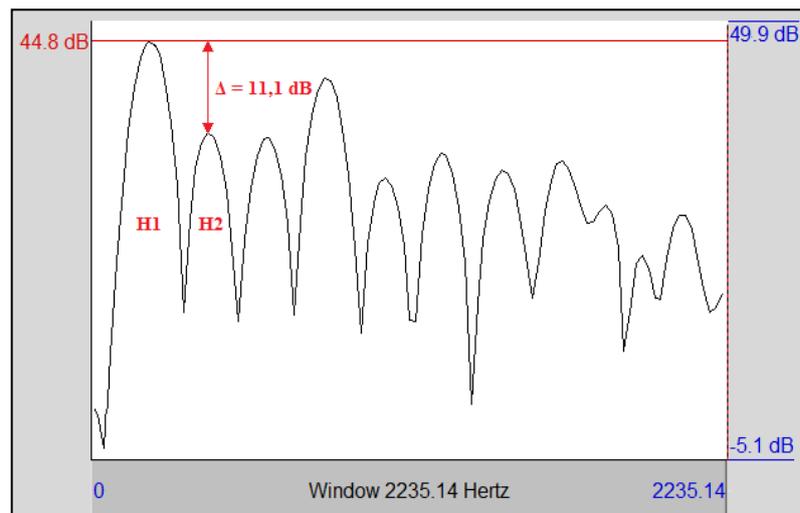


Figure 29 – Spectre de la voyelle [æ] du mot *appy* produit par la locutrice F2AN, obtenu dans le logiciel Praat. Les deux premiers harmoniques (H1 et H2) sont visibles sur la partie gauche du spectre. La différence d'intensité (en dB) entre ces deux items est ici mentionnée en rouge.

⁹⁷ Sur une voyelle fermée, par exemple, le premier et/ou le deuxième harmonique entrera(en)t en résonance à cause du premier formant, situé aux alentours des 300 Hz.

⁹⁸ Correspondant donc approximativement à 5 cycles laryngés. Une telle durée permet de faire ressortir clairement les harmoniques sur le spectre.

Comme l'illustre la figure 29, la différence entre l'intensité de H1 et l'intensité de H2 (exprimée en dB), était ensuite calculée grâce aux informations fournies par le spectre.

L'analyse des données a maintenant été décrite. La section suivante est dédiée aux résultats de cette analyse acoustique.

3. RESULTATS

Dans cette section sont présentés les résultats de l'analyse acoustique pour les hommes et les femmes francophones parisiens et anglophones américains du nord-est des Etats-Unis. Une première partie est consacrée aux résultats bruts, une seconde est dédiée à l'analyse acoustique de ces données.

3.1 Résultats bruts

Les résultats de l'analyse acoustique pour les consonnes, les voyelles et les mots entiers sont présentés ici de manière brute et synthétique⁹⁹.

3.1.1 Analyse acoustique des consonnes

L'analyse acoustique des consonnes initiales a été multidimensionnelle. Dans un premier temps, les résultats pour le pic spectral et le centre de gravité seront donnés, suivis de ceux correspondant au *VOT*, au F0 et enfin à la durée. Les valeurs obtenues pour les locuteurs francophones et celles obtenues pour les anglophones seront présentées alternativement.

3.1.1.1 Pic spectral et centre de gravité

La valeur moyenne¹⁰⁰ du *pic spectral* de chacune des consonnes initiales de mot chez les locuteurs francophones apparaît dans le tableau 3 ci-après. Ces résultats sont donnés, d'une part, pour chaque locuteur et d'autre part, pour les deux femmes et les deux hommes réunis.

⁹⁹ Les fichiers contenant les résultats détaillés (item par item) sont disponibles en ligne : voir annexe 8.

¹⁰⁰ Calculée sur l'ensemble des occurrences de chaque consonne.

Tableau 3 – Pic spectral moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs francophones.

<i>Segment</i>	Pic spectral moyen (Hz)						
	F1FR	F2FR	Moy. loc.¹⁰¹ F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[t]	4617	4217	4417	4383	3950	4167	4292
[k]	3250	3750	3500	2550	2583	2567	3033
Moy. occl.¹⁰² sourdes	3933	3983	3958	3467	3267	3367	3663
[d]	3483	3150	3317	3083	3150	3117	3217
[g]	2783	2617	2700	2450	2383	2417	2558
Moy. occl. voisées	3133	2883	3008	2767	2767	2767	2888
Moy. toutes occl.	3533	3433	3483	3117	3017	3067	3275
[s]	6717	9450	8083	7417	5450	6433	7258
[ʃ]	3283	2783	3033	2850	2417	2633	2833
Moy. fric.¹⁰³ sourdes	5000	6117	5558	5133	3933	4533	5046
[z]	8150	9783	8967	7750	7250	7500	8233
[ʒ]	3450	2317	2883	2750	3050	2900	2892
Moy. fric. voisées	5800	6050	5925	5250	5150	5200	5563
Moy. toutes fric.	5400	6083	5742	5192	4542	4867	5304
Moy. toutes consonnes	4467	4758	4613	4154	3779	3967	4290

On remarque que les pics spectraux des consonnes *se situent dans des fréquences globalement plus élevées chez les locutrices*, de l'ordre de 16 % en moyenne. Ces différences sont plus fortes sur les fricatives sourdes (23 %) et dans une moindre mesure sur les occlusives sourdes (18 %), que sur les consonnes voisées (14 % pour les fricatives, 9 % pour les occlusives). La différence maximale est atteinte sur le son [s], pour lequel la fréquence moyenne du pic spectral est plus élevée de 26 % chez les femmes.

Les résultats correspondants pour les locuteurs anglophones figurent dans le tableau 4 ci-après.

¹⁰¹ L'abréviation « moy. » correspond à « moyenne », « loc. » correspond à « locuteurs ».

¹⁰² L'abréviation « occl. » renvoie à « occlusive ».

¹⁰³ L'abréviation « fric. » correspond à « fricative ».

Tableau 4 – Pic spectral moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs anglophones.

Segment	Pic spectral moyen (Hz)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[t ^h]	2817	4983	3900	4883	1983	3433	3667
[k ^h]	2350	3150	2750	1883	2450	2167	2458
Moy. occl. aspirées	2583	4067	3325	3383	2217	2800	3063
[d]	2450	3250	2850	2550	2417	2483	2667
[g]	2517	2917	2717	2183	2517	2350	2533
Moy. occl. non-aspirées	2483	3083	2783	2367	2467	2417	2600
Moy. toutes occl.	2533	3575	3054	2875	2342	2608	2831
[s]	7117	8783	7950	6283	9217	7750	7850
[ʃ]	5617	3800	4708	2950	4583	3767	4238
Moy. fric. sourdes	6367	6292	6329	4617	6900	5758	6044
[z]	7617	7583	7600	7350	8683	8017	7808
[ʒ]	3983	3583	3783	2417	2517	2467	3125
Moy. fric. voisées	5800	5583	5692	4883	5600	5242	5467
Moy. toutes fric.	6083	5938	6010	4750	6250	5500	5755
Moy. toutes consonnes	4308	4756	4532	3813	4296	4054	4293

Tout comme chez les francophones, les pics spectraux moyens des consonnes initiales produites par les auditeurs anglophones *se trouvent dans des fréquences globalement plus élevées chez les locutrices*. Cette différence semble cependant de moindre ampleur que pour les francophones : elle est d'environ 12 %, toutes consonnes confondues. L'écart femmes / hommes est maximal pour la catégorie des occlusives sourdes aspirées (19 % en moyenne et près de 27 % pour la consonne [k^h]), et minimal sur les fricatives voisées (environ 9 % en moyenne) : notons que le pic spectral est même en moyenne légèrement plus élevé chez les locuteurs masculins sur la consonne [z].

Les résultats de l'analyse du *centre de gravité spectral* des consonnes initiales de mot chez les locuteurs francophones sont présentés dans le tableau 5. Ces moyennes sont calculées sur l'ensemble des occurrences (n=3) de chaque consonne.

Tableau 5 – Centre de gravité moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs francophones.

<i>Segment</i>	Centre de gravité moyen (Hz)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[t]	4921	4570	4745	4139	2588	3363	4054
[k]	3132	4595	3864	3173	2583	2878	3371
Moy. occl. sourdes	4026	4583	4305	3656	2586	3121	3713
[d]	616	777	697	423	336	379	538
[g]	574	939	757	413	318	366	561
Moy. occl. voisées	595	858	727	418	327	372	550
Moy. toutes occl.	2311	2721	2516	2037	1456	1747	2131
[s]	8124	9268	8696	8037	7539	7788	8242
[ʃ]	4976	5031	5004	4623	4831	4727	4866
Moy. fric. sourdes	6550	7150	6850	6330	6185	6258	6554
[z]	4478	4649	4564	2156	1936	2046	3305
[ʒ]	1934	1205	1569	1578	1086	1332	1451
Moy. fric. voisées	3206	2927	3066	1867	1511	1689	2378
Moy. toutes fric.	4878	5038	4958	4099	3848	3973	4466
Moy. toutes consonnes	3594	3879	3737	3068	2652	2860	3298

A l'instar du pic spectral, le centre de gravité des consonnes se trouve en moyenne dans des fréquences plus élevées chez les locuteurs féminins : + 31 % par rapport aux chiffres des locuteurs masculins, toutes consonnes confondues. Ces différences sont particulièrement élevées sur les occlusives sourdes (+ 38 %), de même que sur les fricatives voisées (+ 82 %), avec un maximum de + 123 % pour le son [z].

Les résultats correspondants pour les locuteurs anglophones sont présentés dans le tableau 6.

Tableau 6 – Centre de gravité moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs anglophones.

<i>Segment</i>	Centre de gravité moyen (Hz)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[t ^h]	3062	5833	4447	4540	2434	3487	3967
[k ^h]	2099	3381	2740	1278	2625	1951	2346
Moy. occl. aspirées	2581	4607	3594	2909	2529	2719	3156
[d]	3293	5477	4385	2526	2283	2405	3395
[g]	2925	3185	3055	2178	2105	2142	2598
Moy. occl. non-aspirées	3109	4331	3720	2352	2194	2273	2997
Moy. toutes occl.	2845	4469	3657	2630	2362	2496	3076
[s]	7372	9851	8612	7291	8922	8107	8359
[ʃ]	5463	5791	5627	4423	5062	4743	5185
Moy. fric. sourdes	6417	7821	7119	5857	6992	6425	6772
[z]	2005	4028	3017	1082	1976	1529	2273
[ʒ]	744	737	741	501	673	587	664
Moy. fric. voisées	1375	2383	1879	791	1325	1058	1468
Moy. toutes fric.	3896	5102	4499	3324	4158	3741	4120
Moy. toutes consonnes	3370	4785	4078	2977	3260	3119	3598

La tendance observée sur les locuteurs francophones se retrouve également chez leurs homologues anglophones : *le centre de gravité moyen des consonnes se trouve dans des fréquences globalement plus élevées chez les locutrices*. Cet écart est d'environ 31 % par rapport aux chiffres des locuteurs masculins, toutes consonnes confondues. La différence reste relativement élevée sur tous les types de consonnes et est particulièrement marquée sur les fricatives voisées : + 76 % en moyenne. L'écart maximum est d'ailleurs atteint sur la consonne [z], comme chez les locuteurs francophones, avec une différence de 97 %. D'autre part, notons qu'il existe une importante différence entre les deux locutrices : le centre de gravité moyen des consonnes produites par F2AN se situe dans des fréquences nettement plus élevées que celui des consonnes prononcées par la locutrice F1AN (+ 42 % en moyenne).

Le tableau 7, ci-dessous, présente les fréquences de résonance moyennes des consonnes (pic spectral et centre de gravité) pour les quatre locuteurs francophones en fonction de la taille des individus. Cette dernière mesure donne une indication fiable de la longueur relative des conduits vocaux¹⁰⁴, dont on sait qu'elle conditionne les fréquences de résonance : toutes choses égales par ailleurs, plus un conduit vocal est court et plus les fréquences de résonance seront globalement élevées (voir Fant, 1966, 1970).

Tableau 7 – Taille des locuteurs francophones (cm) et fréquences de résonance moyennes (pic spectral et centre de gravité, en Hz) des consonnes initiales.

	Taille des locuteurs et fréquences de résonance des consonnes			
	F1FR	F2FR	H1FR	H2FR
<i>Taille (cm)</i>	162	175	183	182
Pic spectral moy. toutes consonnes (Hz)	4467	4758	4154	3779
Centre de gravité moy. toutes consonnes (Hz)	3594	3879	3068	2652

On constate qu'en dépit d'une taille inférieure et donc d'un conduit vocal probablement plus court, les fréquences de résonance de la locutrice F1FR, toutes consonnes confondues, sont globalement plus basses que celles de la locutrice F2FR. Chez les deux locuteurs masculins, malgré une taille équivalente, leurs fréquences de résonance moyennes sont sensiblement différentes : elles sont en moyenne plus élevées chez le locuteur H1FR.

Les fréquences de résonance moyennes des consonnes initiales produites par les locuteurs anglophones en fonction de la taille des individus apparaissent dans le tableau 8, ci-après.

¹⁰⁴ La longueur du conduit vocal est en effet très fortement corrélée à la taille globale de l'individu, y compris au sein d'une population du même genre (voir Fitch & Giedd, 1999 ; Cherng et al., 2002).

Tableau 8 – Taille des locuteurs anglophones (cm) et fréquences de résonance moyennes (pic spectral et centre de gravité, en Hz) des consonnes initiales.

	Taille des locuteurs et fréquences de résonance des consonnes			
	F1AN	F2AN	H1AN	H2AN
<i>Taille (cm)</i>	174	169	178	185
Pic spectral moy. toutes consonnes (Hz)	4308	4756	3813	4296
Centre de gravité moy. toutes consonnes (Hz)	3370	4785	2977	3260

Pour les locutrices anglophones, il apparaît que les fréquences de résonance de F1AN sont en moyenne plus basses que celles de F2AN. Cela semble cohérent avec la longueur relative estimée des conduits vocaux : F2AN étant de taille inférieure, elle dispose vraisemblablement d'un conduit vocal plus court. En revanche, chez les locuteurs masculins, c'est H1AN, dont la taille est sensiblement inférieure à celle de H2AN, qui possède les fréquences de résonance moyennes les plus basses.

3.1.1.2 VOT

Les *voice onset times* moyens des locuteurs et locutrices francophones sur les occlusives en initiale de mots sont donnés dans le tableau 9 ci-après. Ces chiffres sont calculés sur les trois occurrences de chacune des consonnes.

Tableau 9 – VOT moyen (ms) de chacune des consonnes initiales pour les quatre locuteurs francophones.

Segment	VOT moyen (ms)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[t]	60	58	59	35	39	37	48
[k]	55	60	58	49	43	46	52
Moy. occl. sourdes	58	59	58	42	41	41	50
[d]	-91	-72	-81	-60	-57	-59	-70
[g]	-79	-63	-71	-60	-57	-59	-66
Moy. occl. voisées	-85	-67	-76	-60	-57	-59	-68

On remarque que *le délai d'établissement du voisement est en moyenne plus élevé chez les locutrices pour les occlusives sourdes (+ 41 %)*. C'est en revanche le contraire *pour les occlusives voisées : le VOT est globalement plus court chez les femmes que chez les hommes francophones, de l'ordre de 29 %*. Le contraste entre occlusives sourdes et voisées semble donc être plus fortement accentué chez les locutrices : 134 ms en moyenne contre 100 ms pour les hommes.

Les délais d'établissement du voisement des locuteurs anglophones pour les consonnes initiales de mot apparaissent dans le tableau 10, ci-après.

Tableau 10 – VOT moyen (ms) de chacune des consonnes initiales pour les quatre locuteurs anglophones.

<i>Segment</i>	VOT moyen (ms)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[t ^h]	86	75	80	54	64	59	69
[k ^h]	90	83	86	54	75	65	76
Moy. occl. aspirées	88	79	83	54	69	62	73
[d]	26	27	27	15	17	16	21
[g]	28	28	28	26	24	25	27
Moy. occl. non-aspirée	27	28	27	20	20	20	24

Chez les locuteurs anglophones *le VOT est en moyenne plus élevé chez les locutrices que chez les locuteurs sur les occlusives aspirées (+ 34 %)*, mais également sur *les occlusives non-aspirées (+ 35 %)*. Néanmoins, le contraste entre occlusives aspirées et non-aspirées demeure au final globalement plus marqué chez les locutrices ($\Delta=56$ ms) que chez les locuteurs ($\Delta=42$ ms).

3.1.1.3 F0

Le F0 des locuteurs francophones, pour chacune des consonnes voisées présentes en initiale de mot, est donné dans le tableau 11. Ces chiffres moyens sont calculés sur les trois occurrences de chaque consonne.

Tableau 11 – F0 moyen (Hz) de chacune des consonnes initiales voisées pour les quatre locuteurs francophones.

Segment	F0 moyen (Hz)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[d]	170	170	170	108	122	115	143
[g]	175	173	174	111	124	118	146
Moy. occl.	173	171	172	110	123	116	144
[z]	173	171	172	104	121	112	142
[ʒ]	169	177	173	103	122	113	143
Moy. fric.	171	174	172	103	122	112	142
Moy. toutes consonnes	172	172	172	106	122	114	143

Fort logiquement, la fréquence fondamentale des locutrices francophones est supérieure à celle des locuteurs sur l'ensemble des consonnes : en moyenne, elle est plus élevée de 51 %. Cet écart reste relativement stable d'une consonne à l'autre.

Pour les locuteurs anglophones, rappelons que seules les fricatives [z] et [ʒ] sont effectivement voisées¹⁰⁵ : les mesures de fréquence fondamentale sur les consonnes initiales n'ont donc pu porter que sur les productions de ces deux items. Les résultats correspondants apparaissent dans le tableau 12 ci-après.

Tableau 12 – F0 moyen (Hz) de chacune des consonnes initiales voisées pour les quatre locuteurs anglophones.

Segment	F0 moyen (Hz)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[z]	161	202	181	107	89	98	140
[ʒ]	166	203	185	110	92	101	143
Moy. toutes consonnes	163	203	183	109	91	100	141

Comme pour les locuteurs francophones, la fréquence fondamentale moyenne des anglophones est globalement plus élevée chez les femmes que chez les hommes sur les consonnes voisées : elle est en moyenne supérieure de 83 %. Ici encore, l'écart ne varie pratiquement pas d'une consonne à l'autre.

¹⁰⁵ Voir section 2.3.

3.1.1.4 Durée

Les durées moyennes des consonnes initiales de mot pour les locuteurs francophones sont présentées dans le tableau 13. Ces moyennes sont calculées sur l'ensemble des occurrences (n=3) de chaque consonne.

Tableau 13 – Durée moyenne (ms) de chacune des consonnes initiales pour les quatre locuteurs francophones.

Segment	Durée moyenne (ms)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[t]	127	124	126	93	89	91	108
[k]	125	112	119	106	94	100	109
Moy. occl. sourdes	126	118	122	100	91	96	109
[d]	137	107	122	83	78	81	101
[g]	119	102	110	90	81	86	98
Moy. occl. voisées	128	104	116	87	80	83	100
Moy. toutes occl.	127	111	119	93	85	89	104
[s]	193	195	194	128	144	136	165
[ʃ]	164	167	165	136	135	136	150
Moy. fric. sourdes	179	181	180	132	140	136	158
[z]	122	155	138	85	108	97	118
[ʒ]	122	125	123	95	112	104	113
Moy. fric. voisées	122	140	131	90	110	100	115
Moy. toutes fric.	150	160	155	111	125	118	137
Moy. toutes consonnes	139	136	137	102	105	104	120

On remarque que chez les francophones, *la durée des consonnes est globalement plus élevée chez les locutrices*, d'environ 32 % en moyenne. Cette différence est minimale pour les occlusives sourdes (+ 27 %), et maximale pour les occlusives voisées (+ 40 %).

Les résultats correspondants pour les locuteurs et locutrices anglophones figurent dans le tableau 14.

Tableau 14 – Durée moyenne (ms) de chacune des consonnes initiales pour les quatre locuteurs anglophones.

Segment	Durée moyenne (ms)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[t ^h]	166	159	163	115	126	121	142
[k ^h]	173	163	168	117	142	130	149
Moy. occl. aspirées	169	161	165	116	134	125	145
[d]	95	135	115	69	116	93	104
[g]	90	121	106	86	101	93	100
Moy. occl. non-aspirées	92	128	110	77	109	93	102
Moy. toutes occl.	131	145	138	97	121	109	123
[s]	190	192	191	148	181	165	178
[ʃ]	215	187	201	151	167	159	180
Moy. fric. sourdes	202	190	196	149	174	162	179
[z]	201	151	176	110	143	127	151
[ʒ]	178	154	166	126	127	127	146
Moy. fric. voisées	189	153	171	118	135	127	149
Moy. toutes fric.	196	171	184	134	155	144	164
Moy. toutes consonnes	163	158	161	115	138	127	144

Pour les anglophones également, *les consonnes produites par les locutrices sont globalement plus longues que celles produites par les locuteurs*. Cette différence atteint 27 % en moyenne sur l'ensemble des consonnes initiales. Les écarts maximums s'observent sur les occlusives aspirées (+ 32 %) et les fricatives voisées (+ 35 %). A l'inverse, ce sont les occlusives non-aspirées qui présentent la plus faible différence femmes / hommes (+ 18 %).

3.1.2 Analyse acoustique des voyelles

Les résultats de l'analyse acoustique des voyelles seront exposés en quatre temps. La première sous-section est consacrée aux valeurs des formants vocaliques (F1, F2 et F3), la seconde à la qualité de voix, la troisième à la fréquence fondamentale et la dernière à la durée. Les chiffres obtenus pour les locuteurs francophones et ceux obtenus pour les anglophones seront là encore présentés alternativement.

3.1.2.1 Formants vocaliques (F1, F2, F3)

Les valeurs moyennes des trois premiers formants des voyelles en syllabe initiale pour les locuteurs francophones sont présentées dans le tableau 15 ci-après. Les écarts-types ainsi que les différences femmes-hommes (sous forme de ratio), y figurent également. Ces valeurs ont été calculées sur la base des neuf occurrences de chaque voyelle par locuteur.

Tableau 15 – Fréquence moyenne (Hz), écart-type (Hz) et ratio femmes / hommes pour les F1, F2 et F3 des voyelles en syllabe initiale, chez les locuteurs francophones.

Voyelle	Formant	Fréquence moyenne du formant (Hz)							Ratio F/H
		F1FR	F2FR	Tous loc. F	H1FR	H2FR	Tous loc. H	Moy. tous loc.	
[i]	F1	307	291	299	304	305	305	302	0,98
	σ	11	7	12	5	10	7	10	
	F2	2678	2572	2625	2088	1992	2040	2332	1,29
	σ	31	54	69	61	76	83	306	
	F3	3594	3539	3567	3089	3096	3092	3329	1,15
	σ	50	42	53	44	89	68	248	
[a]	F1	601	631	616	544	623	584	600	1,05
	σ	81	84	81	34	37	54	70	
	F2	1886	1939	1913	1668	1494	1581	1747	1,21
	σ	75	81	81	88	65	117	195	
	F3	2954	2979	2967	2627	2578	2603	2785	1,14
	σ	42	59	51	84	167	131	209	
[u]	F1	315	306	310	314	314	314	312	0,99
	σ	6	3	6	8	4	6	7	
	F2	1220	1029	1124	1111	874	992	1058	1,13
	σ	146	116	161	120	38	149	167	
	F3	2876	2372	2624	2286	2069	2178	2401	1,20
	σ	73	150	283	83	34	127	313	

Plusieurs éléments intéressants ressortent à la lecture de ce tableau. Tout d'abord, les valeurs des deuxièmes et troisièmes formants sont globalement plus élevées chez les femmes, pour chacune des voyelles testées. Ces différences sont maximales pour le F2 de la voyelle [i] : sa fréquence moyenne est plus élevée de 29 % chez les locutrices. A l'inverse, les différences inter-genres sont extrêmement faibles pour le premier formant. De plus, ces faibles écarts ne vont pas systématiquement dans le même sens : pour [i] et [u], la fréquence du F1 est très légèrement inférieure chez les femmes (respectivement - 2 et - 1 %), alors qu'elle est supérieure pour la voyelle [a] (+ 5 %).

Une représentation des résultats pour F1 et F2 est visible ci-après sous forme de triangles vocaliques¹⁰⁶ (figure 30). Ce type de graphique, qui présente la valeur de F1 sur l'axe vertical (croissante du haut vers le bas), et celle de F2 sur l'axe horizontal, permet de rendre compte approximativement du lieu d'articulation de la voyelle : la fréquence du premier formant étant proportionnelle au degré d'aperture, celle du deuxième formant proportionnelle au degré d'antériorité, tout en subissant une diminution avec l'arrondissement des lèvres (voir Stevens, 1999). Les valeurs de F1 et F2 sont exprimées en hertz dans le triangle de gauche et en barks dans le triangle de droite. Cette dernière unité permet de rendre compte des différences effectivement perçues par l'auditeur. La formule utilisée pour la conversion hertz / barks est celle de Traummüller (1990)¹⁰⁷.

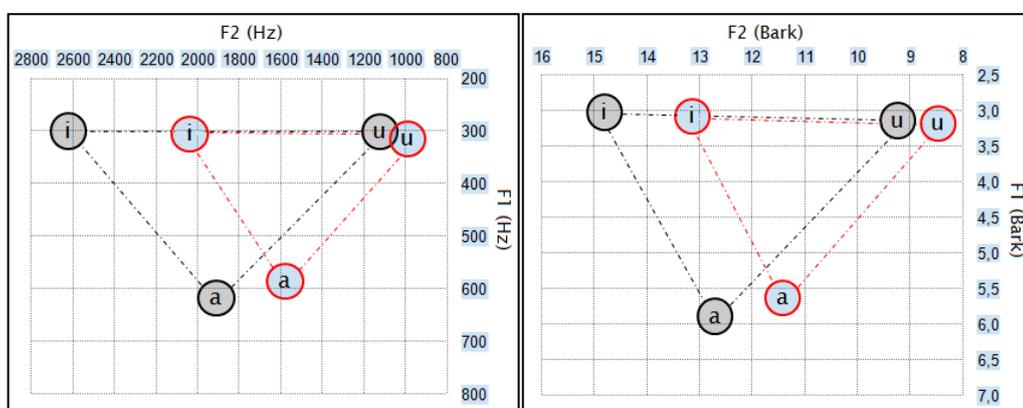


Figure 30 – Triangles vocaliques moyens des hommes (n=2), en trait rouge, et des femmes francophones (n=2), en trait noir, pour les trois voyelles cardinales [i],[a] et [u]. Les valeurs sont données en hertz dans la partie gauche, en barks dans la partie droite. Le premier formant est représenté sur l'axe vertical et le deuxième formant sur l'axe horizontal.

La dissymétrie des différences hommes-femmes sur le plan du deuxième formant, évoquée précédemment, apparaît ici clairement : c'est bien sur les voyelles antérieures qu'elle est la plus élevée, et tout particulièrement sur le [i].

Les valeurs moyennes des trois premiers formants des voyelles en syllabe initiale pour les locuteurs anglophones figurent dans le tableau 16 ci-après. Tout comme pour les sujets francophones, les valeurs ont été calculées sur la base de neuf occurrences de chaque voyelle par locuteur et les écarts-types ainsi que les différences femmes-hommes (sous forme de ratio) sont également précisés.

¹⁰⁶ Les figures représentant des triangles vocaliques dans cette étude ont été obtenues en modifiant des modèles contenus dans le logiciel SaRP 3.0 (voir Nikolov et al., 2011).

¹⁰⁷ $z = 26,81/(1+(1960 / f))-0,53$ (avec f = fréquence en hertz)

Tableau 16 – Fréquence moyenne (Hz), écart-type (Hz) et ratio femmes/hommes pour les F1, F2 et F3 des voyelles en syllabe initiale et accentuée, chez les locuteurs anglophones.

		Fréquence moyenne du formant (Hz)							
Voyelle	Formant	F1AN	F2AN	Tous loc. F	H1AN	H2AN	Tous loc. H	Moy. tous loc.	Ratio F/H
[i:]	F1	325	328	327	303	305	304	315	1,08
	σ	8	10	9	7	2	5	14	
	F2	2554	2572	2563	2007	2387	2197	2380	1,17
	σ	72	33	55	53	65	203	237	
	F3	3526	3382	3454	2763	3057	2910	3182	1,19
	σ	70	120	120	55	54	160	309	
[æ]	F1	828	792	810	670	688	679	745	1,19
	σ	35	17	33	14	11	15	71	
	F2	1822	1813	1817	1615	1701	1658	1738	1,10
	σ	30	31	30	31	16	50	90	
	F3	2734	2657	2695	2344	2558	2451	2573	1,10
	σ	76	41	71	76	56	128	160	
[u:]	F1	339	359	349	314	326	320	335	1,09
	σ	5	5	11	8	9	11	18	
	F2	1487	1642	1565	1221	1235	1228	1396	1,27
	σ	275	157	231	198	183	185	268	
	F3	2564	2627	2596	2241	2347	2294	2445	1,13
	σ	79	68	78	64	46	77	171	

Chez les anglophones, les valeurs moyennes des trois premiers formants vocaliques (F1, F2, F3) sont globalement plus élevées chez les locutrices. Contrairement aux francophones, F1 semble donc varier également en fonction du genre : cela est tout particulièrement vrai sur la voyelle ouverte [æ], pour laquelle le premier formant est en moyenne dans des fréquences 19 % plus élevées chez les femmes. Autre élément important pour la comparaison inter-langues, ce n'est pas ici le deuxième formant du [i:] qui présente les différences maximales (+ 17 %), mais le deuxième formant de la voyelle [u:] (+ 27 %), à l'inverse des francophones, chez qui le F2 de la voyelle fermée postérieure [u] présentait de faibles différences inter-genres. Notons par ailleurs que c'est sur ce même F2 de la voyelle [u:] que les écarts-types sont les plus élevés chez les anglophones, ce qui dénote des variations inter- et intra-locuteurs importantes. Une représentation des résultats pour les deux premiers formants est visible ci-après sous forme de triangles vocaliques (figure 31).

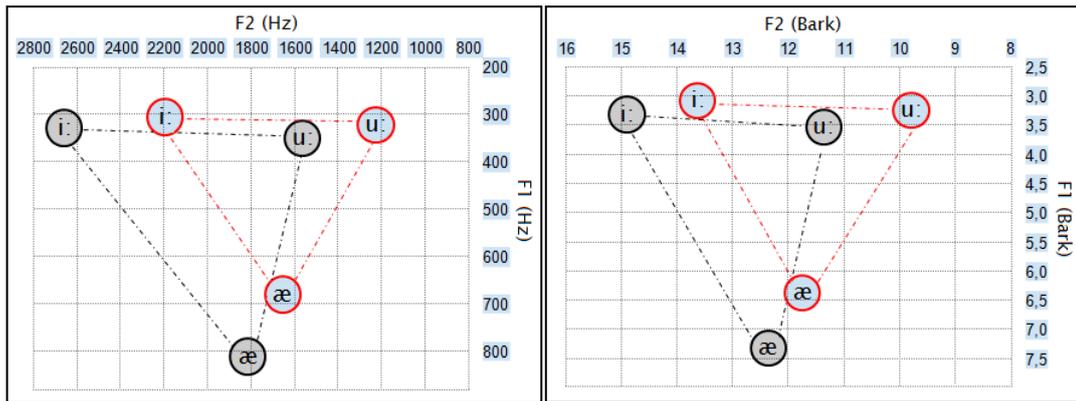


Figure 31 – Triangles vocaliques moyens des hommes (n=2), en trait rouge, et des femmes anglophones (n=2), en trait noir, pour les trois voyelles [i:], [æ] et [u:]. Les valeurs sont données en hertz dans la partie gauche, en barks dans la partie droite. Le premier formant est représenté sur l'axe vertical et le deuxième formant sur l'axe horizontal.

On observe ici de manière claire les importantes différences femmes-hommes pour le F1 de la voyelle [æ], ainsi que pour le F2 des trois voyelles cardinales, y compris le [u:]. Cette dernière voyelle présente même la différence inter-genre la plus forte pour le deuxième formant, lorsque les valeurs sont exprimées en barks.

Le tableau 17, ci-après, présente les valeurs formantiques moyennes (toutes voyelles confondues) pour les quatre locuteurs francophones en fonction de la taille des individus¹⁰⁸.

Tableau 17 – Taille des locuteurs francophones et valeurs moyennes des trois premiers formants vocaliques, toutes voyelles ([i], [a], [u]) confondues.

	Taille des locuteurs et formants vocaliques			
	F1FR	F2FR	H1FR	H2FR
Taille (cm)	162	175	183	182
F1 moy. toutes voyelles (Hz)	408	409	388	414
F2 moy. toutes voyelles (Hz)	1928	1847	1622	1453
F3 moy. toutes voyelles (Hz)	3141	2963	2667	2581

¹⁰⁸ Pour rappel, la longueur d'un conduit vocal est très fortement corrélée à la taille de l'individu (voir Cherng et al., 2002).

Contrairement aux constatations faites sur les consonnes¹⁰⁹, la locutrice F1FR présente ici des fréquences de résonance moyennes légèrement plus élevées que son homologue F2FR. Cela est en adéquation avec la longueur relative estimée de leurs conduits vocaux : la taille de F2FR est en effet nettement supérieure à celle de F1FR. Chez les locuteurs masculins, on constate, à l’instar des consonnes, que malgré la taille équivalente des deux individus, les valeurs formantiques moyennes de H1FR sont globalement plus élevées que celles de H2FR, en particulier pour le deuxième formant.

Les fréquences de résonance moyennes des consonnes initiales produites par les locuteurs anglophones en fonction de la taille des individus apparaissent dans le tableau 18, ci-après.

Tableau 18 – Taille des locuteurs anglophones et valeurs moyennes des trois premiers formants vocaliques, toutes voyelles ([i:], [æ], [u:]) confondues.

	Taille des locuteurs et formants vocaliques			
	F1AN	F2AN	H1AN	H2AN
<i>Taille (cm)</i>	174	169	178	185
F1 moy. toutes voyelles (Hz)	497	493	429	440
F2 moy. toutes voyelles (Hz)	1954	2009	1614	1774
F3 moy. toutes voyelles (Hz)	2941	2889	2449	2654

Pour les locuteurs anglophones, on remarque que les valeurs formantiques moyennes des deux femmes sont proches, ce qui concorde avec l’assez faible différence de taille qu’il existe entre ces deux locutrices. Chez leurs homologues masculins, en revanche, on constate une différence assez importante : les valeurs formantiques de H1AN sont globalement plus basses que celles de H2AN, et ce malgré un conduit vocal probablement plus court, ce qui rejoint en tout point les observations faites sur ces deux locuteurs pour les fréquences de résonance des consonnes.

¹⁰⁹ Voir sous-section 3.1.1.1.

3.1.2.2 Qualité de voix

Comme nous l'avons vu précédemment¹¹⁰, la mesure acoustique la plus fiable pour estimer la qualité de voix est la différence entre l'intensité (en dB) du premier harmonique et celle du deuxième harmonique (H1-H2) : un chiffre élevé correspond à une voix *breathy*, et un chiffre négatif indique lui une voix *creaky* (ou *fry*). L'intensité relative de H1 par rapport à H2 pour les quatre locuteurs francophones est visible dans le tableau 19 ci-après. Ces mesures ont été réalisées sur les neuf occurrences de la voyelle [a].

Tableau 19 – Différence d'intensité entre le premier et le deuxième harmonique pour les neuf occurrences de la voyelle [a] produites par les locuteurs francophones.

Segment et contexte	Différence d'intensité H1-H2 (dB)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[<u>a</u> pi]	8,9	9,8	9,8	0,7	0,1	0,4	4,9
[<u>ʃ</u> api]	2,6	6,8	4,7	-3,6	-0,2	-1,9	1,4
[<u>d</u> api]	5,6	2,7	4,2	-3,8	-1,2	-2,5	0,8
[<u>g</u> api]	6,0	2,3	4,2	-2,4	-0,2	-1,3	1,4
[<u>ʒ</u> api]	2,3	8,6	5,5	-2,7	-0,4	-1,6	2,0
[<u>k</u> api]	2,9	0,9	1,9	-2,0	-1,1	-1,6	0,2
[<u>s</u> api]	4,4	5,1	4,8	-1,4	-1,2	-1,3	1,7
[<u>t</u> api]	0,9	4,3	2,6	-2,2	-3,4	-2,8	-0,1
[<u>z</u> api]	2,2	4,4	3,3	-2,3	0,5	-0,9	1,2
Moy. [a]	4,0	5,0	4,5	-2,2	-0,8	-1,5	1,5

On note une forte différence femmes-hommes : le premier harmonique est nettement plus fort chez les locutrices que chez les locuteurs, cette différence atteint en moyenne 6 dB. Cela suggère que le quotient d'ouverture glottique¹¹¹ est supérieur chez les locutrices, dont les voix peuvent donc être considérées comme plus *breathy* que celle des hommes francophones. Le rapport H1-H2 moyen chez les hommes est légèrement négatif¹¹², ce qui est caractéristique d'une voix *modale* ou faiblement *creaky*.

¹¹⁰ Voir section 2.4.

¹¹¹ Temps pendant lequel les plis vocaux sont en position ouverte durant un cycle laryngé.

¹¹² Cela signifie que H2 est plus intense que H1.

Les données équivalentes pour les locuteurs anglophones sont présentées dans le tableau 20, ci-après. Ces mesures ont été effectuées sur les neuf occurrences de la voyelle [æ] pour chacun des locuteurs.

Tableau 20 – Différence d'intensité entre le premier et le deuxième harmonique pour les neuf occurrences de la voyelle [æ] produites par les locuteurs anglophones.

Segment et contexte	Différence d'intensité H1-H2 (dB)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[æpi]	4,1	11,1	7,6	-5,6	-3,2	-4,4	1,6
[jæpi]	4,2	10,2	7,2	-5,1	-0,7	-2,9	2,2
[dæpi]	4,0	2,4	3,2	-4,2	-5,2	-4,7	-0,8
[gæpi]	2,4	5,2	3,8	-5,3	-4,7	-5,0	-0,6
[zæpi]	3,4	6,1	4,8	-3,8	1,8	-1,0	1,9
[k ^h æpi]	3,8	5,5	4,7	-3,9	1,9	-1,0	1,8
[sæpi]	2,4	4,0	3,2	-4,1	-0,1	-2,1	0,6
[t ^h æpi]	5,0	4,6	4,8	-4,8	-1,5	-3,2	0,8
[zæpi]	6,3	5,4	5,9	-6,7	-2,3	-4,5	0,7
Moy. [æ]	4,0	6,1	5,0	-4,8	-1,6	-3,2	0,9

Chez les anglophones le premier harmonique est en moyenne nettement plus intense chez les femmes. Cette différence est sensiblement plus importante que chez les francophones : elle de 8,2 dB en moyenne. Les voix des locutrices sont donc clairement plus *breathy* que celles des hommes. De plus, le rapport H1-H2 est ici fortement négatif (en moyenne -3,2 dB) chez les locuteurs masculins, cela indique une qualité de voix *creaky*. Notons que c'est d'ailleurs sur les voix d'hommes que la différence avec les francophones est la plus importante : les voix des locuteurs anglophones semblent globalement plus *creaky* de celles des hommes francophones.

3.1.2.3 F0

Le F0 moyen des locuteurs francophones, pour chacune des voyelles présentes en première syllabe de mot, est donné dans le tableau 21. Ces chiffres sont calculés sur les neuf occurrences de chaque voyelle.

Tableau 21 – F0 moyen (Hz) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs francophones.

Segment	F0 moyen (Hz)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[i]	286	251	268	148	166	157	213
[a]	275	224	250	140	147	144	197
[u]	297	257	277	158	166	162	219
Moy. toutes voyelles	286	244	265	149	160	154	210

On constate que *la fréquence fondamentale des locutrices francophones est supérieure à celle des locuteurs sur l'ensemble des voyelles* : en moyenne, elle est plus élevée de 72 %, soit un écart nettement plus important que celui observé précédemment sur les consonnes voisées (+ 51 %). Ces différences restent relativement stables d'une voyelle à l'autre.

Les F0 moyens des locuteurs anglophones pour les trois voyelles [i:], [æ] et [u:] apparaissent dans le tableau 22, ci-après.

Tableau 22 – F0 moyen (Hz) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs anglophones.

Segment	F0 moyen (Hz)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[i:]	251	236	244	142	110	126	185
[æ]	209	212	211	130	96	113	162
[u:]	255	234	245	143	112	128	186
Moy. toutes voyelles	238	227	233	138	106	122	178

On retrouve ici les mêmes tendances que chez les locuteurs francophones : *la fréquence fondamentale moyenne des voyelles est globalement plus élevée chez les femmes anglophones que chez les hommes*, de l'ordre de 91 %. Cet écart est supérieur à celui observé précédemment sur les consonnes initiales, et ne varie que faiblement d'une voyelle à une autre.

3.1.2.4 Durée

Les durées moyennes des voyelles [i], [a] et [u]¹¹³, produites par les locuteurs et locutrices francophones sont données dans le tableau 23, ci-dessous. Ces chiffres sont calculés sur la base des neuf occurrences de chaque voyelle¹¹⁴.

Tableau 23 – Durée moyenne (ms) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs francophones.

Segment	Durée moyenne (ms)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[i]	104	78	91	89	109	99	95
[a]	102	98	100	98	113	106	103
[u]	110	72	91	83	112	97	94
Moy.toutes voyelles	105	83	94	90	111	101	97

Contrairement aux consonnes initiales, on constate *une très faible différence hommes-femmes sur le plan de la durée de ces voyelles* : elle est en moyenne supérieure d'environ 7 % pour les locuteurs masculins. Cette différence demeure relativement constante d'une voyelle à l'autre.

Les résultats correspondants pour les locutrices et locuteurs anglophones figurent dans le tableau 24.

Tableau 24 – Durée moyenne (ms) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs anglophones.

Segment	Durée moyenne (ms)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[i:]	106	118	112	107	105	106	109
[æ]	139	147	143	125	119	122	132
[u:]	98	109	104	106	91	99	101
Moy.toutes voyelles	115	125	120	112	105	109	114

¹¹³ Présentes en syllabe initiale de mot.

¹¹⁴ Voir annexe 8 pour le téléchargement des résultats détaillés pour chaque occurrence.

Là encore, les résultats pour les anglophones sont très similaires à ceux obtenus chez les francophones. On note en effet *une faible différence inter-genres sur la durée des voyelles* : ces dernières sont en moyenne 10 % plus longues chez les locutrices. Cette différence, bien qu'allant dans le même sens, est de moindre ampleur que celle observée sur les consonnes initiales. D'autre part, on remarque que l'écart femmes-hommes est nettement plus élevé sur la voyelle [æ] (+ 17 % en moyenne) que sur les deux autres voyelles.

3.1.3 Analyse acoustique des mots dissyllabiques entiers

Cette dernière sous-partie comprend les résultats de l'analyse acoustique pour les mots dissyllabiques entiers. Le F0 moyen, la plage de variation de F0, la durée moyenne ainsi que la répartition temporelle consonnes / voyelles seront présentés. Comme précédemment, les résultats seront donnés successivement pour les francophones parisiens et les anglophones américains du nord-est des Etats-Unis.

3.1.3.1 F0 moyen

Les fréquences fondamentales moyennes des quatre locuteurs francophones parisiens sur les 27 (pseudo-)mots apparaissent dans le tableau 25 ci-après.

Tableau 25 – F0 moyen (Hz) pour chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs francophones.

Mot	F0 moyen (Hz)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[api]	232	231	232	138	139	139	185
[ipi]	245	240	243	127	157	142	192
[upi]	255	224	240	122	155	139	189
[ʒapi]	254	251	253	135	156	146	199
[ʃipi]	259	224	242	125	157	141	191
[ʒupi]	269	218	244	128	170	149	196
[dapi]	220	195	208	133	134	134	171
[dipi]	236	218	227	134	143	139	183
[dupi]	240	225	233	128	136	132	182
[gapi]	219	201	210	121	138	130	170
[gipi]	235	224	230	142	149	146	188
[gupi]	244	225	235	142	151	147	191
[ʒapi]	225	216	221	125	143	134	177
[ʒipi]	230	205	218	137	146	142	180
[ʒupi]	233	241	237	140	142	141	189
[kapi]	256	209	233	136	149	143	188
[kipi]	266	238	252	145	151	148	200
[kupi]	269	231	250	138	151	145	197
[sapi]	259	254	257	136	141	139	198
[sipi]	252	243	248	136	158	147	197
[supi]	285	243	264	143	147	145	205
[tapi]	249	205	227	125	138	132	179
[tipi]	267	236	252	128	153	141	196
[tupi]	274	238	256	148	162	155	206
[zapi]	217	210	214	129	127	128	171
[zipi]	237	217	227	137	142	140	183
[zupi]	237	218	228	140	148	144	186
Moy. tous mots	247	225	236	134	148	141	188

Comme cela était prévisible, la fréquence fondamentale moyenne des locutrices est nettement supérieure à celle des locuteurs sur l'ensemble des 27 mots : en moyenne, elle est plus élevée de 67 %.

Les F0 moyens des quatre locuteurs anglophones sur les 27 (pseudo-)mots sont présentés dans le tableau 26 ci-après.

Tableau 26 – F0 moyen (Hz) pour chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs anglophones.

Mot	F0 moyen (Hz)						
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
[æpi]	224	235	230	132	97	115	172
[i:pi]	213	251	232	136	98	117	175
[u:pi]	229	242	236	154	105	130	183
[ʃæpi]	178	236	207	103	96	100	153
[ʃi:pi]	230	255	243	125	103	114	178
[ʃu:pi]	226	250	238	116	101	109	173
[dæpi]	201	233	217	102	98	100	159
[di:pi]	213	247	230	117	99	108	169
[du:pi]	224	243	234	126	95	111	172
[gæpi]	229	232	231	144	94	119	175
[gi:pi]	231	245	238	119	99	109	174
[gu:pi]	225	242	234	108	104	106	170
[zæpi]	179	226	203	118	97	108	155
[zi:pi]	207	230	219	133	99	116	167
[zu:pi]	191	229	210	117	98	108	159
[k ^h æpi]	238	238	238	116	104	110	174
[k ^h i:pi]	223	256	240	159	98	129	184
[k ^h u:pi]	249	247	248	142	97	120	184
[sæpi]	251	248	250	116	100	108	179
[si:pi]	208	252	230	133	104	119	174
[su:pi]	229	252	241	125	118	122	181
[t ^h æpi]	222	237	230	123	100	112	171
[t ^h i:pi]	239	249	244	130	118	124	184
[t ^h u:pi]	242	252	247	122	105	114	180
[zæpi]	187	220	204	115	96	106	155
[zi:pi]	196	234	215	121	96	109	162
[zu:pi]	205	238	222	114	96	105	163
Moy. tous mots	218	241	230	125	101	113	171

Sans surprise, la fréquence fondamentale moyenne des femmes anglophones est très largement supérieure à celle des hommes, sur chacun des 27 mots et pseudo-mots de cette étude. Cet écart atteint en moyenne + 104 %.

3.1.3.2 Plage de variation de F0

La plage de variation moyenne de F0¹¹⁵ à l'échelle du mot dissyllabique pour les locuteurs francophones figure dans le tableau 27, ci-après. Ces chiffres sont la moyenne des plages de variation observées sur chacun des 27 mots dissyllabiques utilisés dans

¹¹⁵ Pour rappel, cette mesure correspond à l'écart entre la fréquence fondamentale la plus basse et la fréquence fondamentale la plus élevée atteintes dans un domaine temporel donné, ici le mot dissyllabique.

cette étude¹¹⁶, ils sont exprimés en hertz, mais également en demi-tons : cette deuxième unité de mesure permet de rendre compte de la variation perçue par l'oreille humaine.

Tableau 27 – Plage de variation moyenne de F0 sur les 27 mots dissyllabiques prononcés par les locuteurs francophones, exprimée en hertz et en demi-tons.

Plage de variation moyenne de la fréquence fondamentale sur mot-dissyllabique							
<i>Unité de mesure</i>	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
Hertz (Hz)	110	96	103	50	49	50	77
Demi-ton (dt)	7,93	7,52	7,73	6,59	5,83	6,21	6,97

Lorsqu'elle est exprimée en hertz, la plage de variation moyenne des locutrices francophones est donc environ deux fois supérieure à celle des locuteurs masculins. L'écart diminue fortement lorsque ces chiffres sont convertis en demi-tons. Néanmoins, la plage de variation des femmes demeure sensiblement plus élevée que celle des hommes, d'environ 1,5 dt.

Les résultats équivalents pour les quatre locuteurs anglophones américains, sont présentés dans le tableau 28.

Tableau 28 – Plage de variation moyenne de F0 sur les 27 mots dissyllabiques prononcés par les locuteurs anglophones, exprimée en hertz et en demi-tons.

Plage de variation moyenne de la fréquence fondamentale sur mot-dissyllabique							
<i>Unité de mesure</i>	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	Moy. tous loc.
Hertz (Hz)	88	71	80	45	32	38	59
Demi-ton (dt)	6,78	5,11	5,95	6,54	5,79	6,17	6,06

Chez les anglophones, la plage de variation de F0 est en moyenne environ deux fois plus grande chez les locutrices lorsque l'unité de mesure utilisée est le hertz. En revanche, une fois ces chiffres convertis en demi-tons, la différence devient minime et s'inverse : elle est d'environ 0,2 dt, en faveur des locuteurs masculins.

¹¹⁶ Les résultats détaillés sont là encore disponibles en ligne : voir annexe 8.

3.1.3.3 Durée des mots entiers et répartition temporelle intra-mots

La durée des 27 mots et pseudo-mots produits par les quatre locuteurs francophones est visible dans le tableau 29 ci-dessous.

Tableau 29 – *Durée moyenne (ms) de chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs francophones.*

<i>Mot</i>	Durée (ms)						
	F1FR	F2FR	Moy. loc. F	H1FR	H2FR	Moy. loc. H	Moy. tous loc.
[api]	574	455	515	332	397	365	440
[ipi]	575	379	477	314	406	360	419
[upi]	553	426	490	380	413	397	443
[ʃapi]	601	488	545	430	536	483	514
[ʃipi]	647	502	575	406	576	491	533
[ʃupi]	660	510	585	446	554	500	543
[dapi]	741	560	651	416	472	444	547
[dipi]	694	502	598	377	462	420	509
[dupi]	735	527	631	365	467	416	524
[gapi]	736	497	617	423	508	466	541
[gipi]	735	560	648	331	406	369	508
[gupi]	691	465	578	372	462	417	498
[ʒapi]	611	544	578	429	526	478	528
[ʒipi]	590	589	590	366	517	442	516
[ʒupi]	585	513	549	363	537	450	500
[kapi]	687	483	585	403	502	453	519
[kipi]	690	475	583	394	480	437	510
[kupi]	662	469	566	375	584	480	523
[sapi]	718	547	633	408	527	468	550
[sipi]	685	567	626	411	543	477	552
[supi]	682	515	599	428	550	489	544
[tapi]	606	465	536	350	543	447	491
[tipi]	665	484	575	421	468	445	510
[tupi]	580	476	528	339	489	414	471
[zapi]	567	549	558	370	500	435	497
[zipi]	633	591	612	381	648	515	563
[zupi]	601	538	570	379	550	465	517
Moy. tous mots	648	507	577	386	505	445	511

On remarque que *la durée des mots est globalement plus élevée chez les femmes francophones que chez les hommes* : c'est le cas sur l'ensemble des 27 items. Cette différence atteint en moyenne + 30 %. Néanmoins, ces résultats sont dus principalement à la locutrice F1FR, qui a produit des mots d'une durée importante, et au locuteur H1FR, dont les mots sont particulièrement courts : les mots produits par les locuteurs F2FR et H2FR ont quant à eux une durée moyenne pratiquement identique (respectivement 507 et 505 ms).

La durée des 27 mots et pseudo-mots produits par les quatre locuteurs anglophones américains est visible dans le tableau 30 ci-dessous.

Tableau 30 – Durée moyenne (ms) de chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs anglophones.

Mot	Durée (ms)						Moy. tous loc.
	F1AN	F2AN	Moy. loc. F	H1AN	H2AN	Moy. loc. H	
[æpi]	450	535	493	372	419	396	444
[i:pi]	460	517	489	335	454	395	442
[u:pi]	403	537	470	367	438	403	436
[ʃæpi]	635	641	638	498	522	510	574
[ʃi:pi]	651	654	653	466	537	502	577
[ʃu:pi]	567	679	623	502	547	525	574
[dʌæpi]	506	641	574	407	509	458	516
[di:pi]	520	668	594	400	516	458	526
[du:pi]	500	635	568	405	508	457	512
[gæpi]	543	660	602	417	473	445	523
[gi:pi]	466	626	546	433	490	462	504
[gu:pi]	581	645	613	448	444	446	530
[zæpi]	593	648	621	489	519	504	562
[zi:pi]	597	652	625	426	500	463	544
[zu:pi]	663	653	658	460	456	458	558
[k ^h æpi]	603	641	622	428	466	447	535
[k ^h i:pi]	534	648	591	429	510	470	530
[k ^h u:pi]	507	598	553	435	476	456	504
[sæpi]	613	650	632	502	499	501	566
[si:pi]	540	731	636	488	562	525	580
[su:pi]	569	689	629	514	554	534	582
[t ^h æpi]	556	638	597	411	478	445	521
[t ^h i:pi]	579	702	641	434	479	457	549
[t ^h u:pi]	565	620	593	436	447	442	517
[zæpi]	646	670	658	472	513	493	575
[zi:pi]	681	690	686	437	585	511	598
[zu:pi]	575	655	615	442	478	460	538
Moy. tous mots	559	642	600	439	496	467	534

Chez les locuteurs anglophones, on constate également que la durée des mots est plus élevée chez les femmes que chez les hommes, et ce sur la totalité des 27 items. Cette différence est en moyenne de + 28 %.

Outre la durée globale des énoncés, j'ai souhaité par ailleurs observer la répartition temporelle consonnes / voyelles au sein des mots de type CVCV¹¹⁷. Cette répartition moyenne est visible dans la figure 32 ci-dessous, pour les quatre locuteurs francophones.

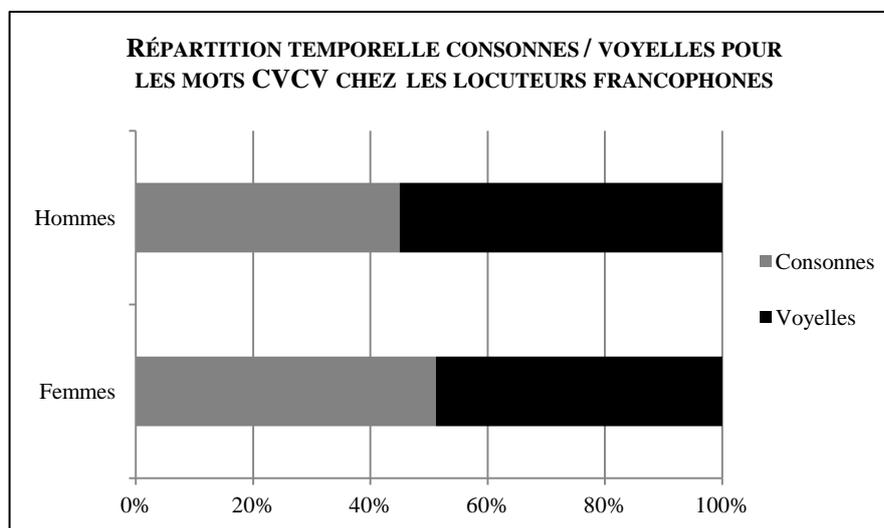


Figure 32 – Répartition temporelle consonnes / voyelles exprimée en pourcentages moyens, pour les 24 mots de type CVCV produits par les hommes (n=2) et les femmes francophones (n=2).

On remarque une assez faible différence femmes-hommes sur le plan de la répartition temporelle entre consonnes et voyelles : les consonnes sont proportionnellement plus longues chez les locutrices que chez les locuteurs francophones. Elles représentent 52 % de la durée totale moyenne des mot chez les femmes, contre 46 % chez les hommes.

Le graphique correspondant pour les quatre locuteurs anglophones américains est présenté ci-après (figure 33).

¹¹⁷ Les pseudo-mots [api], [ipi] et [upi], pour les francophones, et [æpi], [i:pi] et [u:pi] pour les anglophones ne sont donc pas inclus dans ces calculs.

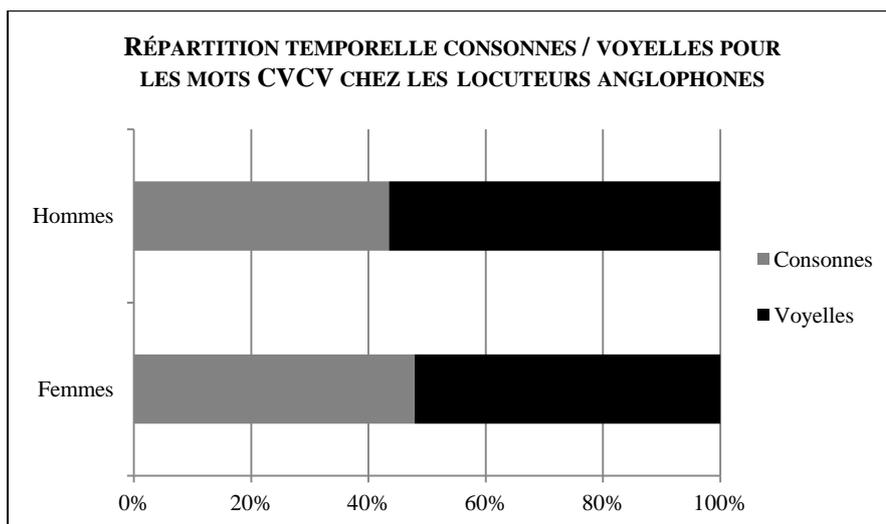


Figure 33 – Répartition temporelle consonnes / voyelles exprimée en pourcentages moyens, pour les 24 mots de type ‘CVCV produits par les hommes (n=2) et les femmes anglophones (n=2).

Pour les locuteurs anglophones, on constate également une légère différence inter-genre sur la répartition temporelle entre consonnes et voyelles dans les 24 mots de type CVCV : les consonnes sont proportionnellement plus longues chez les femmes que chez les hommes. Dans le détail, on constate que les consonnes représentent une portion légèrement plus grande du mot chez les locutrices : elles constituent en moyenne 48 % de la durée totale, contre environ 43 % pour les locuteurs.

Les résultats bruts ont maintenant été présentés. La section suivante est consacrée à leur analyse statistique.

3.2 Analyse statistique

Afin d'établir si les tendances décrites dans la section précédente sont significatives et de vérifier diverses interactions possibles entre les facteurs, j'ai procédé à une analyse statistique des résultats à l'aide du logiciel *StatView*¹¹⁸.

3.2.1 Analyse statistique des données obtenues pour les consonnes

Cette sous-section est dédiée à l'analyse statistique des données obtenues sur les consonnes initiales. Dans un premier temps, les résultats pour le pic spectral et le centre de gravité seront exposés, suivi de ceux correspondant au *VOT*, au *F0* et enfin à la durée. Les résultats seront donnés alternativement pour les locuteurs francophones et anglophones.

3.2.1.1 Pic spectral et centre de gravité

Afin de tester l'influence du genre du locuteur sur *le pic spectral* des consonnes initiales de mot produites par les quatre locuteurs francophones parisiens, une ANOVA à deux facteurs (« genre du locuteur » et « consonne ») a été effectuée sur ces données. Un graphique illustrant cette analyse est visible ci-après (figure 34).

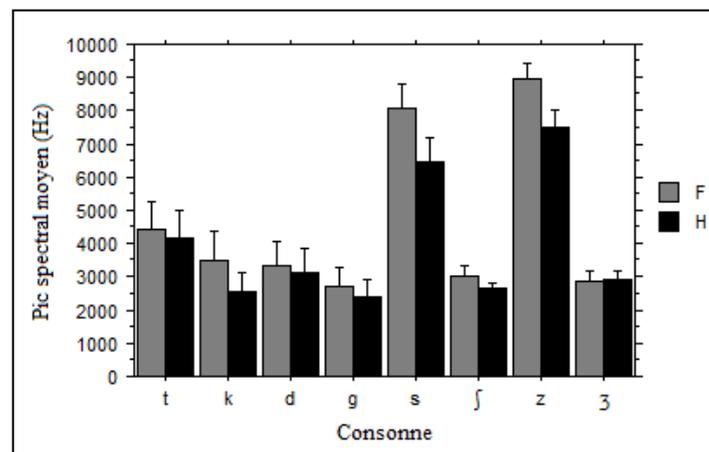


Figure 34 – Graphique des interactions représentant le pic spectral moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs francophones, avec les barres d'erreur-type.

¹¹⁸ *StatView* version 5.0.

Le résultat de cette ANOVA montre qu'il existe bien un effet global significatif du facteur « genre du locuteur » sur la fréquence du pic spectral des consonnes initiales : $F(1,80)=4,566$ avec $p<0,05$. *Le pic spectral des consonnes initiales se situe donc en moyenne dans des fréquences significativement plus élevées chez les locutrices francophones que chez leurs homologues masculins.*

Bien qu'il n'existe pas d'interaction significative entre les facteurs « genre du locuteur » et « consonne » ($F(7,80)=0,546$; $p>0,7$) pour les pics spectraux des consonnes initiales, j'ai souhaité conduire une ANOVA à un facteur (« genre du locuteur ») de manière individuelle pour chacune des consonnes. Ces analyses révèlent *une différence femmes-hommes significative pour le pic spectral de la consonne [z]* ($F(1,10)=5,110$; $p<0,05$) et non-significative pour les consonnes [t] ($F(1,10)=0,045$; $p>0,8$), [k] ($F(1,10)=0,811$; $p>0,3$), [d] ($F(1,10)=0,034$; $p>0,8$), [g] ($F(1,10)=0,136$; $p>0,7$), [s] ($F(1,10)=2,568$; $p>0,1$), [ʃ] ($F(1,10)=1,591$; $p>0,2$) et [ʒ] ($F(1,10)=0,002$; $p>0,9$).

Pour les quatre locuteurs anglophones, une ANOVA à deux facteurs (« genre du locuteur » et « consonne ») a été conduite sur le pic spectral des consonnes initiales de mot. Le graphique correspondant à cette analyse est présenté ci-après (figure 35).

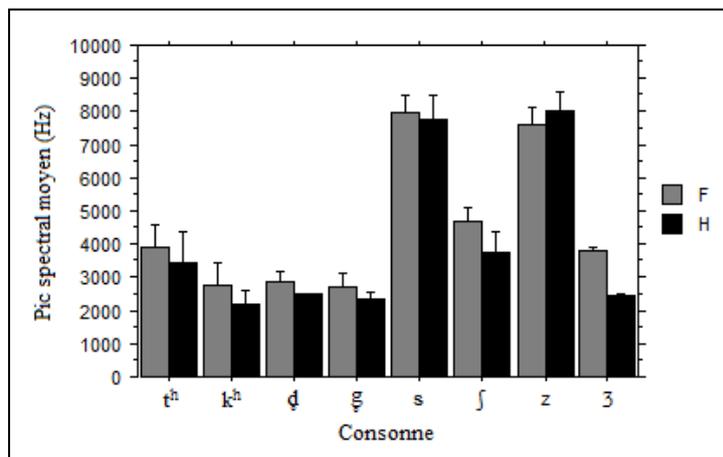


Figure 35 – Graphique des interactions représentant le pic spectral moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs anglophones, avec les barres d'erreur-type.

Le résultat de l'ANOVA indique une absence d'effet global significatif du facteur « genre du locuteur » sur le pic spectral des consonnes initiales produites par les locuteurs anglophones : $F(1,80)=3,373$ avec $p>0,05$. *Le pic spectral n'est donc pas, de manière globale, significativement plus élevé chez les locutrices que chez les locuteurs*

anglophones. Notons que cette absence de significativité s'explique en grande partie par une importante variabilité intra- et inter-sujets.

L'ANOVA à deux facteurs ne fait pas apparaître d'interaction significative entre les facteurs « genre du locuteur » et « consonne » ($F(7,80)=0,483$; $p>0,8$). J'ai néanmoins réalisé, comme pour les locuteurs francophones, une ANOVA à un facteur (« genre du locuteur ») de manière individuelle pour chacune des consonnes initiales. Ces analyses mettent en évidence *une différence femmes-hommes significative pour le pic spectral de la consonne [ʒ]* ($F(1,10)=99,696$; $p<0,0001$) et non-significative pour les consonnes [t^h] ($F(1,10)=0,696$; $p>0,6$), [k^h] ($F(1,10)=0,508$; $p>0,4$), [d] ($F(1,10)=1,217$; $p>0,2$), [g] ($F(1,10)=0,699$; $p>0,4$), [s] ($F(1,10)=0,051$; $p>0,8$), [ʃ] ($F(1,10)=1,557$; $p>0,2$) et [z] ($F(1,10)=0,614$; $p>0,6$).

Pour tester les différences inter-genres sur *le centre de gravité spectral* des consonnes initiales de mot prononcées par les quatre locuteurs francophones, une ANOVA à deux facteurs (« genre du locuteur » et « consonne ») a été réalisée sur ces données. Un graphique illustrant cette analyse est visible ci-dessous (figure 36).

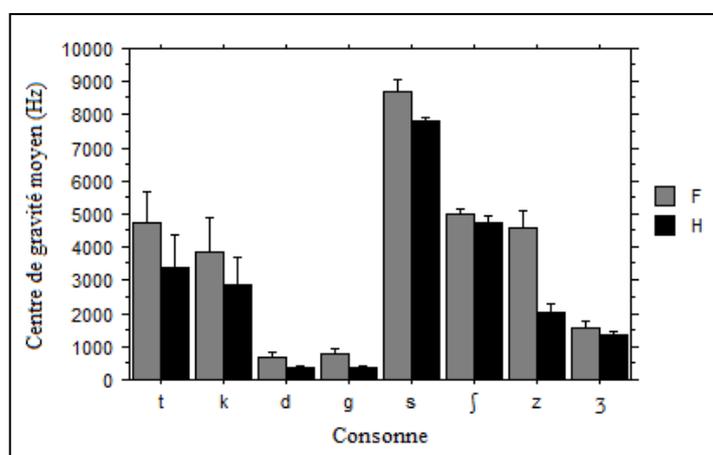


Figure 36 – Graphique des interactions représentant le centre de gravité moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs francophones, avec les barres d'erreur-type.

Le résultat de ce test indique qu'il existe un effet global très significatif du facteur « genre du locuteur » sur la fréquence du centre de gravité des consonnes initiales de mot produites par les francophones : $F(1,80)=11,501$ avec $p<0,01$. *Le centre de gravité des consonnes initiales se situe donc globalement dans des fréquences significativement plus élevées chez les femmes francophones que chez leurs homologues masculins.*

Il n'existe pas, chez les locuteurs francophones, d'interaction significative entre les facteurs « genre du locuteur » et « consonne » ($F(7,80)=1,143$; $p>0,3$) pour le centre de gravité des consonnes initiales. J'ai cependant souhaité conduire une ANOVA à un facteur (« genre du locuteur ») de manière individuelle pour chacune des consonnes. Ces analyses révèlent *une différence femmes-hommes significative pour le centre gravité des consonnes [d] ($F(1,10)=4,989$; $p<0,05$), [s] ($F(1,10)=6,168$; $p<0,05$) et [z] ($F(1,10)=16,614$; $p<0,01$)* mais non-significative pour les consonnes [t] ($F(1,10)=0,995$; $p>0,3$), [k] ($F(1,10)=0,580$; $p>0,4$), [g] ($F(1,10)=4,539$; $p>0,05$), [ʃ] ($F(1,10)=1,217$; $p>0,2$) et [ʒ] ($F(1,10)=1,079$; $p>0,3$).

Pour les locuteurs anglophones américains ($n=4$), une ANOVA à deux facteurs (« genre du locuteur » et « consonne ») a été effectuée sur le centre de gravité des consonnes initiales de mot. Le graphique correspondant à cette analyse est présenté ci-après (figure 37).

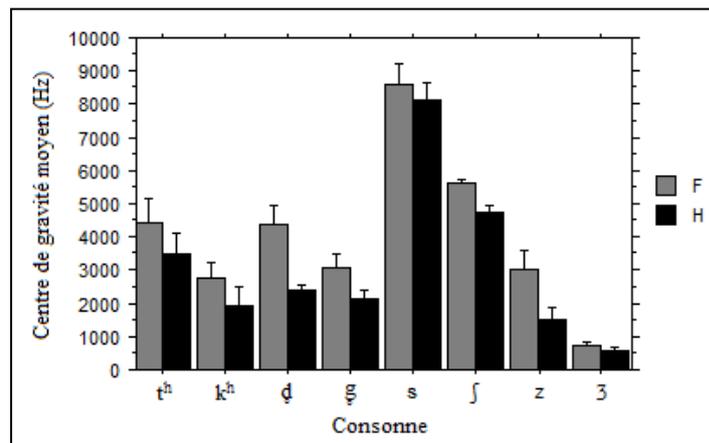


Figure 37 – Graphique des interactions représentant le centre de gravité moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs anglophones, avec les barres d'erreur-type.

Le résultat de l'ANOVA montre qu'il existe chez les anglophones un effet global très significatif du facteur « genre du locuteur » sur le centre de gravité spectral des consonnes initiales de mot : $F(1,80)=18,863$ avec $p<0,0001$. A l'instar des locuteurs francophones, *le centre de gravité spectral des consonnes initiales se situe donc globalement dans des fréquences significativement plus élevées chez les femmes anglophones que chez leurs homologues masculins.*

L'ANOVA ne met pas en évidence d'interaction significative entre les facteurs « genre du locuteur » et « consonne » ($F(7,80)=0,811$; $p>0,5$), pour le centre de gravité des consonnes initiales. J'ai néanmoins réalisé, comme pour les locuteurs francophones, une ANOVA à un facteur (« genre du locuteur ») de manière individuelle pour chacune des consonnes initiales. Ces analyses mettent en évidence *une différence femmes-hommes significative pour le centre de gravité des consonnes [d] ($F(1,10)=12,04$; $p<0,01$), [f] ($F(1,10)=12,818$; $p<0,01$) et [z] ($F(1,10)=5,348$; $p<0,05$), mais non-significative pour les consonnes [t^h] ($F(1,10)=1,029$; $p>0,3$), [k^h] ($F(1,10)=1,207$; $p>0,2$), [g] ($F(1,10)=3,480$; $p>0,05$), [s] ($F(1,10)=0,389$; $p>0,5$) et [ʒ] ($F(1,10)=1,771$; $p>0,2$).*

Taille des locuteurs et fréquences de résonance des consonnes initiales

Outre les différences inter-genres, plusieurs analyses ont enfin été effectuées dans les deux langues pour établir *la potentielle significativité des différences inter-locuteurs*, en particulier entre ceux du même genre. Il s'agit de quatre ANOVA à deux facteurs, « locuteur » et « consonne », menées successivement sur le pic spectral des consonnes initiales produites par les francophones, puis par les anglophones, ainsi que sur le centre de gravité de ces mêmes consonnes, chez les francophones puis chez les anglophones. Au delà de l'effet global du facteur « locuteur », les résultats du test PLSD de Fisher pour les différents locuteurs pris deux à deux seront riches en informations pertinentes, à mettre en regard avec la taille des locuteurs¹¹⁹.

Chez les francophones, l'effet global du facteur « locuteur » n'est pas significatif pour le pic spectral des consonnes initiales ($F=1,79$; $p>0,1$), il est en revanche significatif pour leur centre de gravité ($F=3,924$; $p<0,02$). Le PLSD de Fisher révèle des différences significatives entre F1FR et H2FR ($p<0,02$), F2FR et H1FR ($p<0,05$) et F2FR et H2FR ($p<0,01$) pour le centre de gravité. *Il n'existe donc, pour les locuteurs francophones, aucune différence significative entre les locuteurs d'un même genre pour le pic spectral et le centre de gravité des consonnes initiales.*

Chez les locuteurs anglophones, l'effet global du facteur « locuteur » est significatif pour le pic spectral des consonnes ($F=2,901$; $p<0,05$) et pour leur centre de gravité spectral ($F=24,154$; $p<0,001$). Un test a posteriori de Fisher indique qu'il existe

¹¹⁹ Voir section 3.1.1.1 du présent chapitre.

des différences significatives entre F2AN et H1AN pour le pic spectral ($p < 0,01$), ainsi qu'entre F2AN et H1AN ($p < 0,0001$), F2AN et H2AN ($p < 0,0001$), F1AN et F2AN ($p < 0,0001$) pour le centre de gravité. *On note donc une différence significative entre les locutrices F1AN et F2AN pour le centre de gravité des consonnes : ce dernier se situe en moyenne dans des fréquences plus élevées chez F2AN, ce qui semble en adéquation avec la taille de cette locutrice, qui est inférieure à celle de F1AN (169 cm contre 174 cm)¹²⁰.*

3.2.1.2 VOT

Pour tester la significativité des différences de VOT entre femmes et hommes francophones, j'ai réalisé plusieurs ANOVA à un facteur, le facteur « genre du locuteur ». Une première analyse porte sur le VOT des occlusives [t] et [k]. Le résultat de ce test montre que *le VOT des locutrices francophones est significativement plus long que celui des locuteurs pour ces occlusives sourdes*, avec $F(1,22)=4,332$ et $p < 0,05$. Une analyse équivalente, portant cette fois-ci *sur les occlusives voisées ([d] et [g])*, a été effectuée : *il apparaît que le VOT des locutrices est significativement plus court que celui des hommes francophones* : $F(1,22)=9,87$ avec $p < 0,01$. La figure 38, ci-après, présente une synthèse de ces deux analyses menées sur les locuteurs francophones ($n=4$).

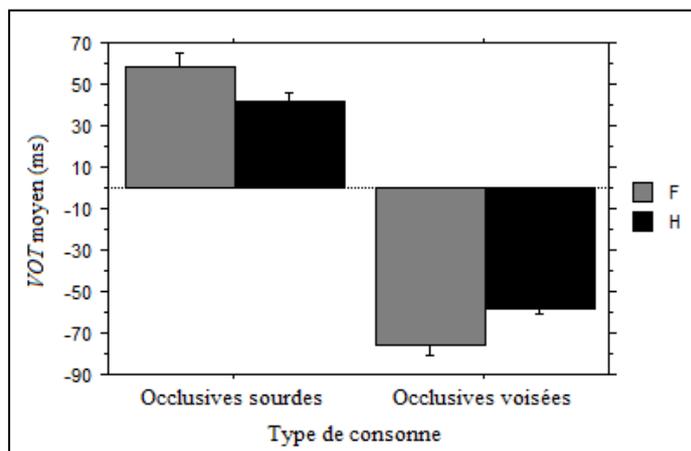


Figure 38 – VOT moyen (ms) en fonction du type de consonne initiale et du genre du locuteur (femme ou homme) pour les quatre locuteurs francophones, avec les barres d'erreur-type.

¹²⁰ Comme nous l'avons vu précédemment, la taille des locuteurs permet de déduire la longueur relative de leur conduit vocal (très forte corrélation positive entre ces deux variables, voir Fitch & Giedd, 1999). Toutes choses égales par ailleurs, plus le conduit vocal est long et plus les fréquences de résonance tendent à être basses (Fant, 1966, 1970).

Chez les locuteurs anglophones (n=4), les mêmes analyses ont été conduites. L'ANOVA à un facteur, « genre du locuteur » sur le *VOT* des occlusives aspirées ($[t^h]$ et $[k^h]$) met en évidence une différence très significative en fonction du genre : le *VOT* des locutrices anglophones est significativement plus long que celui des hommes, avec $F(1,22)=29,584$ et $p<0,0001$. Concernant les occlusives non-aspirées ($[d]$ et $[g]^{121}$), le résultat du test statistique montre, là encore, une différence significative sur le plan du *VOT* en fonction du genre du locuteur. Contrairement aux francophones, il est ici significativement plus long chez les locutrices, avec $F(1,22)=10,42$ et $p<0,01$. Le graphique correspondant à ces deux analyses est visible ci-dessous (figure 39).

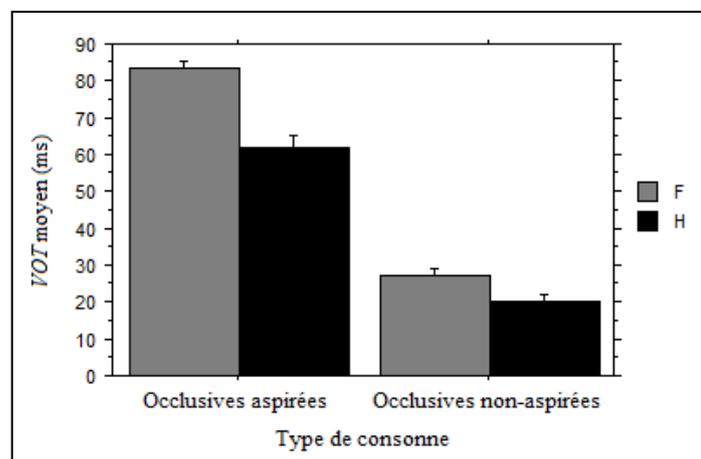


Figure 39 – *VOT* moyen (ms) en fonction du type de consonne initiale et du genre du locuteur (femme ou homme) pour les quatre locuteurs anglophones, avec les barres d'erreur-type.

Dans un deuxième temps, je me suis intéressé plus précisément au contraste entre les consonnes occlusives voisées / non-voisées (pour les francophones) et aspirées / non-aspirées (pour les anglophones) en terme de différence de *VOT*¹²², pour tester s'il existe sur ce point des différences liées au genre du locuteur. Une ANOVA à un facteur, « genre du locuteur », a donc été effectuée sur le contraste de *VOT* entre les consonnes sourdes et voisées chez les francophones. L'analyse révèle un effet très significatif de ce facteur, avec $F(1,22)=18,195$ et $p<0,001$. Ainsi, le contraste entre occlusive sourde et voisée est,

¹²¹ Pour rappel, ces occlusives ne peuvent être ici considérées comme des occlusives « voisées ». Voir section 2.3 du présent chapitre.

¹²² Par exemple, différence entre le *VOT* (en ms) de la consonne initiale dans le mot [kapi] et celui de la consonne initiale dans le mot [gapi], chez les francophones.

au plan du VOT, significativement plus grand chez les femmes que chez les hommes francophones.

Une analyse similaire portant sur le contraste aspirées / non-aspirées pour les locuteurs anglophones américains, montre également un effet significatif du facteur « genre du locuteur » : $F(1,22)=10,816$ et $p<0,01$. Comme chez les francophones, le contraste entre les deux types d'occlusives est, en termes de VOT, plus grand chez les locutrices que chez les locuteurs anglophones.

3.2.1.3 F0

Une ANOVA à deux facteurs, « genre du locuteur » et « consonne », a été conduite sur la fréquence fondamentale moyenne des consonnes voisées initiales produites par les quatre locuteurs francophones. Le graphique correspondant à cette analyse est présenté ci-après (figure 40).

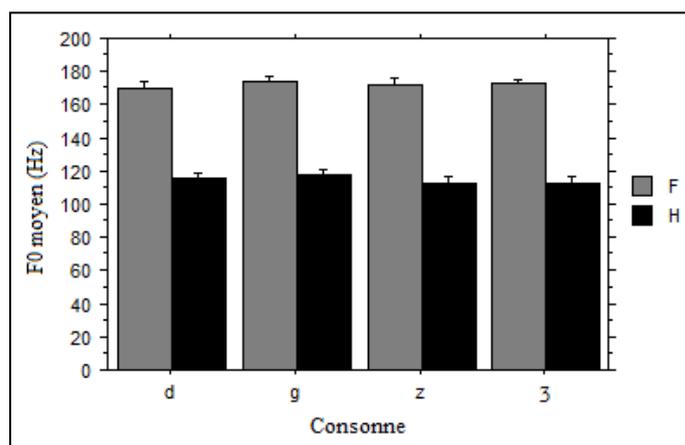


Figure 40 – Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les locuteurs francophones ($n=4$), avec les barres d'erreur-type.

Il ressort de cette analyse un effet global très significatif du facteur « genre du locuteur » sur le F0 moyen des consonnes initiales voisées chez les francophones, avec $F(1,40)=465,177$ et $p<0,0001$. Il n'existe aucune interaction entre les facteurs « consonne » et « genre du locuteur » avec $F(3,40)=0,254$ et $p>0,85$. Cela signifie que le F0 moyen des femmes francophones est significativement plus élevé que celui de leurs homologues masculins sur les consonnes initiales voisées, et que cette différence ne varie pas de manière significative d'une consonne à l'autre.

Une analyse similaire a été effectuée pour les quatre locuteurs anglophones américains, sur le F0 moyen des consonnes initiales voisées. Il s'agit là encore d'une ANOVA à deux facteurs : « genre du locuteur » et « consonne ». Ce deuxième facteur ne présente ici que deux niveaux, contre quatre chez les francophones : comme nous l'avons vu précédemment¹²³, chez les anglophones, seules les consonnes fricatives [z] et [ʒ] sont effectivement voisées dans cette position accentuée, pré-vocalique et initiale de mot. Le graphique résultant de ce test statistique apparaît en figure 41, ci-dessous.

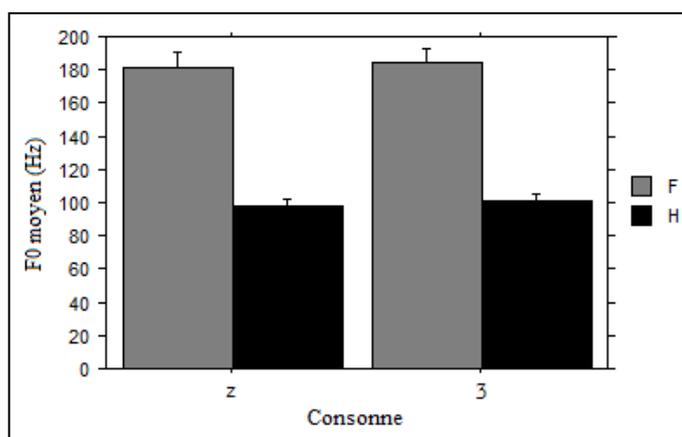


Figure 41 – Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de consonne et du genre du locuteur (femme ou homme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.

Tout comme chez les francophones, il existe un effet global fortement significatif du facteur « genre du locuteur » sur le F0 moyen des consonnes initiales produites par les locuteurs anglophones, avec $F(1,20)=141,075$ et $p<0,0001$. De même il n'existe aucune interaction entre les facteurs « consonne » et « genre du locuteur » : $F(1,20)=0,001$ et $p>0,95$. *Le F0 moyen des locutrices anglophones sur les consonnes initiales voisées est donc significativement plus élevé que celui des locuteurs masculins, avec une différence qui reste stable d'un type de consonne à l'autre.*

3.2.1.4 Durée

Afin de tester l'influence du genre du locuteur sur la durée des consonnes initiales produites par les quatre locuteurs francophones parisiens, une ANOVA à deux facteurs

¹²³ Voir section 2.3 du présent chapitre.

(« genre du locuteur » et « consonne ») a été menée sur ces données. Un graphique illustrant cette analyse est visible ci-dessous (figure 42).

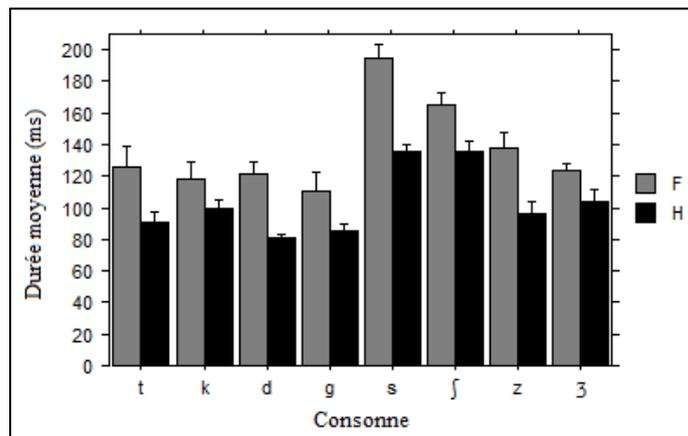


Figure 42 – Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.

Le résultat de ce test indique qu'il existe chez les francophones un effet global très significatif du facteur « genre du locuteur » sur la durée moyenne des consonnes, avec $F(1,80)=70,167$ et $p<0,0001$. D'autre part, il n'existe pas d'interaction significative entre les facteurs « consonne » et « genre du locuteur » avec $F(7,80)=1,379$ et $p>0,2$. *La durée des consonnes initiales est donc significativement plus élevée chez les locutrices que chez les locuteurs francophones et cette différence demeure relativement stable d'un type de consonne à un autre.*

Une analyse statistique équivalente (ANOVA à deux facteurs, « genre du locuteur » et « consonne ») a été réalisée sur la durée des consonnes initiales de mot produites par les quatre locuteurs anglophones. Le graphique correspondant à cette analyse est présenté ci-après (figure 43).

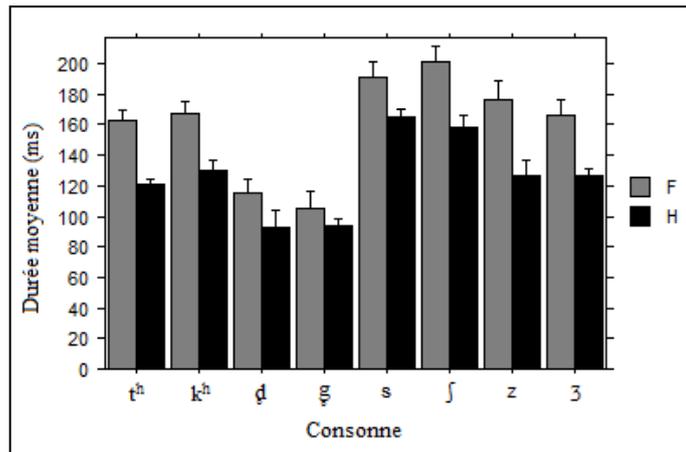


Figure 43 – Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.

A l'instar de leurs homologues francophones, il existe chez les locuteurs anglophones américains un effet global très significatif du facteur « genre du locuteur » sur la durée moyenne des consonnes initiales, avec $F(1,80)=54,058$ et $p<0,0001$. On ne note aucune interaction entre les facteurs « consonne » et « genre du locuteur », avec $F(7,80)=0,895$ et $p>0,5$. Pour les anglophones, les consonnes initiales de mot sont donc d'une longueur supérieure chez les locutrices, avec un écart qui reste équivalent sur les différentes consonnes.

3.2.2 Analyse statistique des données obtenues pour les voyelles

Les résultats de l'analyse statistique des données obtenues pour les voyelles seront présentés en quatre temps. La première sous-section est consacrée aux valeurs des formants vocaliques (F1, F2 et F3), la seconde à la qualité de voix, la troisième à la fréquence fondamentale et la dernière à la durée. Les chiffres obtenus pour les locuteurs francophones et ceux obtenus pour les locuteurs anglophones seront de nouveau présentés alternativement.

3.2.2.1 Formants vocaliques (F1, F2, F3)

Plusieurs tests statistiques ont été conduits sur les données relatives aux formants vocaliques des locuteurs francophones. Dans un premier temps, j'ai effectué une ANOVA à deux facteurs, « voyelle » et « genre du locuteur », sur la fréquence du premier formant vocalique (F1). Les résultats indiquent une absence d'effet significatif du facteur « genre

du locuteur » : $F(1,102)=0,914$ avec $p>0,3$. *Il n'existe donc pas de différence globale femmes-hommes significative sur le plan du F1 des voyelles produites par les francophones.* On note par ailleurs une absence d'interaction significative entre les facteurs « voyelle » et « genre du locuteur » ($F(2,102)=2,494$; $p>0,05$), la différence hommes-femmes a néanmoins été testée individuellement pour le F1 de chacune des trois voyelles cardinales, en procédant à des ANOVA à un facteur (« genre du locuteur »). Ces trois tests révèlent des différences inter-genres non-significatives, tant pour la voyelle [i] ($F(1,34)=2,869$; $p>0,1$), que pour le [a] ($F(1,34)=1,935$; $p>0,1$) et le [u] ($F(1,34)=3,512$; $p>0,05$).

Des analyses similaires ont été réalisées pour le deuxième formant (F2) des voyelles cardinales prononcées par les locuteurs francophones. Tout d'abord, une ANOVA à deux facteurs, « voyelle » et « genre du locuteur », a été conduite sur la fréquence de F2. Contrairement au F1, l'effet global du facteur « genre du locuteur » est ici très fortement significatif : $F(1,102)=247,477$ et $p<0,0001$. *Il existe donc une différence globale femmes-hommes très significative sur le plan du F2 des voyelles produites par les francophones, ce dernier se situe en moyenne dans des fréquences significativement plus élevées chez les locutrices.* D'autre part, on constate une interaction significative entre les facteurs « voyelle » et « genre du locuteur » ($F(2,102)=34,684$; $p<0,0001$) : j'ai par conséquent souhaité tester la différence hommes-femmes individuellement pour le F2 de chacune des trois voyelles cardinales, *via* des ANOVA à un facteur (« genre du locuteur »). *Cette différence est extrêmement significative pour la voyelle [i] ($F(1,34)=525,914$; $p<0,0001$), fortement significative pour le [a] ($F(1,34)=98,642$; $p<0,0001$) et faiblement significative pour la voyelle arrière [u] ($F(1,34)=6,521$; $p<0,02$).*

Les données relatives au troisième formant (F3) des voyelles produites par les francophones ont également fait l'objet des mêmes analyses statistiques. L'ANOVA globale à deux facteurs (« voyelle » et « genre du locuteur »), révèle un effet fortement significatif du genre du locuteur sur la fréquence du troisième formant vocalique, toutes voyelles confondues : $F(1,102)=240,17$ avec $p<0,0001$. *Il existe donc bien une différence globale inter-genre très significative sur le plan du F3 des voyelles prononcées par les locuteurs francophones parisiens : sa fréquence est en moyenne significativement plus élevée chez les femmes, toutes voyelles confondues.* On note une absence d'interaction

significative entre les facteurs « voyelle » et « genre du locuteur » ($F(2,102)=1,433$; $p>0,2$). La différence hommes-femmes a néanmoins été testée individuellement pour le F3 de chacune des trois voyelles cardinales, en procédant à des ANOVA à un facteur (« genre du locuteur »). Ces trois analyses font apparaître des *différences inter-genres fortement significatives, tant pour la voyelle avant [i] ($F(1,34)=541,408$; $p<0,0001$), que pour la voyelle ouverte [a] ($F(1,34)=121,171$; $p<0,0001$) et la voyelle arrière [u] ($F(1,34)=37,159$; $p<0,0001$).* Trois graphiques illustrant de manière synthétique l'ensemble des analyses effectuées sur les formants des voyelles produites par les locuteurs francophones ($n=4$) sont présentés en figure 44, ci-dessous.

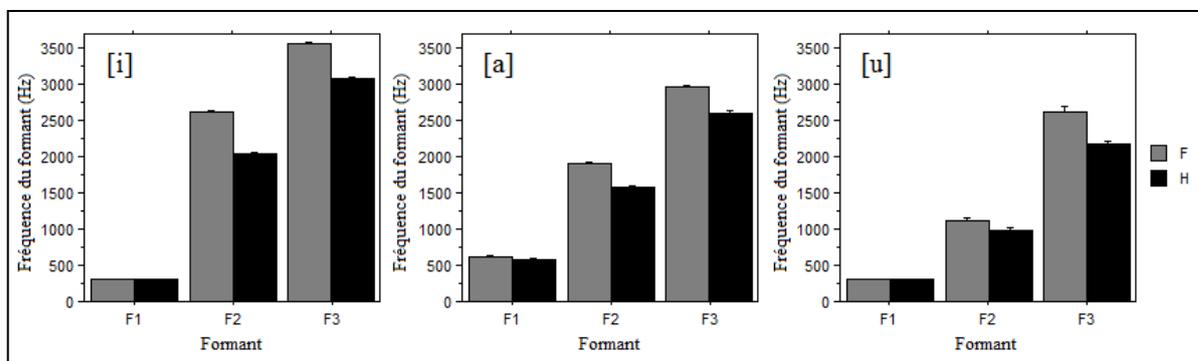


Figure 44 – Graphiques représentant la fréquence moyenne (en Hz) des trois premiers formants vocaliques (F1, F2, F3) pour les trois voyelles cardinales [i], [a], [u] produites par les locutrices ($n=2$) et locuteurs ($n=2$) francophones, avec les barres d'erreur-type.

Le même ensemble de tests statistiques a été conduit sur les données relatives aux formants vocaliques des locuteurs anglophones américains. J'ai tout d'abord effectué une ANOVA à deux facteurs, « voyelle » et « genre du locuteur », sur la fréquence du premier formant vocalique (F1), toutes voyelles confondues. Les résultats mettent en évidence un effet très significatif du facteur « genre du locuteur » : $F(1,102)=364,857$ avec $p<0,0001$. Contrairement aux francophones, *il existe donc une différence globale femmes-hommes significative sur le F1 des voyelles produites par les anglophones : la fréquence de ce formant est globalement plus élevée chez les locutrices.* On note, d'autre part, une interaction significative entre les facteurs « voyelle » et « genre du locuteur » ($F(2,102)=121,354$; $p<0,0001$), la différence hommes-femmes a par conséquent été testée individuellement pour le F1 de chacune des trois voyelles cardinales, en procédant à des ANOVA à un facteur (« genre du locuteur »). Ces tests révèlent *une différence inter-genres importante et très fortement significative pour la*

voyelle ouverte [æ] ($F(1,34)=236,665$; $p<0,0001$) et des différences plus faibles mais néanmoins significatives pour les voyelles fermées [i:] ($F(1,34)=92,298$; $p<0,0001$) et [u:] ($F(1,34)=62,373$; $p<0,001$).

Des analyses du même type ont été effectuées pour le deuxième formant (F2) des voyelles cardinales prononcées par les locuteurs anglophones. Dans un premier temps, une ANOVA à deux facteurs, « voyelle » et « genre du locuteur », a été réalisée sur la fréquence de F2. A l'instar du premier formant, l'effet global du facteur « genre du locuteur » est, ici encore, très fortement significatif : $F(1,102)=98,541$ et $p<0,0001$. *On note donc une différence globale hommes-femmes très significative pour le F2 des voyelles produites par les anglophones, ce dernier se situe en moyenne dans des fréquences significativement plus hautes chez les locutrices que chez les locuteurs.* D'autre part, on constate une interaction faible mais significative entre les facteurs « voyelle » et « genre du locuteur » ($F(2,102)=5,002$; $p<0,01$) : j'ai donc souhaité tester la différence inter-genres individuellement pour le F2 de chacune des trois voyelles cardinales, à l'aide d'ANOVA à un facteur (« genre du locuteur ») à deux niveaux (« femme », « homme »). *Cette différence est nettement significative, tant pour la voyelle antérieure [i:] ($F(1,34)=54,372$; $p<0,0001$), que pour le [æ] ($F(1,34)=132,237$; $p<0,0001$) et la voyelle arrière [u:] ($F(1,34)=23,207$; $p<0,0001$).*

Les données relatives au troisième formant (F3) des voyelles produites par les anglophones ont elles aussi fait l'objet des mêmes analyses statistiques. L'ANOVA globale à deux facteurs (« voyelle » et « genre du locuteur ») révèle, comme chez les francophones, un effet fortement significatif du genre du locuteur sur la fréquence du troisième formant vocalique : $F(1,102)=290,178$ et $p<0,0001$. *Il existe donc une différence globale inter-genre très significative sur le F3 des voyelles prononcées par les locuteurs anglophones : sa fréquence est en moyenne significativement plus élevée chez les femmes, toutes voyelles confondues.* On note par ailleurs l'existence d'une interaction significative entre les facteurs « voyelle » et « genre du locuteur » ($F(2,102)=18,578$; $p<0,0001$). La différence hommes-femmes a donc été testée individuellement pour le F3 de chacune des trois voyelles cardinales, en procédant de nouveau à des ANOVA à un facteur (« genre du locuteur »). Ces trois analyses mettent en évidence *une différence inter-genres importante et fortement significative pour les voyelles fermées [i:] ($F(1,34)=132,54$; $p<0,0001$) et [u:] ($F(1,34)=135,443$; $p<0,0001$), ainsi qu'une*

différence plus faible mais toujours significative pour le [æ] ($F(1,34)=50,129$; $p<0,0001$). La figure 45, ci-après, contient trois graphiques illustrant de manière synthétique l'ensemble des analyses effectuées sur les formants des voyelles produites par les locuteurs anglophones ($n=4$).

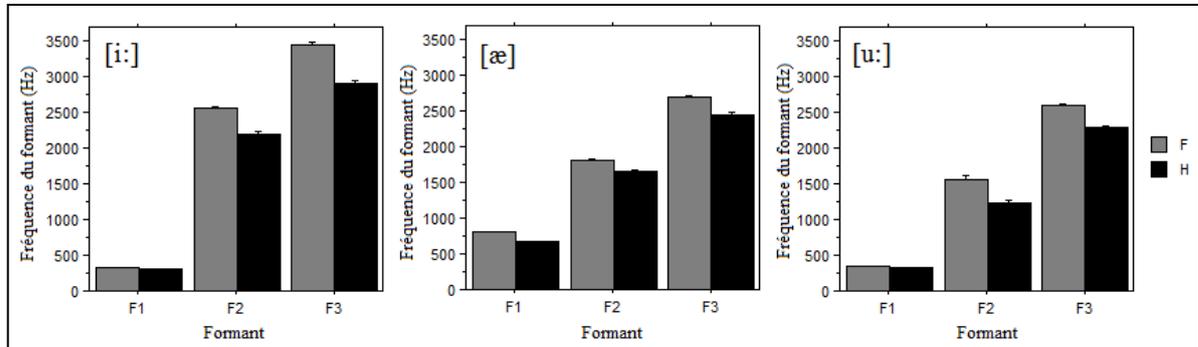


Figure 45 – Graphiques représentant la fréquence moyenne (en Hz) des trois premiers formants vocaliques (F1, F2, F3) pour les trois voyelles [i:], [æ], [u:] produites par les locutrices (n=2) et locuteurs (n=2) anglophones, avec les barres d'erreur-type.

Taille des locuteurs et formants vocaliques

Comme cela a été fait pour les pics spectraux et les centres de gravité des consonnes initiales¹²⁴, j'ai conduit plusieurs analyses pour établir l'éventuelle significativité des différences inter-locuteurs, en particulier entre ceux du même genre, pour les valeurs formantiques des voyelles. Il s'agit de deux ANOVA à trois facteurs, « locuteur », « formant » et « voyelle », menées successivement sur l'ensemble des valeurs formantiques des voyelles produites par les francophones, puis sur celles des voyelles prononcées par les locuteurs anglophones. En complément des résultats concernant l'effet global du facteur « locuteur », je m'intéresserai particulièrement aux conclusions du test PLSD de Fisher pour les différents locuteurs pris deux à deux, qui seront mis en parallèle avec la taille des locuteurs¹²⁵.

Sans surprise, l'effet global du facteur « locuteur » sur la fréquence des formants vocaliques (F1, F2 et F3) est fortement significatif ($F(3,288)=372,287$ avec $p<0,0001$) chez les locuteurs francophones. Le PLSD de Fisher révèle, fort logiquement, des

¹²⁴ Voir section 3.2.1.1 du présent chapitre.

¹²⁵ Voir section 3.1.2.1 du présent chapitre.

différences inter-locuteurs très significatives entre les locuteurs de genre opposé, c'est-à-dire entre F1FR et H1FR ($p < 0,0001$), F1FR et H2FR ($p < 0,0001$), F2FR et H1FR ($p < 0,0001$) et F2FR et H2FR ($p < 0,0001$), mais le test fait également apparaître des différences significatives entre F1FR et F2FR ($p < 0,0001$) ainsi qu'entre H1FR et H2FR ($p < 0,0001$). *Les valeurs formantiques de F1FR sont, en effet, significativement plus élevées que celles de F2FR, ce qui semble en adéquation avec la taille relative de ces locutrices (162 cm pour F1FR, 175 cm pour F2FR¹²⁶). Par ailleurs, les formants vocaliques des voyelles produites par H1FR sont significativement plus hauts que ceux des voyelles prononcées par le locuteur H2FR, malgré la taille quasi-identique de ces deux locuteurs (respectivement 183 cm et 182 cm).*

Pour les locuteurs anglophones, l'effet global du facteur « locuteur » sur les valeurs formantiques (F1, F2, F3) des voyelles est également très significatif ($F(3,288)=248,875$ avec $p < 0,0001$). Un test a posteriori de Fisher indique qu'il existe des différences hautement significatives entre les locuteurs anglophones de genre opposé, que ce soit entre F1AN et H1AN ($p < 0,0001$), F1AN et H2AN ($p < 0,0001$), F2AN et H1AN ($p < 0,0001$) ou entre F2AN et H2AN ($p < 0,0001$). Plus intéressant, le résultat de l'analyse indique également une différence significative entre les valeurs formantiques des locuteurs H1AN et H2AN ($p < 0,0001$). *En dépit d'une taille plus élevée (185 cm contre 178 cm pour H1AN) et donc d'un conduit vocal probablement plus long, les formants des voyelles produites par H2AN se situent globalement dans des fréquences significativement plus élevées que ceux des voyelles prononcées par le locuteur H1AN. Cela suggère que les deux locuteurs adoptent des conduites articulatoires sensiblement différentes lors de la production des voyelles.*

3.2.2.2 Qualité de voix

Afin de savoir si les différences observées sur le plan de la qualité de voix¹²⁷ sont significatives, j'ai conduit une ANOVA à un facteur, « genre du locuteur » (homme ou femme), sur la différence d'intensité (exprimée en dB) entre le premier et le deuxième

¹²⁶ Ces données sur la taille des deux locutrices laissent en effet supposer que le conduit vocal de F2FR est plus long que celui de F1FR (voir Fitch & Giedd, 1999) et que par conséquent les fréquences de résonance (dont les formants vocaliques) tendraient à être plus basses chez F2FR, toutes choses égales par ailleurs (Fant, 1966).

¹²⁷ Terme employé ici comme synonyme de « type de phonation ».

harmonique des voyelles ouvertes [a] produites par les francophones. Il en ressort un effet très significatif, avec $F(1,34)=69,516$ et $p<0,0001$. *La différence d'intensité H1-H2 est significativement plus élevée chez les locutrices que chez les locuteurs francophones.*

Une analyse similaire a été réalisée sur les locuteurs anglophones américains, pour tester l'influence du facteur « genre du locuteur » sur la différence d'intensité entre le premier et le deuxième harmonique des réalisations de la voyelle ouverte [æ]. Là encore l'effet est fortement significatif : $F(1,34)=101,079$ avec $p<0,0001$. *Pour les anglophones, c'est donc à nouveau chez les locutrices que la différence H1-H2 est la plus élevée, ce qui caractérise un QOG¹²⁸ plus élevé et donc une voix plus breathy.*

Dans un troisième temps, j'ai souhaité effectuer un nouveau test permettant de réunir les données des anglophones et de francophones et d'observer des éventuelles différences inter-langues significatives. Pour cela, une ANOVA à un facteur à quatre niveaux a été effectuée : il s'agit du facteur « type de locuteur » (homme anglophone, homme francophone, femme anglophone, femme francophone). La variable dépendante étant la différence d'intensité (en dB) entre le premier et le deuxième harmonique des réalisations de voyelles ouvertes. Le graphique correspondant à cette analyse est présenté ci-dessous (figure 46).

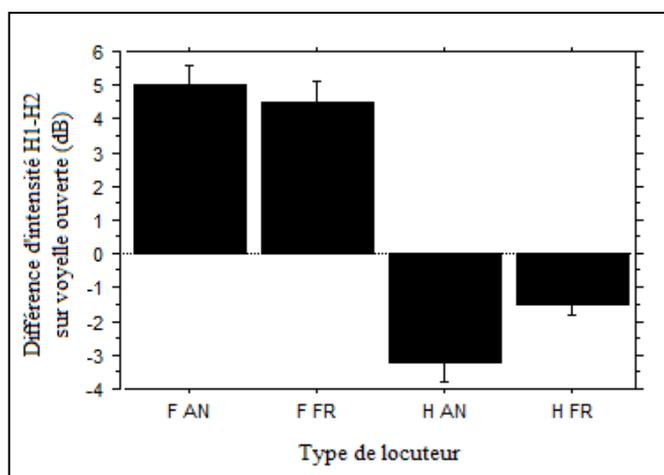


Figure 46 – Différence d'intensité H1-H2 (dB) sur les voyelles ouvertes en fonction du type de locuteur : femme anglophone (n=2), femme francophone (n=2), homme anglophone (n=2), homme francophone (n=2), avec les barres d'erreur-type.

¹²⁸ « Quotient d'Ouverture Glottique » ou « *Glottal Open Quotient* ». Voir les sections 2.4 du présent chapitre et 3.5 du chapitre 1 pour plus de détails.

Sans surprise, l'effet global du facteur « type de locuteur » sur la différence d'intensité H1-H2 des réalisations de voyelles ouvertes est nettement significatif : $F(3,68)=58,62$ et $p<0,0001$. Deux résultats intéressants ressortent du test PLSD de Fisher. Tout d'abord, on note une absence de différence significative entre les femmes anglophones et les femmes francophones ($p>0,4$). En revanche, *il existe bien une différence significative entre les hommes anglophones et les hommes francophones ($p<0,05$) : la différence H1-H2 est significativement plus faible chez les anglophones que chez les francophones, ce qui implique un QOG très faible et donc une voix plus creaky.*

3.2.2.3 F0

Une ANOVA à deux facteurs, « genre du locuteur » et « voyelle », a été réalisée sur la fréquence fondamentale moyenne des voyelles présentes en syllabe initiale et produites par les locuteurs francophones parisiens ($n=4$). Le graphique correspondant à cette analyse est visible ci-après (figure 47).

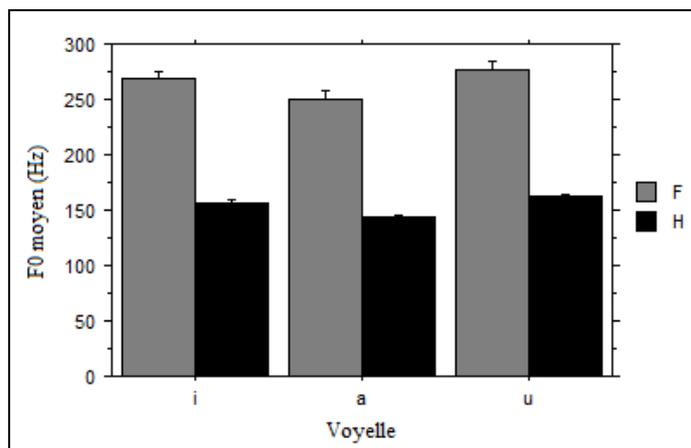


Figure 47 – Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs francophones ($n=4$), avec les barres d'erreur-type.

Sans surprise, l'ANOVA indique un effet global très significatif du facteur « genre du locuteur » sur le F0 moyen des voyelles chez les francophones, avec $F(1,102)=678,694$ et $p<0,0001$. Il n'existe aucune interaction entre les facteurs « voyelle » et « genre du locuteur » avec $F(2,102)=0,382$ et $p>0,65$. *Le F0 moyen des voyelles est donc ici significativement plus élevé chez les locutrices que chez les locuteurs francophones, avec un écart qui demeure relativement stable d'une voyelle à l'autre.*

Une analyse similaire a été conduite pour les quatre locuteurs anglophones américains, sur le F0 moyen des voyelles présentes en syllabe initiale. Il s'agit à nouveau d'une ANOVA à deux facteurs : « genre du locuteur » et « voyelle ». Le graphique résultant de ce test statistique apparaît en figure 48, ci-dessous.

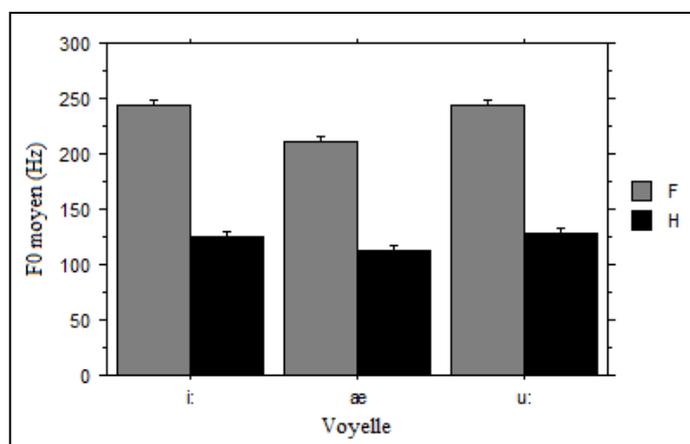


Figure 48 – Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.

Comme chez les francophones, il existe un effet global fortement significatif du facteur « genre du locuteur » sur le F0 moyen des voyelles produites par les locuteurs anglophones américains, avec $F(1,102)=867,92$ et $p<0,0001$. De même, il n'existe pas d'interaction significative entre les facteurs « voyelle » et « genre du locuteur » : $F(2,102)=3,015$ et $p>0,05$. *La fréquence fondamentale des voyelles est donc significativement plus élevée lorsque ces dernières sont produites par les locutrices anglophones, et cette différence femmes-hommes demeure stable d'une voyelle à l'autre.*

3.2.2.4 Durée

Afin de tester l'influence du genre du locuteur sur la durée des voyelles en syllabe initiale produites par les locuteurs francophones (n=4), une ANOVA à deux facteurs (« genre du locuteur » et « voyelle ») a été effectuée sur ces données. Un graphique illustrant cette analyse est visible ci-après (figure 49).

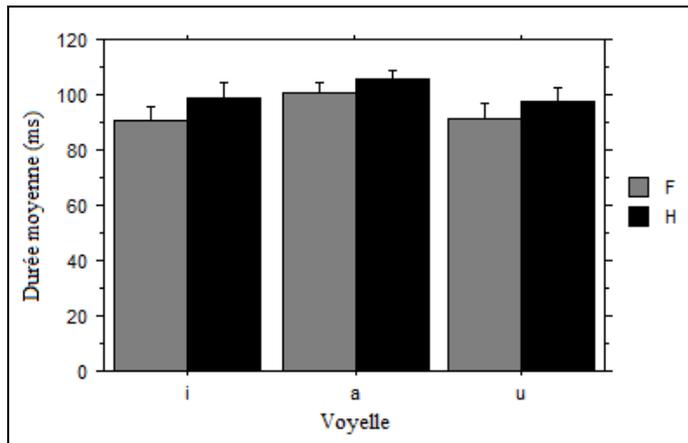


Figure 49 – Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.

Contrairement aux consonnes, l'effet global du facteur « genre du locuteur » sur la durée des voyelles n'est pas significatif chez les francophones : $F(1,102)=2,77$ et $p>0,05$. Il n'existe, d'autre part, aucune d'interaction entre les facteurs « voyelle » et « genre du locuteur » : $F(2,102)=0,042$ et $p>0,95$. *Il n'existe donc pas de différence de durée significative entre les voyelles produites par les locutrices francophones et celle produites par les locuteurs.*

Une analyse statistique de même type (ANOVA à deux facteurs, « genre du locuteur » et « voyelle ») a été conduite sur la durée des voyelles en syllabe initiale de mot produites par les locuteurs anglophones (n=4). Le graphique correspondant à cette analyse est présenté ci-après (figure 50).

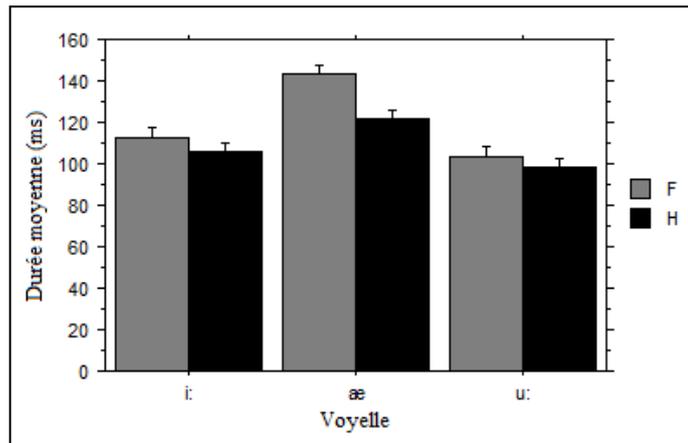


Figure 50 – Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.

Il existe, chez les locuteurs anglophones, un effet global significatif du facteur « genre du locuteur » sur la durée moyenne des voyelles, avec $F(1,102)=9,324$ et $p<0,01$. L'analyse indique, par ailleurs, l'absence d'interaction significative entre les facteurs « voyelle » et « genre du locuteur » $F(2,102)=2,174$ et $p>0,1$. *La durée des voyelles en syllabe initiale accentuée est donc significativement plus élevée chez les femmes que chez les hommes anglophones, avec une différence qui reste globalement stable d'une voyelle à une autre.*

3.2.3 Analyse statistique des données obtenues pour les mots dissyllabiques entiers

Cette dernière sous-section comprend les résultats de l'analyse statistique pour les données obtenues sur les mots dissyllabiques entiers : le F0 moyen, la plage de variation de F0, la durée moyenne ainsi que la répartition temporelle consonnes / voyelles seront présentés. Les résultats sont donnés successivement pour les francophones parisiens et les anglophones américains.

3.2.3.1 F0 moyen

Un test statistique de type ANOVA à un facteur, « genre du locuteur » (femme / homme), a été réalisé sur la fréquence fondamentale moyenne des mots dissyllabiques produits par les locuteurs francophones (n=4). Le graphique correspondant à cette analyse est présenté ci-après (figure 51).

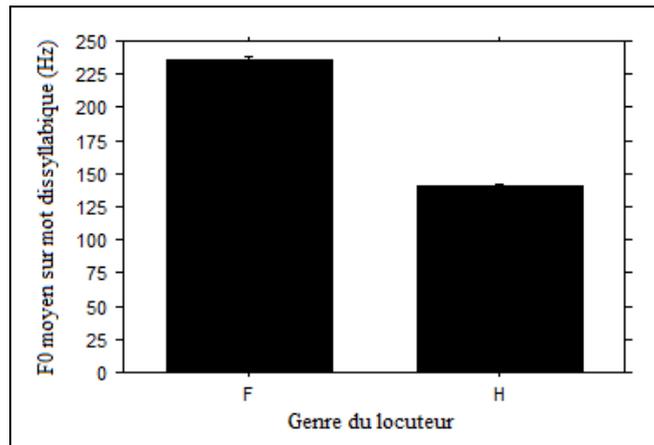


Figure 51 – F0 moyen sur les mots dissyllabiques (exprimé en Hz) en fonction du genre du locuteur (femme ou homme) pour les locuteurs francophones ($n=4$), avec les barres d'erreur-type.

Le résultat du test est particulièrement clair : il existe un effet fortement significatif du facteur « genre du locuteur », avec $F(1,106)=951,013$ et $p<0,0001$. *La fréquence fondamentale moyenne des femmes francophones sur les mots dissyllabiques est donc très significativement supérieure à celle de leurs homologues masculins.*

J'ai procédé à une analyse équivalente pour les quatre locuteurs anglophones. Il s'agit d'une ANOVA à un facteur, « genre du locuteur », sur la fréquence fondamentale moyenne des mots dissyllabiques. La figure 52, ci-dessous, illustre le résultat de ce test statistique.

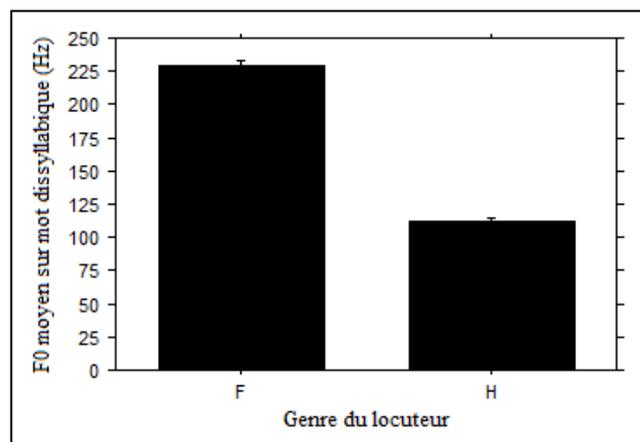


Figure 52 – F0 moyen sur les mots dissyllabiques (exprimé en Hz) en fonction du genre du locuteur (femme ou homme) pour les locuteurs anglophones ($n=4$), avec les barres d'erreur-type.

Comme cela était prévisible, il existe chez les anglophones un effet significatif du facteur « genre du locuteur » sur le F0 moyen des mots dissyllabiques : $F(1,106)=1159,938$ et $p<0,0001$. A l'instar des francophones, *la fréquence fondamentale moyenne des femmes anglophones est donc significativement plus haute que celle des hommes sur les mots dissyllabiques.*

Afin de mettre en évidence d'éventuelles différences inter-langues sur le plan du F0 moyen (sur les mots dissyllabiques), j'ai conduit une ANOVA à un facteur, "type de locuteur", à quatre niveaux : femme anglophone, femme francophone, homme anglophone et homme francophone. L'effet global du facteur est naturellement significatif ($F=729,559$; $p<0,0001$) et le test PLSD de Fisher montre que *le F0 des hommes anglophones est significativement plus bas que celui de leurs homologues francophones* ($p<0,0001$). En revanche, aucune différence significative entre les F0 moyen des femmes anglophones et francophones ($p>0,05$).

3.2.3.2 Plage de variation de F0

Afin de tester l'influence du genre du locuteur sur la plage de variation pour les mots dissyllabiques¹²⁹ produits par les locuteurs francophones, une ANOVA à un facteur (« genre du locuteur ») a été effectuée sur ces données. Lorsqu'elle est exprimée en hertz, la plage de variation moyenne de F0 sur les mots dissyllabiques varie de manière très significative en fonction du genre du locuteur, chez les francophones ($F(1,106)=248,105$; $p<0,0001$) : elle est globalement plus grande chez les locutrices que chez les locuteurs.

Pour vérifier s'il en est de même lorsque ces données sont converties en demi-tons¹³⁰, j'ai conduit une deuxième ANOVA pour tester l'effet du facteur « genre du locuteur » sur la « plage de variation en demi-tons sur les mots dissyllabiques ». Le graphique correspondant est présenté ci-après (figure 53).

¹²⁹ C'est-à-dire l'écart entre la fréquence fondamentale (ou la tonalité) la plus basse et la fréquence fondamentale (ou la tonalité) la plus haute atteinte au sein de chaque mot dissyllabique.

¹³⁰ La conversion en demi-tons permet de rendre compte de la variation réellement perçue par l'auditeur (voir notamment Henton, 1989a).

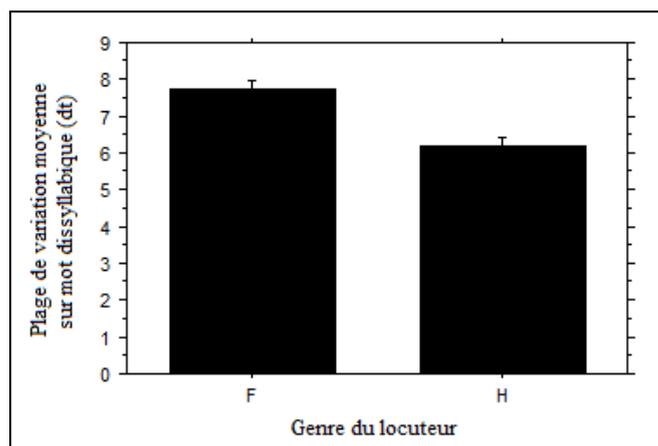


Figure 53 – *Plage de variation moyenne sur les mots dissyllabiques (exprimée en dt) en fonction du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.*

Le résultat de cette analyse est très net : il existe une influence très significative du facteur « genre du locuteur » sur la plage de variation exprimée en demi-tons, avec $F(1,106)=22,489$ et $p<0,0001$. *La plage de variation moyenne des femmes francophones sur les mots dissyllabiques est donc significativement plus grande que celle des hommes, qu'elle soit exprimée en hertz ou en demi-tons.*

Chez les locuteurs anglophones (n=4), la plage de variation moyenne de F0 sur les mots dissyllabiques varie également de manière très significative en fonction du genre du locuteur, lorsqu'elle est exprimée en hertz : $F(1,106)=128,831$ avec $p<0,0001$. Comme pour les locuteurs francophones, j'ai réalisé une deuxième analyse afin de vérifier si la différence était significative lorsque la plage de variation est exprimée en demi-tons. Le graphique correspondant à cette analyse est visible ci-après.

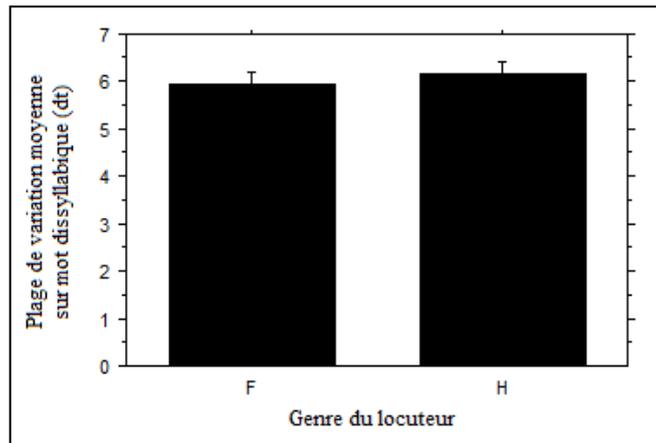


Figure 54 – *Plage de variation moyenne sur les mots dissyllabiques (exprimée en dt) en fonction du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.*

Le résultat du test montre que, contrairement aux francophones, la différence n'est plus significative lorsque l'on exprime la plage de variation en demi-tons : $F(1,106)=0,383$ avec $p>0,5$. *La plage de variation des locuteurs anglophones américains sur les mots dissyllabiques est donc supérieure chez les femmes lorsqu'elle est donnée en hertz, mais elle est en revanche du même ordre chez les hommes et les femmes lorsqu'elle est calculée en demi-tons.*

3.2.3.3 *Durée des mots entiers et répartition temporelle intra-mots*

Une analyse statistique de type ANOVA à un facteur, « genre du locuteur » (femme / homme), a été conduite sur la durée des mots dissyllabiques produits par les locuteurs francophones (n=4). Le graphique correspondant à cette analyse est présenté ci-après (figure 55).

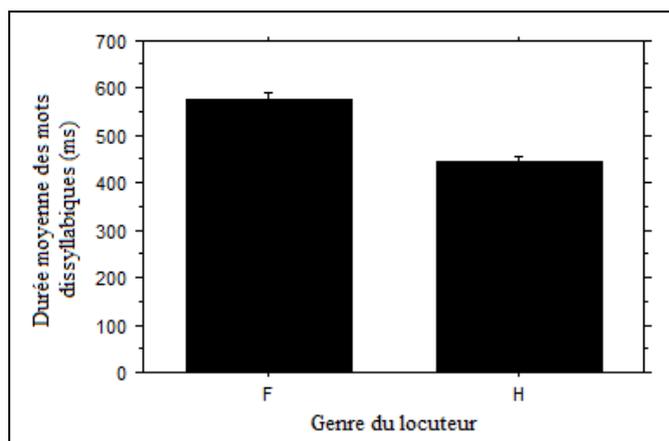


Figure 55 – *Durée moyenne des mots dissyllabiques (exprimée en ms) en fonction du genre du locuteur (femme ou homme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.*

Le résultat de l'ANOVA fait apparaître un effet très significatif du facteur « genre du locuteur » : $F(1,106)=67,524$ avec $p<0,0001$. *Les mots dissyllabiques produits par les locutrices francophones sont donc significativement plus longs que ceux prononcés par les locuteurs masculins.*

Une analyse du même type a été effectuée pour les locuteurs anglophones (n=4). Il s'agit à nouveau d'une ANOVA à un facteur, « genre du locuteur », sur la fréquence fondamentale moyenne des mots dissyllabiques. La figure 56, ci-après, illustre le résultat de ce test statistique.

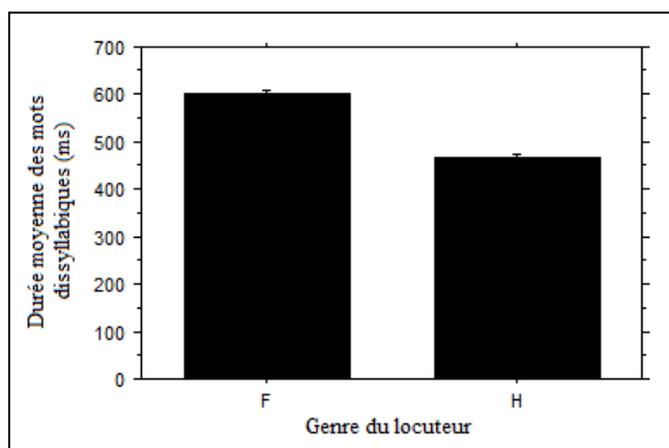


Figure 56 – *Durée moyenne des mots dissyllabiques (exprimée en ms) en fonction du genre du locuteur (femme ou homme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.*

Il existe bien un effet largement significatif du facteur « genre du locuteur » sur la durée des mots dissyllabiques produits par les anglophones : $F(1,106)=123,6$ avec $p<0,0001$. La tendance observée chez les francophones se retrouve donc également *pour les locuteurs anglophones américains : les mots prononcés par les femmes sont significativement plus longs que ceux produits par les hommes.*

Au-delà de la durée globale des énoncés, j'ai souhaité par ailleurs m'intéresser à la répartition temporelle consonnes / voyelles au sein des 24 mots dissyllabiques de type CVCV¹³¹. Une légère différence inter-genres avait été observée dans la sous-section précédente¹³², les consonnes occupant une place plus importante dans les mots produits par les femmes, aussi bien chez les francophones (52 % du mot en moyenne contre 46 % pour les hommes) que chez les anglophones (48 % contre 43 %).

Afin de tester si ces différences sont significatives, j'ai conduit une ANOVA à un facteur (« genre du locuteur ») à deux niveaux (« femme », « homme »), sur la part de la durée du mot (exprimée en pourcentage) occupée par les consonnes pour chacune des réalisations des 24 mots CVCV. Chez les locuteurs francophones, on note un effet global très significatif du facteur « genre du locuteur » ($F(1,94)=17,409$; $p<0,0001$). *En termes de durée, les consonnes occupent donc une place significativement plus importante dans les mots dissyllabiques CVCV produits par les femmes, que dans ceux produits par les hommes francophones.* Un effet nettement significatif du facteur « genre du locuteur » se retrouve également sur les données relatives aux mots produits par les anglophones ($F(1,94)=17,975$ et $p<0,0001$). *Les consonnes sont donc, de manière significative, proportionnellement plus longues dans les mots CVCV produits par les femmes anglophones, que dans ces mêmes mots produits par leurs homologues masculins.*

¹³¹ Ce qui exclut de cette analyse les trois mots de type VCV, c'est-à-dire [api], [ipi] et [upi] pour les francophones, et [æpi], [i:pi], [u:pi] pour les anglophones.

¹³² Voir section 3.1.3.3.

4. DISCUSSION - CONCLUSION PARTIELLE

L'analyse acoustique réalisée dans ce chapitre, portant sur les productions de mots dissyllabiques par des locutrices et locuteurs francophones parisiens et anglophones américains du nord-est des Etats-Unis, a permis de mettre en évidence un certain nombre de résultats intéressants, qu'il convient de synthétiser, d'interpréter et de mettre en perspective.

Concernant la fréquence fondamentale moyenne, qui est une estimation du nombre de cycles de vibration des plis vocaux effectués en une seconde et qui correspond d'un point de vue perceptif à la hauteur de la voix, des différences inter-genres significatives sont apparues pour les locuteurs des deux langues et à tous les niveaux : que ce soit sur les consonnes voisées initiales de mot et sur les voyelles de la syllabe initiale prises isolément, ou sur les mots dissyllabiques dans leur ensemble. Dans tous les cas, la fréquence fondamentale moyenne des locutrices est nettement supérieure à celle des hommes : à l'échelle des mots dissyllabiques, elle atteint en moyenne 230 Hz pour les femmes anglophones et 236 Hz pour les francophones, contre respectivement 113 Hz et 141 Hz chez leurs homologues masculins. Ces résultats confirment largement les études antérieures menées sur ce point (Takefuta et al., 1972 ; Boë et al., 1975). On note par ailleurs que la fréquence fondamentale des hommes anglophones américains est significativement plus basse que celle des hommes francophones : il paraît néanmoins hasardeux de tirer des conclusions générales sur ce point, quand on connaît la forte variabilité inter-sujets (au sein d'un même genre) qu'il existe sur le plan de la fréquence fondamentale moyenne (Pegoraro-Krook, 1988), même si les critères de sélection des locuteurs (non-fumeurs, même tranche d'âge, milieux socioculturels proches, etc.) ont probablement permis de limiter ce phénomène.

La plage de variation de la fréquence fondamentale, qui correspond ici à l'écart entre la fréquence (ou tonalité) la plus basse et la fréquence (ou tonalité) la plus élevée dans chaque mot dissyllabique, a également été analysée. Dans les deux langues, la plage de variation est significativement et largement plus grande chez les femmes que chez les hommes lorsqu'elle est mesurée en hertz : 103 Hz contre 50 Hz en moyenne pour les francophones, 80 Hz et 38 Hz chez les anglophones. Mais comme nous l'avons vu précédemment, la perception des fréquences par l'oreille humaine n'est pas linéaire mais

logarithmique, si l'on considère par exemple une variation brute de 50 Hz, la différence perçue sera plus grande entre 100 et 150 Hz qu'entre 250 et 300 Hz. Il est donc possible que les locutrices, ayant une fréquence fondamentale moyenne plus élevée que les hommes, utilisent une plage de variation brute (i.e. en hertz) plus grande afin d'arriver au même résultat perceptif (Henton, 1989a ; Simpson, 2009).

Il convient donc de considérer dans un deuxième temps la plage de variation des locutrices et locuteurs lorsque celle-ci est convertie en demi-tons, une échelle qui restitue la variation de hauteur perçue. Les résultats montrent une absence de différence significative chez les locuteurs anglophones : la plage de variation atteint 5,95 dt pour les femmes et 6,17 dt pour les hommes. L'explication avancée ci-dessus pourrait donc être avérée pour les anglophones américains. En revanche, chez les francophones, la plage de variation demeure significativement plus large chez les locutrices (7,73 dt) que chez les locuteurs (6,21 dt), même lorsqu'elle est exprimée en demi-tons. Une réelle différence femme-homme au plan de l'intonation semble donc exister chez les francophones. Cela tend à confirmer les résultats obtenus par Pépiot (2010, 2011), selon lesquels l'intonation (c'est-à-dire les variations de F0 dans le temps) jouerait un rôle dans l'identification du genre par la voix chez les auditeurs francophones : ces derniers tendraient à associer les phrases avec contour intonatif plat à une voix d'homme. Cependant, les plages de variation ont ici été mesurées sur des mots dissyllabiques et non sur des phrases, ce qui doit conduire à une certaine prudence dans l'interprétation des données, même si le contexte utilisé (*Il a dit « MOT » deux fois*) permet de considérer ces items comme des énoncés à part entière.

Sur le plan des formants vocaliques, nous avons noté chez les locuteurs francophones l'absence de différence inter-genres significative pour le premier formant, mais de fortes différences pour F2 et F3, situés dans des fréquences nettement plus élevées chez les locutrices. Dans le détail, on constate que la différence maximale pour F2 est atteinte sur la voyelle antérieure [i], alors que le [u], voyelle postérieure, présente une très faible différence hommes-femmes sur ce formant. Cela va dans le sens des observations faites notamment par Hillenbrand et al. (1995) et Pépiot (2009). Selon Nordström (1977), cette asymétrie dans les différences formantiques inter-genres pourrait être d'origine physiologique ; Fant (1966, 1975) évoque quant à lui une possible explication articulatoire, à savoir une plus forte protrusion des lèvres chez les locutrices

lors de la production des voyelles arrières arrondies. Pour le troisième formant, les différences sont relativement homogènes d'une voyelle à l'autre. Les résultats sont sensiblement différents chez les anglophones américains. En effet, il existe dans ce groupe une forte différence sur le premier formant de la voyelle ouverte [æ], avec des fréquences plus hautes chez les locutrices. Une différence inter-genres plus faible mais néanmoins significative existe également sur les voyelles fermées [i:] et [u:]. A l'instar des francophones, les fréquences de F2 et F3 sont là encore globalement plus élevées chez les femmes. Néanmoins, les différences sur le deuxième (et le troisième) formant sont cette fois-ci du même ordre pour les trois voyelles [i:], [æ] et [u:].

Ces fortes différences sur le F2 de la voyelle arrière, non-observées chez les locuteurs francophones, suggèrent que des variations articulatoires entrent ici en jeu : les locutrices anglophones ont probablement antériorisé légèrement leur voyelle [u:], ou effectué une moindre protrusion des lèvres (Fant, 1966, 1975). Un autre élément semble remettre en cause l'explication purement physiologique de ces différences formantiques : les formants du locuteur H1FR sont significativement plus élevés que ceux de H2FR, alors que leur taille (183 cm contre 182 cm) et donc probablement la longueur de leur conduit vocal (voir Fitch & Giedd, 1999 ; Cherng et al., 2002) sont quasi-identiques. Plus surprenant encore, les formants du locuteur H2AN (mesurant 185 cm) se situent dans des fréquences significativement plus élevées que ceux de H1AN (178 cm), en dépit d'un conduit vocal très probablement plus long qui devrait donc entraîner, à geste articulatoire identique, des fréquences de résonance plus basses (Fant, 1966). Notons d'autre part que les valeurs de F2 et F3 obtenues ici sont, pour les locuteurs des deux genres, globalement plus élevées que les valeurs de référence de Calliope (1989) pour le français et de Peterson & Barney (1952) et Hillenbrand et al. (1995) pour l'anglais américain. Cela s'explique par une antériorisation induite par la présence d'un [i] en finale des mots (dissyllabiques) du corpus.

Des différences inter-genres ont également été observées sur les fréquences de résonance des consonnes initiales de mot. Deux mesures ont été effectuées : le pic spectral et le centre de gravité. Chez les francophones, les fréquences sont globalement significativement plus élevées chez les femmes, sur ces deux paramètres acoustiques, ce qui va dans le sens des résultats obtenus notamment par Schwartz (1968), Nittrouer (1995) et Whiteside (1998b). En considérant les différentes consonnes prises

individuellement, on constate que les différences sont significatives pour le pic spectral de [z] et le centre de gravité de [d], [s] et [z]. Pour les locuteurs anglophones, il existe une différence globale (toutes consonnes confondues) pour le centre de gravité (fréquences plus élevées chez les femmes) mais pas pour le pic spectral. L'absence de différence globale pour ce dernier paramètre s'explique par une importante variabilité intra-sujet. En prenant les consonnes une à une, on note une différence inter-genres significative pour le pic spectral du [ʒ], ainsi que pour le centre de gravité de [d], [ʃ] et [z]. On remarque que les différences significatives sont, dans les deux langues, plutôt présentes sur les fricatives que sur les occlusives. D'autre part, chez les francophones, ces différences significatives se situent uniquement sur les consonnes alvéo-dentales, alors que chez les anglophones, elles apparaissent également sur les consonnes post-alvéolaires ([ʃ] et [ʒ]). Ces disparités inter-langues laissent une nouvelle fois supposer l'existence de variantes articulatoires d'ordre culturel entre les deux genres, ce qui rejoint les résultats obtenus par Heffernan (2004) sur des productions de la fricative [s] par des locutrices et locuteurs anglophones et japonophones.

L'analyse menée sur la qualité de voix des locuteurs, à travers la mesure de l'amplitude relative des deux premiers harmoniques (H1-H2) des voyelles ouvertes [a] et [æ], fournit des résultats concluants. Chez les locuteurs anglophones comme francophones, la différence H1-H2 est significativement plus élevée chez les femmes, ce qui est caractéristique d'un quotient d'ouverture glottique plus important et donc d'une voix de type *breathy* (ou soufflée). Cela étaye les résultats de Titze (1989) et Simpson (2009), pour qui cette différence aurait une origine physiologique : les plis vocaux des femmes, plus fins que ceux des hommes, auraient une tendance naturelle à ne pas se fermer complètement durant les cycles laryngés. En outre, la différence H1-H2 est significativement plus faible chez les hommes anglophones américains (-3,2 dB) que chez leurs homologues francophones (-1,5 dB), indiquant ainsi une qualité de voix plus *creaky* (ou craquée). L'utilisation de ce type de phonation pourrait donc bien être un marqueur socio-phonétique, variant notamment en fonction de la langue des locuteurs, comme le laissent entendre Henton & Bladon (1988) et Henton (1989b). Notons enfin que le contraste de qualité de voix femmes-hommes, plus important chez les anglophones que chez les francophones, pourrait en partie expliquer pourquoi les différences inter-genres au plan du F1 sont significatives uniquement chez les locuteurs anglophones. En effet, selon un modèle mis au point par Barney et al. (2004), le QOG influencerait la valeur du

premier formant : un QOG élevé entraînerait une augmentation de la valeur de ce dernier et inversement.

Concernant le *VOT* (ou DEV) des consonnes occlusives, défini comme le délai entre le relâchement de la consonne et l'entrée en vibration des plis vocaux, d'importantes différences inter-genres sont apparues. Chez les locuteurs francophones, le *VOT* des femmes est significativement plus long que celui des hommes sur les occlusives sourdes, et significativement plus court sur les occlusives voisées. Le contraste entre ces deux types de consonnes est donc très nettement accentué chez les locutrices par rapport aux locuteurs masculins. Pour les anglophones américains, le *VOT* est significativement plus long chez les femmes sur les occlusives aspirées comme sur les occlusives non-aspirées, ce qui va dans le sens des observations faites notamment par Swartz (1992), Whiteside & Irving (1997) ou encore Robb et al. (2005). Cependant, le contraste entre ces deux types de consonnes demeure, comme pour les francophones, significativement plus grand chez les locutrices que chez les locuteurs masculins. Ainsi, dans les deux langues, les femmes marquent une distinction plus forte que les hommes entre les deux types de consonnes occlusives (voisées / non-voisées ou aspirées / non-aspirées). Ces différences pourraient relever de facteurs culturels et socio-phonétiques, notamment d'une tendance à une articulation plus « soignée » chez les locutrices (voir Ditcheva & Dommergues, 2008 et Simpson, 2009).

Outre le *VOT*, plusieurs autres paramètres temporels ont été analysés, à savoir la durée des consonnes initiales de mot, la durée des voyelles en syllabe initiale, la durée globale des mots dissyllabiques et la répartition temporelle consonnes / voyelles au sein de ces mots. Tout d'abord, les consonnes initiales sont significativement plus longues chez les femmes que chez les hommes dans les deux langues. Quant aux voyelles en syllabe initiale, elles sont plus longues pour les femmes chez les anglophones, mais aucune différence inter-genre significative n'apparaît chez les francophones. A l'échelle des mots dissyllabiques entiers, les résultats sont similaires dans les deux langues : la durée de ces items est significativement plus élevée chez les locutrices que chez les locuteurs masculins (+ 30 % pour les francophones, + 28 % chez les anglophones). Cela rejoint des constatations faites dans plusieurs études antérieures (Byrd, 1994 ; Whiteside, 1995, 1996 ; Fitzsimons et al., 2001) et semble donc mettre en évidence, là encore, une

tendance à une articulation plus minutieuse chez les locutrices, tout du moins lors d'une tâche expérimentale de lecture, comme c'était le cas dans la présente étude.

Autre élément intéressant, on constate que les consonnes occupent une proportion temporelle plus importante du mot chez les femmes que chez les hommes, tant pour les francophones (52 % contre 46 %) que pour les anglophones (48 % contre 43 %). Les locutrices tendraient donc à faire durer plus longtemps les consonnes des mots (et donc moins longtemps les voyelles) que ne le font leurs homologues masculins. Ces différences, qui contredisent en partie une étude menée précédemment par Simpson et Ericsson (2003) sur des anglophones américains, relèvent très probablement d'habitudes articulatoires d'ordre socio-phonétique. Ces données pourraient être liées, une nouvelle fois, à la plus grande recherche d'intelligibilité chez les locutrices : Owren et Cardillo (2006) ont en effet mis en évidence le rôle primordial joué par les consonnes dans l'identification des mots par l'auditeur.

Cette analyse acoustique multiparamétrique aura donc permis d'observer dans le détail les différences inter-genres, validant ainsi les **hypothèses générales 1 et 2**, mais également leurs variations inter-langues chez les locuteurs anglophones américains du nord-est et les francophones parisiens. L'**hypothèse 3**, selon laquelle les différences acoustiques entre voix de femmes et voix d'hommes varient en fonction de la langue, est donc largement validée. Au final, il apparaît de manière claire que les différences physiologiques et anatomiques ne permettent pas de rendre compte de l'ensemble des différences inter-genres. Une partie non négligeable des différences observées, notamment la tendance à l'articulation plus claire et plus soignée des locutrices, relève très certainement de la construction sociale genrée. Néanmoins, les résultats obtenus dans cette étude doivent être interprétés avec précaution. Tout d'abord, seuls deux hommes et deux femmes ont été enregistrés dans chaque langue, ce qui malgré les critères de sélection très précis et les faibles différences observées entre les individus du même genre, ne permet pas de généraliser les résultats obtenus à l'ensemble des femmes et des hommes locuteurs de ces langues. D'autre part, le corpus analysé ici est constitué uniquement de mots dissyllabiques lus : il n'est absolument pas certain que des résultats similaires auraient été obtenus avec de la parole spontanée ou semi-spontanée.

Après avoir investigué en détail les différences acoustiques femmes-hommes, le chapitre suivant est consacré à deux expériences d'identification du genre par la voix, l'une menée sur des auditeurs américains et l'autre sur des francophones. Les items ayant été analysés acoustiquement dans le présent chapitre sont réutilisés comme stimuli pour ces deux expériences.

CHAPITRE 3

**EXPERIENCE D'IDENTIFICATION DU GENRE
PAR LA VOIX CHEZ DES AUDITEURS
ANGLOPHONES ET FRANCOPHONES**

RESUME

L'étude présentée ici est une double expérience d'identification du genre par la voix, menée conjointement sur 25 auditeurs francophones parisiens avec des stimuli en français, et sur 25 auditeurs anglophones américains avec des stimuli en anglais. Des extraits de mots et pseudo-mots dissyllabiques de type (C)VCV ont été présentés aux participants en utilisant la technique du *gating*. Ces derniers avaient pour tâche d'identifier le genre du locuteur ayant produit le stimulus et d'indiquer leur degré de certitude. Chez les locuteurs des deux langues, les pourcentages de bonnes réponses sont significativement supérieurs au seuil de chance dès la présentation d'une consonne sourde initiale de mot et la présentation d'une voyelle initiale suffit pour obtenir des scores proches des 100 %. Au regard de l'analyse acoustique des stimuli, on constate que les paramètres acoustiques n'ont pas influencé de la même manière les auditeurs des deux langues. Ainsi, la fréquence fondamentale moyenne et la qualité de voix (H1-H2) semblent avoir plus fortement influencé les auditeurs anglophones que les francophones, à l'inverse de la fréquence des formants vocaliques et de la plage de variation de F0.

ABSTRACT

The current study is a gender identification from speech experiment. It was conducted jointly on 25 Parisian French native listeners with French stimuli, and 25 American English listeners with English stimuli. Extracts from (C)VCV words and pseudo-words were played to the participants, by using the gating paradigm. The listeners had to identify the speaker's gender and indicate their degree of certainty. In both languages, percentages of correct identifications were significantly above chance for initial voiceless consonants, and close to 100 % with initial vowels. Considering the acoustic analysis performed on the stimuli, it was found that American English and French listeners did not use the same strategies. Thus, mean F0 and voice quality (H1-H2) had more influence on American English listeners' judgements than on French listeners', contrary to vowel formant frequencies and F0 range.

MOTS-CLES : voix et genre, expérience de perception, identification du genre par la voix, variations inter-langues, français parisien, anglais américains.

KEYWORDS: voice and gender, perceptual experiment, gender identification from speech, cross-language variations, French speakers, American English speakers.

Ce troisième chapitre est dédié à l'expérience de *gating* appliqué à l'identification du genre par la voix, que j'ai menée d'une part avec des stimuli en anglais sur des auditeurs anglophones et d'autre part avec des stimuli en français sur des auditeurs francophones. Une première section est consacrée à la méthode, la seconde présente les résultats, tout d'abord de manière brute, puis à travers leur analyse statistique.

1. METHODE

Dans cette section sont décrits le choix du matériau linguistique et des locuteurs, la procédure d'enregistrement, l'obtention des stimuli, le choix des participants et la procédure expérimentale.

1.1 Matériau linguistique

Du matériau linguistique français et anglais était nécessaire pour entreprendre cette expérience. L'utilisation de mots isolés m'a paru la plus adéquate pour le paradigme du *gating* appliqué à l'identification du genre par la voix, afin d'établir la quantité acoustique nécessaire pour effectuer cette identification. J'ai donc opté pour des *mots* et *pseudo-mots*¹³³ *dissyllabiques*, paradigme qui me permet de tester un nombre important de combinaisons de phonèmes. Du fait de la segmentation réalisée *a posteriori* sur ces items, l'utilisation de phrases aurait généré un nombre de stimuli bien trop important, et n'aurait pas permis de jouer de manière aussi souple sur les combinaisons de phonèmes.

Comme annoncé précédemment, les mots qui serviront ici *sont les mêmes que ceux utilisés dans le chapitre 2* : ils ont ainsi préalablement fait l'objet d'une analyse acoustique, ce qui me permettra de mieux interpréter les résultats de l'expérience d'identification du genre par la voix décrite dans le présent chapitre. Pour rappel, le choix de ces (pseudo)-mots s'est fait sur différents critères : faire apparaître un maximum de consonnes et de voyelles pertinentes tout en limitant le nombre d'items et obtenir une correspondance étroite entre les items français et anglais. Je ne reviendrai pas ici sur les détails des différents choix effectués, *ces derniers ayant déjà été longuement explicités*

¹³³ Items comportant des combinaisons de phonèmes permises par les contraintes phonotactiques d'une langue, mais ne figurant pas dans le lexique de cette dernière.

dans la section 1.1 du chapitre 2. Au final, les combinaisons suivantes, de type CVCV ou VCV¹³⁴, ont été retenues (27 mots dans chacune des deux langues) :

- **Combinaisons de type /C occlusive – V – p – i / :** /tipi/, /tapi/, /tupi/, /dipi/, /dapi/, /dupi/, /kipi/, /kapi/, /kupi/, /gipi/, /gapi/, /gupi/ pour le français, /'ti:pi/, /'tæpi/, /'tu:pi/, /'di:pi/, /'dæpi/, /'du:pi/, /'ki:pi/, /'kæpi/, /'ku:pi/, /'gi:pi/, /'gæpi/, /'gu:pi/ pour l'anglais.
- **Combinaisons de type /C fricative – V – p – i / :** /sipi/, /sapi/, /supi/, /zipi/, /zapi/, /zupi/, /ʃipi/, /ʃapi/, /ʃupi/, /ʒipi/, /ʒapi/, /ʒupi/ pour le français, /'si:pi/, /'sæpi/, /'su:pi/, /'zi:pi/, /'zæpi/, /'zu:pi/, /'ʃi:pi/, /'ʃæpi/, /'ʃu:pi/, /'ʒi:pi/, /'ʒæpi/, /'ʒu:pi/ pour l'anglais.
- **Combinaisons de type /V – p – i / :** /ipi/, /api/, /upi/ pour le français, /'i:pi/, /'æpi/, /'u:pi/ pour l'anglais.

Les 27 mots anglais (lus par des anglophones) seront utilisés pour l'expérience menée sur des auditeurs anglophones, tandis que les 27 mots en français (produits par des francophones) seront utilisés dans l'expérience conduite exclusivement sur des francophones.

1.2 Choix des locuteurs pour les enregistrements

J'ai fait appel à 12 locuteurs au total. Une moitié étant francophone (locuteurs du français parisien¹³⁵), et l'autre moitié anglophone (plus précisément constituée de locuteurs de l'anglais américain originaires d'une même zone du nord-est des Etats-Unis¹³⁶). Les enregistrements des *huit locuteurs du chapitre 2*¹³⁷, qui ont d'ores et déjà fait l'objet d'une analyse acoustique, seront utilisés comme stimuli et pris en compte dans cette expérience. En plus de ces huit locuteurs, j'ai fait appel à deux autres locuteurs dans

¹³⁴ C = consonne, V = voyelle.

¹³⁵ Vivant depuis au moins 10 ans en Ile-de-France et ne présentant pas d'accent régionalement marqué.

¹³⁶ Voir section 1.2 du chapitre 2 pour plus de détails sur le choix des locuteurs anglophones.

¹³⁷ Voir section 1.2 de ce chapitre.

chaque langue, un homme et une femme, dont les enregistrements seront utilisés durant l'expérience uniquement en tant que *warm-ups*¹³⁸. L'ensemble des locuteurs sont monolingues, non-fumeurs et ne présentent aucun trouble de la parole.

Voici les principales caractéristiques pertinentes des différents locuteurs anglophones et francophones :

- **Locutrice francophone numéro 1 (F1FR)** : 27 ans, francophone native, étudiante en sciences du langage niveau M2, F0 moyen : 205 Hz¹³⁹.
- **Locutrice francophone numéro 2 (F2FR)** : 23 ans, francophone native, étudiante en information-communication niveau L3, F0 moyen : 197 Hz.
- **Locuteur francophone numéro 1 (H1FR)** : 23 ans, francophone natif, étudiant en information-communication niveau L2, F0 moyen : 115 Hz.
- **Locuteur francophone numéro 2 (H2FR)** : 24 ans, francophone natif, étudiant en information-communication niveau L3, F0 moyen : 125 Hz.
- **Locutrice francophone pour *warm-ups* (WUFR)** : 22 ans, francophone native, étudiante en sciences du langage niveau L3, F0 moyen : 228 Hz.
- **Locuteur francophone pour *warm-ups* (WUHFR)** : 34 ans, francophone natif, étudiant en sciences du langage niveau M2, F0 moyen : 135 Hz.
- **Locutrice anglophone numéro 1 (FA1)** : 40 ans, anglophone native (anglais américain), originaire¹⁴⁰ de Northampton (Massachusetts), enseignante en anglais, F0 moyen : 195 Hz.
- **Locutrice anglophone numéro 2 (FA2)** : 23 ans, anglophone native (anglais américain), originaire de l'Etat de du Vermont, étudiante en psychologie niveau Master, F0 moyen : 216 Hz.

¹³⁸ Items d'entraînement.

¹³⁹ Les F0 moyens des locuteurs ont tous été mesurés avec le logiciel *Praat*, sur l'ensemble des 54 phrases enregistrées pour cette étude (« il a dit *MOT* deux fois » * 27 mots * 2 lectures : voir section 1.3).

¹⁴⁰ Les locuteurs anglophones ont été enregistrés lorsqu'ils séjournèrent pour une durée plus ou moins longue à Paris.

- **Locuteur anglophone numéro 1 (HA1)** : 39 ans, anglophone natif (anglais américain), originaire de Philadelphie (Pennsylvanie), pasteur en formation, F0 moyen : 116 Hz.
- **Locuteur anglophone numéro 2 (HA2)** : 26 ans, anglophone natif (anglais américain), originaire de Binghamton (Etat de New-York), enseignant en anglais, F0 moyen : 97 Hz.
- **Locuteur anglophone pour *warm-ups* (WUHAN)** : 40 ans, anglophone natif (anglais américain), originaire de Philadelphie (Pennsylvanie), enseignant en anglais, F0 moyen : 136 Hz.
- **Locutrice anglophone pour *warm-ups* (WUFAN)** : 45 ans, anglophone native (anglais américain), originaire de Charlottesville (Virginie), enseignante en anglais, F0 moyen : 214 Hz.

1.3 Procédure d'enregistrement

Cette procédure d'enregistrement est *décrite en détail dans le chapitre 2, section 1.3*. Je ne reviendrai donc ici que sur l'essentiel.

Afin d'homogénéiser les paramètres prosodiques sur les différents items¹⁴¹ devant servir pour l'expérience, chacun d'entre eux a été placé en contexte pour les enregistrements : « *Il a dit MOT deux fois* » pour les mots¹⁴² français, et « *He said MOT twice* » pour les items en anglais, les mots ayant tous été écrits sous forme orthographique¹⁴³. Les locuteurs anglophones ont eu à lire exclusivement les phrases contenant les mots anglais, et les francophones uniquement les phrases contenant les mots français.

Pour limiter tout bruit de fond, les enregistrements ont tous été effectués dans une pièce isolée phoniquement (chambre anéchoïque). Le matériel utilisé est un enregistreur numérique portatif *Edirol R-09HR* de marque Roland, comportant un microphone doté

¹⁴¹ La liste des mots est visible dans la section 1.1.

¹⁴² Par commodité, l'emploi du terme « *mots* » au sens large, c'est-à-dire incluant les *mots* et *pseudo-mots*, sera privilégié dans ce chapitre.

¹⁴³ Voir tableaux 1 et 2 dans la section 1.3 du chapitre 2.

d'une bande passante allant de 20 Hz à 40 kHz. Les enregistrements ont été effectués au format Wave¹⁴⁴.

Voyons maintenant en détail les opérations effectuées sur les enregistrements afin d'obtenir les stimuli de l'expérience.

1.4 Création des stimuli

Afin d'obtenir les stimuli définitifs pour l'expérience de *gating*, chaque mot a dû être extrait des phrases enregistrées, puis segmenté. Ces deux étapes sont décrites dans cette sous-section.

1.4.1 Extraction des mots

Les mots ont dû, dans un premier temps, être extraits de leur contexte « *Il a dit MOT deux fois* » ou « *He said MOT three times* ». Cette opération a été réalisée manuellement dans le logiciel *Praat*. Pour ce faire, je me suis appuyé conjointement sur le spectrogramme, la courbe de F0 et l'onde sonore afin de déterminer précisément les frontières de mots. Une fois ces frontières localisées, j'effectuais la sélection du mot, puis le décalage des bornes gauche et droite de la sélection au passage par zéro le plus proche (opération automatisée grâce à la création d'un script) afin d'éviter les *clics* parasites. Une vérification auditive a permis de confirmer ce découpage.

Chacun des mots a ensuite été exporté dans un fichier au format Wave, en maintenant la fréquence d'échantillonnage à 44 kHz.

1.4.2 Découpage des mots pour l'expérience

Le principe général du *gating* est de segmenter des items en extraits, et de présenter des extraits de plus en plus longs à l'auditeur. Le choix du découpage des mots a donc été un élément crucial de cette étude.

La segmentation des mots en *moments* (consonne 1, consonne 1 plus voyelle 1, etc.) plutôt qu'en extraits de longueur brute équivalente m'a semblé plus pertinente pour

¹⁴⁴ En stéréo (pas d'enregistrement direct en mono possible avec l'*Edirol R-09HR*), avec encodage à 44,1 kHz. Ces fichiers ont ensuite été convertis en mono, en préservant la fréquence d'échantillonnage.

une expérience d'identification du genre par la voix : en effet, les mêmes mots n'étant pas de longueur identique selon les locuteurs l'ayant produit, un découpage temporel brut n'aurait pas été adéquat.

La principale contrainte a été de limiter raisonnablement le nombre de stimuli, qui détermine la durée de l'expérience. Pour rappel, nous avons pour base 27 mots, lus par quatre locuteurs différents dans chaque langue, soit 108 items pour chacune des deux versions de l'expérience (en anglais et en français). A raison d'environ 10 secondes pour traiter chaque stimulus¹⁴⁵, j'ai souhaité limiter le nombre de stimuli à environ 300, pour ne surtout pas dépasser une durée d'expérience égale à une heure. Etant donné que les mots entiers seront bien entendu présentés, cela me laisse donc la possibilité de réaliser au maximum deux *paliers* ou *points de segmentation* par mot.

Tout d'abord, il m'a paru tout à fait indispensable de *présenter isolément la consonne initiale de mot* (que j'appellerai *C1*), afin de tester la capacité des auditeurs à reconnaître le genre par la voix à partir d'une consonne isolée, selon qu'elle est voisée ou non, fricative ou occlusive. Les résultats obtenus à ce point pourront notamment être mis en perspective avec les études de Schwartz (1968) et Whiteside (1998b). Cela constituera donc le premier point de segmentation, et générera 96 stimuli dans chaque langue (24 mots¹⁴⁶ x 4 locuteurs).

Pour le deuxième palier, j'ai opté pour une *segmentation après la première voyelle* (*VI*). Ainsi, pour les 24 mots de type CVCV, cela permettra d'évaluer l'effet de l'ajout d'une voyelle à la consonne 1 sur le taux de bonnes réponses, et de comparer ces résultats selon les différentes combinaisons de consonnes et de voyelles. De plus, grâce aux trois mots de type VCV, ce point de segmentation me permet également de tester l'identification du genre sur des voyelles isolées : l'analyse de ces résultats offrira notamment une possibilité de comparaison avec l'étude menée par Whiteside (1998a). Cette deuxième segmentation implique la création de 108 stimuli supplémentaires pour chaque langue (27 mots x 4 locuteurs).

¹⁴⁵ Voir la procédure expérimentale en 1.6.

¹⁴⁶ Les trois mots débutant directement par une voyelle ([api], [ipi], [upi] pour le français et [æpi], [i:pi], [u:pi] pour l'anglais) ne sont pas concernés.

Enfin, comme je l’ai annoncé précédemment, *les mots entiers seront également présentés* (27 mots x 4 locuteurs soit 108 stimuli). Le total des items expérimentaux s’élève donc à 312 pour chacune des langues. La figure 57 ci-après résume le découpage des mots opéré pour l’expérience.

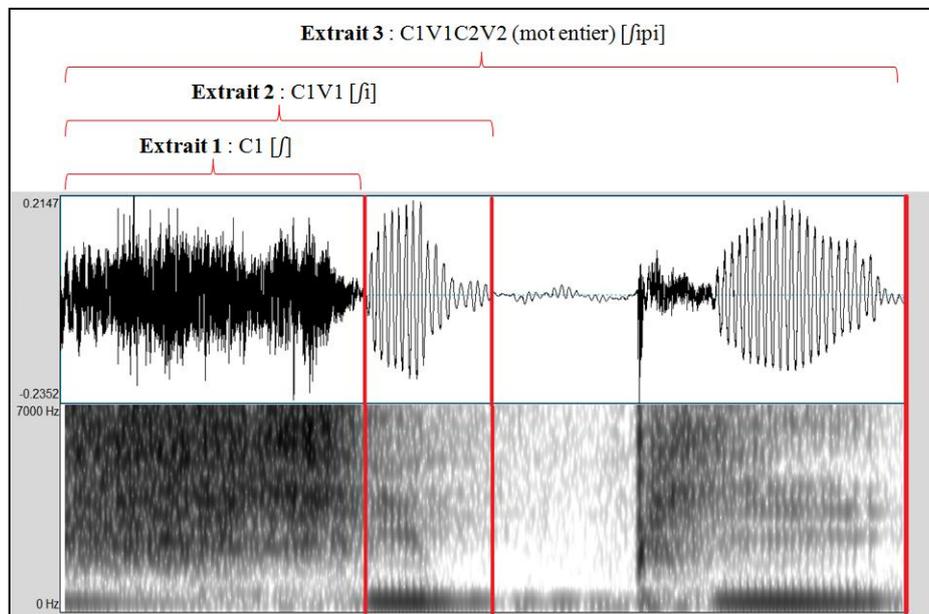


Figure 57 – Représentation schématique des découpages opérés sur chaque mot pour obtenir les stimuli de l’expérience. L’exemple utilisé ici est le mot [ʃipi] produit par la locutrice F2FR : son onde sonore et son spectrogramme, obtenus avec Praat, sont visibles respectivement dans la partie centrale et inférieure de la figure. Les barres verticales indiquent les frontières du découpage.

Pour obtenir les items d’entraînement, seul un mot a été utilisé : le mot [ʃapi] pour l’expérience sur les francophones, et l’équivalent [ʃæpi] pour celle menée sur les anglophones. Ces mots, enregistrés l’un comme l’autre par deux locuteurs (WUFFR et WUHFR pour le mot français, WUFAN et WUHAN pour le mot anglais¹⁴⁷) ont été segmentés de la même manière que les items expérimentaux, donnant ainsi six items par langue. En ajoutant ces items d’entraînement, on arrive par conséquent à un total de 318 stimuli par expérience.

Pour obtenir ces stimuli, j’ai réutilisé les fichiers *TextGrid* réalisés pour l’analyse acoustique du chapitre 2¹⁴⁸ : ces derniers contiennent le découpage précis et l’étiquetage

¹⁴⁷ Voir section 1.2.

¹⁴⁸ Voir section 2.1 du chapitre 2.

en phones de chaque mot. En m'appuyant sur ces fichiers *TextGrid*, j'ai programmé un script permettant d'automatiser l'ensemble des actions suivantes : dans un premier temps, extraire l'élément C1 de chaque mot, l'exporter dans un fichier au format *Wave*, et dans un deuxième temps, extraire les éléments C1 et V1 de chaque mot, les regrouper, exporter le tout dans un fichier au format *Wave*. Ce script a été lancé sur l'ensemble des mots, et m'a ainsi permis de créer 208 fichiers sons par langue, qui ajoutés aux 110 fichiers sons contenant les mots entiers, constitueront les 318 stimuli¹⁴⁹ utilisés pour l'expérience dans chacune des deux langues¹⁵⁰.

1.5 Participants

Les auditeurs ayant pris part à l'expérience peuvent être divisés en deux groupes : un groupe d'anglophones américains natifs, n'ayant pas le français comme autre langue maternelle, et un groupe de francophones natifs (français *parisien*¹⁵¹), n'ayant pas l'anglais comme autre langue maternelle. Tous ne présentent aucun trouble du langage ou de l'audition. Voici de façon sommaire la composition des deux groupes de participants (une description plus détaillée de chacun des participants est présentée dans les annexes 2 et 3) :

- **Groupe des francophones natifs**¹⁵² : 25 participants, 17 femmes et 8 hommes, âgés de 18 à 48 ans. Moyenne d'âge : 24,2 ans (21,9 ans pour les femmes et 29,1 ans pour les hommes).
- **Groupe des anglophones américains natifs**¹⁵³ : 25 participants, 18 femmes et 7 hommes, âgés de 18 à 38 ans. Moyenne d'âge : 24,8 ans (24,3 ans pour les femmes et 26,1 ans pour les hommes).

Le choix des participants à l'expérience a maintenant été exposé. La section suivante décrit la procédure expérimentale.

¹⁴⁹ Items d'entraînement compris.

¹⁵⁰ Les stimuli sont disponibles en ligne. Voir annexe 1.

¹⁵¹ Ces participants vivent en Ile-de-France depuis plus de 5 ans.

¹⁵² Les francophones natifs participent à l'expérience avec les stimuli français.

¹⁵³ Les anglophones américains natifs participent à l'expérience avec les stimuli anglais.

1.6 Procédure expérimentale

La tâche d'identification du genre par la voix (voir Lass et al., 1976 ; Pausewang Gelfer & Mikos, 2005 ; Pépiot, 2010) se caractérise par la présentation de stimuli audio, à partir desquels les participants doivent tenter de reconnaître le genre du locuteur ayant produit ces stimuli. J'ai choisi pour cette étude d'appliquer la technique du *gating*, c'est-à-dire du dévoilement progressif des stimuli (voir Grosjean, 1980). Le participant aura donc pour tâche de reconnaître le genre à partir d'extraits dont la longueur augmentera au fur et à mesure que l'expérience avance. Il devra également accompagner chaque réponse d'un degré de certitude.

Après le découpage des mots, nous avons pour cette étude 312 stimuli expérimentaux et 6 items d'entraînement par langue. Les stimuli expérimentaux se décomposent de la manière suivante :

- **96 items contenant uniquement la C1**¹⁵⁴ ([k], [t], [s], etc.).
- **12 items contenant uniquement la V1**¹⁵⁵. Ces derniers concernent les mots débutant directement par une voyelle : [api], [ipi], [upi] pour le français et [æpi], [i:pi], [u:pi] pour les mots anglais.
- **96 items contenant la C1 plus la V1** ([ka], [ti], [sa], [du], etc.).
- **108 items correspondant aux mots entiers.**

L'ordre de présentation a suivi cette organisation : les quatre sous-ensembles de stimuli ont été présentés dans l'ordre mentionné ci-dessus. A l'intérieur de ces quatre groupes, l'ordre de présentation a été défini en mode aléatoire (et donc différent pour chaque participant). Ainsi, le participant sera dans un premier temps exposé uniquement aux stimuli contenant les consonnes initiales isolées, puis à des voyelles isolées (ces dernières étant censées contenir plus d'informations sur le genre qu'une consonne¹⁵⁶), viennent ensuite les extraits contenant une consonne suivie d'une voyelle, et enfin les

¹⁵⁴ Consonne 1.

¹⁵⁵ Voyelle 1.

¹⁵⁶ Voir Whiteside (1998c) et Pausewang Gelfer & Mikos (2005).

mots complets. Il m'a paru logique de procéder de la sorte afin de bien respecter l'idée de *dévoilement progressif d'informations* : si j'avais opté pour les présentations successives de chaque mot aux trois points de segmentation ([s], [sa], [sapi] ; [k], [ka], [kapi], etc.), l'auditeur aurait obtenu très rapidement des informations suffisantes pour reconnaître aisément les différentes voix utilisées dans l'expérience, ce qui aurait pu fortement biaiser les résultats.

L'expérience a été réalisée à l'aide d'un ordinateur portable et du logiciel *Lancelot 3.0.5.0*¹⁵⁷. La mise en place de l'expérience dans ce logiciel a nécessité la programmation d'un script dédié, et de plusieurs pages HTML¹⁵⁸ associées. Une version anglaise et une version française de l'expérience ont dû être programmées, avec utilisation des stimuli adéquats et traduction du texte devant être affiché. Par ailleurs, je me suis systématiquement assuré de fermer tous les programmes et tâches de fond autres que le logiciel *Lancelot* avant le lancement de l'expérience, dans le but d'éviter tout traitement parallèle qui aurait pu perturber le bon déroulement de l'expérience.

Les passations se sont déroulées dans une pièce calme (salle réservée à cet effet à l'Université Paris 8 pour les francophones, à *AUP*¹⁵⁹ ou *NYU*¹⁶⁰ Paris pour les anglophones), en suivant une procédure rigoureusement identique pour chaque participant. Avant de débiter l'expérience, plusieurs informations sur l'auditeur étaient recueillies et entrées dans le logiciel : son âge, son sexe, ses initiales, sa ou ses langue(s) maternelle(s) et sa profession. Le participant était ensuite invité à s'installer devant l'écran d'ordinateur et à s'équiper d'un casque audio¹⁶¹. Les directives suivantes étaient alors affichées, centrées au milieu à l'écran :

¹⁵⁷ Voir Ghio et al. (2003) et André et al. (2003). Le logiciel est inclus dans le *package Perceval*, téléchargeable gratuitement à l'adresse suivante :

<http://aune.lpl.univ-aix.fr/~lpldev/perceval/percevaldownload.html>

¹⁵⁸ Ce script est visible en annexe 6, de même que le lien permettant de le télécharger directement au format adapté au logiciel *Lancelot*, accompagné des fichiers HTML associés.

¹⁵⁹ *American University of Paris*.

¹⁶⁰ *New York University*.

¹⁶¹ Casque fermé de marque *Sony*, disposant d'une bande passante de 5 Hz à 28 kHz.

- Version française :

« Vous allez entendre des extraits d'enregistrements sonores de différentes personnes. Ces extraits peuvent être de très courte durée. Dans tous les cas, ils seront **présentés deux fois**.

Votre tâche sera de **deviner s'il s'agit d'une voix de femme ou d'une voix d'homme**. Vous devrez également accompagner votre réponse d'un **degré de certitude**, sur une échelle allant de 0 (pas sûr du tout) à 7 (tout à fait sûr).

Vous validerez vos réponses en **cliquant sur le bouton présent en bas de page**.

Avant de commencer, quelques exemples vous seront proposés afin de vous familiariser avec cette tâche.

Cliquez ici pour commencer »

- Version anglaise :

« You will hear extracts from audio recordings obtained from different speakers. These extracts may be very short. In any case, each of them will be **played twice**.

You have to **guess if the voice was a male or a female one**. Then you have to associate with your answer a **degree of certainty**, on a scale going from 0 (not sure at all) to 7 (completely sure).

To confirm your answers, you will need to **click on the button at the bottom of the screen**.

Before starting the actual experiment, a training session will take place in order to get you familiar with the task.

Click here to start »

Après m'être assuré que la consigne avait été parfaitement comprise, la phase d'entraînement, qui contient 6 items, était lancée. La procédure pour chaque item était identique pour toute l'expérience, y compris durant cette phase d'entraînement :

- Affichage d'un écran vierge.
- Présentation d'un son pur à 200 Hz d'une durée de 200 ms avec *fade in* et *fade out*. Ce *bip* a pour but de relever l'attention du participant avant la double présentation du stimulus.
- Après un délai de 700 ms, première présentation du stimulus.

- Après un délai d'environ 1000 ms¹⁶², deuxième présentation du stimulus.
- A la fin de cette deuxième présentation, le formulaire permettant de collecter les réponses du participant apparaît. Une capture d'écran de ce dernier est présentée dans la figure 58 ci-dessous.

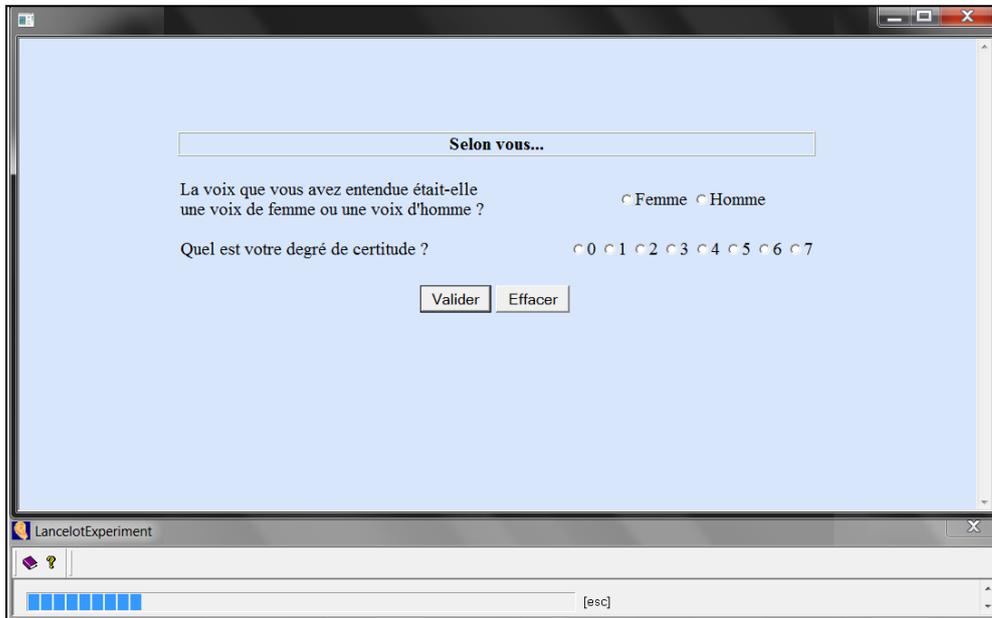


Figure 58 – Capture d'écran de l'interface visuelle de l'expérience d'identification du genre par la voix, dans sa version française. Dans la partie supérieure se trouve le dispositif de réponse et la jauge figurant en bas de l'écran indique l'état d'avancement de l'expérience.

- Le participant doit alors répondre à la question « *Selon vous... La voix que vous avez entendue était-elle une voix de femme ou une fois d'homme ?*¹⁶³ » en cochant la réponse de son choix *via* le curseur de la souris, et indiquer son degré de certitude sur une échelle allant de 0 à 7, en cliquant là encore sur la case appropriée.
- Le participant peut valider ses réponses en cliquant sur le bouton *Valider* (*Confirm*) situé en bas de l'écran. Il peut également remettre à zéro le formulaire s'il s'est trompé lors de la saisie de ses réponses, en cliquant sur le bouton *Effacer*

¹⁶² Le délai est légèrement variable car il doit être fixé dans le script pour toute l'expérience et qu'il est calculé entre le début du premier stimulus (et non la fin) et le début du second. Sa valeur est de 1500 ms : les stimuli les plus longs (mot entiers) étant d'une longueur d'environ 500 ms, cela laisse un silence minimal entre les deux présentations d'environ 1000 ms.

¹⁶³ La traduction anglaise utilisée pour cette question est la suivante : « *According to you... Is it a female or a male voice?* ».

(Erase). Une fois ses réponses validées, l'ensemble de la procédure décrite ici recommence pour l'item suivant. Cette procédure de validation par le participant lui permet de moduler le rythme de l'expérience à sa guise. S'il souhaite le ralentir, il n'a qu'à patienter quelques secondes avant de cliquer sur le bouton de validation, et inversement.

Après la phase d'entraînement, qui contient 6 items, l'expérience était automatiquement suspendue : un écran contenant le texte « *Nous allons commencer l'expérience. Cliquez ici pour démarrer.*¹⁶⁴ » était affiché. A cet instant, je m'assurais auprès du participant que ce dernier ne rencontrait aucune difficulté particulière et n'avait pas de question. Une fois les éventuels éclaircissements apportés, le participant était invité à continuer l'expérience, en sachant qu'il ne pourrait plus me poser de question avant la fin de celle-ci.

Enfin, après la présentation des 312 items expérimentaux, le message de remerciement suivant s'affichait à l'écran : « *L'expérience est maintenant terminée. Merci de votre participation !*¹⁶⁵ ». La durée totale de l'expérience est d'environ 45 minutes, ce temps variant sensiblement d'un participant à l'autre en fonction du rythme de réponse adopté par ce dernier.

La méthodologie de l'expérience a maintenant été exposée. Dans la section suivante, je décris succinctement l'analyse des données.

¹⁶⁴ Ce message est défini automatiquement par le logiciel et n'est pas modifiable. Le texte correspondant pour la version anglaise est le suivant : « *Test step. Click here to start.* »

¹⁶⁵ Texte correspondant pour l'expérience menée sur les anglophones : « *The experiment is over. Thanks for your participation!* »

2. ANALYSE DES DONNEES

A la fin de chaque passation, les réponses des participants (« voix d'homme » / « voix de femme » et degré de certitude), pour l'ensemble des stimuli, ont été automatiquement inscrites par *Perceval* dans un fichier au format compatible avec le logiciel *Excel*.

Grâce au script programmé pour l'expérience, les réponses données par les auditeurs sur le type de voix (« voix de femme » ou « voix d'homme ») ont été systématiquement comparées avec la réponse correcte : si les deux concordent, la réponse du locuteur est catégorisée comme « *ok* », dans le cas contraire, c'est la mention « *err* » qui apparaît. Ces deux termes ont par la suite été respectivement remplacés par « *100* » et « *0* », afin de faciliter leur futur traitement statistique.

Hormis les items d'entraînement, toutes les réponses sans exception ont été prises en compte dans les résultats. Au total, pour les 25 auditeurs francophones ayant pris part à l'expérience, ce sont 7800 catégorisations « voix de femmes » / « voix d'hommes » qui ont été recueillies, avec les 7800 degrés de certitude correspondants. La même quantité de données a été recueillie avec l'expérience conduite sur les anglophones, soit 7800 catégorisations et 7800 degrés de certitude pour l'ensemble des 25 participants américains.

Les résultats bruts ainsi que leur analyse statistique sont présentés dans la section suivante.

3. RESULTATS

Dans un premier temps, les résultats des expériences sont présentés de manière brute et synthétique¹⁶⁶. Une deuxième partie est consacrée à leur analyse statistique. Les résultats sont donnés séparément pour chacun des deux groupes d'auditeurs (francophones et anglophones) dans leur ensemble, c'est-à-dire sans distinction de genre. Il ne semble en effet pas pertinent de traiter isolément les auditeurs masculins et féminins car aucune étude n'a montré de différences significatives hommes-femmes dans des tâches d'identification du genre par la voix (voir Pépiot, 2009). Par ailleurs, les résultats seront au fur et à mesure mis en perspective avec les données obtenues *via* l'analyse acoustique effectuée préalablement sur les stimuli dans le chapitre 2. Tous les nombres donnés dans cette section ont été arrondis à 10^{-2} .

3.1 Résultats bruts

Les résultats bruts des deux expériences d'identification du genre par la voix, menées respectivement sur des auditeurs francophones et anglophones, sont exposés dans cette sous-section. Après une synthèse globale pour l'ensemble des types de stimuli, les résultats seront présentés de manière plus détaillée pour les consonnes initiales, les voyelles initiales, les combinaisons « consonne + voyelle » puis les mots entiers.

3.1.1 Tous stimuli

Les pourcentages d'identifications réussies ainsi que les degrés de certitude moyens¹⁶⁷ obtenus sur les auditeurs francophones pour chacune des cinq grandes catégories de stimuli sont présentés dans le tableau 31, ci-après.

¹⁶⁶ Les fichiers contenant les résultats détaillés, participant par participant, sont disponibles en ligne : voir annexe 9.

¹⁶⁷ Pour rappel, le degré de certitude est exprimé sur une échelle allant de 0 à 7.

Tableau 31 – *Pourcentage d'identifications réussies et degré de certitude moyen pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre par la voix menée sur les francophones.*

<i>Type de stimulus</i>	N¹⁶⁸ d'items	N de réponses correctes	N de réponses incorrectes	Pourcentage d'identifications réussies	Degré de certitude moyen¹⁶⁹
C initiale sourde	1200	788	412	65,67	3,56
C initiale voisée	1200	1134	66	94,50	4,76
V initiale	300	294	6	98,00	6,54
C initiale + V	2400	2358	42	98,25	6,64
Mot dissyllabique	2700	2698	2	99,93	6,92
Tous types	7800	7272	528	93,23	5,97

On note un assez fort pourcentage d'identifications réussies dès la présentation d'une consonne initiale sourde (supérieur à 65 %), en dépit d'un degré de certitude relativement bas (3,56). Lorsque l'indice supplémentaire que constitue le F0 apparaît (consonnes initiales voisées), le score augmente fortement, pour passer à 94,5 %. En revanche, le degré de certitude moyen s'améliore dans une bien moindre mesure, en passant à 4,76 / 7. Le pourcentage d'identifications réussies augmente à nouveau quand une voyelle est ajoutée aux consonnes initiales (plus de 98 %), tout comme le degré de certitude (6,64), pour atteindre un score frôlant les 100 % et un degré de certitude proche du maximum (6,92) sur les mots dissyllabiques. La présentation isolée d'une voyelle initiale entraîne quant à elle un pourcentage d'identifications correctes (98 %) et un degré de certitude moyen (6,54) extrêmement élevés : ces chiffres sont supérieurs à ceux obtenus sur les consonnes isolées.

Les résultats correspondants pour l'expérience menée sur les auditeurs anglophones américains figurent dans le tableau 32.

¹⁶⁸ L'abréviation « N » correspond à « nombre ».

¹⁶⁹ Pour rappel, le degré de certitude a été exprimé sur une échelle allant de 0 à 7.

Tableau 32 – Pourcentage d'identifications réussies et degré de certitude moyen pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre par la voix menée sur les anglophones.

Type de stimulus	N d'items	N de réponses correctes	N de réponses incorrectes	Pourcentage d'identifications réussies	Degré de certitude moyen
C initiale sourde ¹⁷⁰	1800	1377	423	76,50	2,85
C initiale voisée	600	596	4	99,33	6,00
V initiale	300	293	7	97,67	6,33
C initiale + V	2400	2364	36	98,50	6,39
Mot dissyllabique	2700	2684	16	99,41	6,54
Tous types	7800	7314	486	93,77	5,60

Pour les auditeurs anglophones américains, le pourcentage d'identifications réussies est d'ores et déjà très élevé sur les consonnes sourdes (76,5 %), avec un score sensiblement supérieur à celui obtenu par les francophones pour les stimuli de même type. Le degré de certitude moyen est en revanche très faible (2,85). La présence de voisement sur les consonnes initiales fait augmenter de manière particulièrement forte le pourcentage d'identifications correctes, qui frôle les 100 % pour cette catégorie de stimuli (99,33 %). Le degré de certitude moyen passe quant à lui à 6 / 7, soit un niveau nettement plus élevé que celui des auditeurs francophones pour les items équivalents. La présence d'une voyelle à la suite de la consonne (séquence C + V) fait encore monter légèrement ce degré de certitude (qui atteint 6,33) et maintient le score d'identifications correctes à un niveau proche du maximum (98,5 %). Lorsque les mots dissyllabiques sont présentés dans leur intégralité, le degré de certitude progresse de nouveau (6,54 / 7) et le pourcentage de bonnes réponses plafonne à 99,41 % : ces tendances sont similaires à celles observées chez les francophones. Enfin, concernant les voyelles isolées, on note que contrairement aux auditeurs francophones, bien que le degré de certitude soit très légèrement en hausse par rapport aux consonnes voisées, le pourcentage de bonnes réponses, déjà très élevé, ne progresse pas : il est même légèrement plus faible (97,67 %).

¹⁷⁰ Dans les résultats de l'expérience conduite sur les anglophones américains, les consonnes dévoisées [d̥] et [t̥] sont classées dans la catégorie des consonnes initiales *sourdes*. Voir section 2.3 du chapitre 2.

Le tableau 33, ci-après, donne les résultats pour les auditeurs francophones, sur les cinq grandes catégories d'items, en fonction du locuteur ayant produit les stimuli (H1FR, H2FR, F1FR, F2FR).

Tableau 33 – Répartition des réponses incorrectes pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre menée sur les francophones en fonction des locuteurs ayant produit les stimuli.

<i>Type de stimulus</i>	N de réponses correctes	N de réponses incorrectes selon voix						
		F1	F2	Total F	H1	H2	Total H	Total toutes voix
C initiale sourde	788	69	99	168	169	75	244	412
C initiale voisée	1134	29	27	56	5	5	10	66
V initiale	294	0	2	2	0	4	4	6
C initiale + V	2358	5	29	34	4	4	8	42
Mot dissyllabique	2698	0	1	1	1	0	1	2
Tous types	7272	103	158	261	179	88	267	528

Sur l'ensemble de l'expérience, le nombre d'erreurs commises est donc très similaire pour les voix de femmes (261) et pour les voix d'hommes (267). Cependant, ce n'est pas le cas si l'on considère une à une les cinq catégories de stimuli. Ainsi, pour les consonnes sourdes présentées isolément, les voix d'hommes ont entraîné un plus grand nombre de réponses erronées, tout particulièrement la voix H1FR qui a entraîné à elle seule 169 erreurs sur les 244 commises au total. Concernant les consonnes voisées, ce sont sur les voix de femmes qu'ont porté la grande majorité des réponses incorrectes (56 sur 66), ces erreurs étant réparties de manière équivalente sur les deux voix féminines (29 pour F1FR, 27 pour F2FR). Enfin, notons que pour les combinaisons « consonne + voyelle », les réponses erronées ont été majoritairement données sur les voix de femmes (34 sur un total de 42), et plus précisément sur les stimuli produits par la locutrice F2FR (29 erreurs à eux seuls).

Au regard de l'analyse acoustique effectuée dans le chapitre 2, on constate que *les consonnes sourdes produites par le locuteur H1FR*, qui ont entraîné la majorité des identifications incorrectes commises sur ce type de stimulus, *ont leur pic spectral et leur*

centre de gravité dans des fréquences nettement plus élevées que celles produites par l'autre locuteur masculin¹⁷¹. D'autre part, pour ce qui est des nombreuses erreurs portant sur les combinaisons « consonne + voyelle » produites par la locutrice F2FR, il apparaît que les voyelles produites par cette femme possèdent des valeurs formantiques globalement plus basses que celles produites par l'autre locutrice, en particulier sur la voyelle [u]¹⁷².

Les résultats en fonction des différentes voix pour l'expérience conduite sur les anglophones figurent dans le tableau 34.

Tableau 34 – Répartition des réponses incorrectes pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre menée sur les anglophones en fonction des locuteurs ayant produit les stimuli.

Type de stimulus	N de réponses correctes	N de réponses incorrectes selon voix						
		F1	F2	Total F	H1	H2	Total H	Total toutes voix
C initiale sourde	1377	77	54	131	106	186	292	423
C initiale voisée	596	4	0	4	0	0	0	4
V initiale	293	7	0	7	0	0	0	7
C initiale + V	2364	17	17	34	0	2	2	36
Mot dissyllabique	2684	13	3	16	0	0	0	16
Tous types	7314	118	74	192	106	188	294	486

Chez les auditeurs anglophones américains, la grande majorité des erreurs commises sur les items de type « consonne initiale sourde » portent sur des voix d'hommes et se concentrent principalement sur les consonnes produites par le locuteur H2AN. A l'inverse, sur toutes les autres catégories de stimuli (consonnes voisées, voyelles, séquence C + V et mot dissyllabiques) ce sont les voix de femmes qui ont entraîné la quasi-totalité des erreurs d'identification (61 sur un total de 63) : cela rejoint les tendances observées auprès des auditeurs francophones. Enfin, tous stimuli confondus,

¹⁷¹ Voir section 3.1.1.1 du chapitre 2 pour plus de détails.

¹⁷² Voir section 3.1.2.1 du chapitre 2.

on constate que les stimuli produits par la locutrice F1AN ont entraîné sensiblement plus de réponses erronées que ceux produits par F2AN.

La mise en parallèle avec l'analyse acoustique des stimuli effectuée dans le chapitre 2 met évidence certains éléments intéressants. Tout d'abord, on note que *les consonnes sourdes produites par le locuteur H2AN*, sur lesquelles ont porté près de la moitié des erreurs d'identification pour ce type de stimulus, *ont leur pic spectral et leur centre de gravité dans des fréquences globalement plus élevées que celles produites par H1AN*¹⁷³. On constate par ailleurs que *la locutrice F1AN*, dont les stimuli ont entraîné de manière générale plus de réponses erronées que ceux produits par son homologue F2AN, *présente un F0 globalement plus bas que cette dernière*¹⁷⁴, *une différence H1-H2 plus faible sur les voyelles ouvertes*¹⁷⁵ (i.e. *une voix moins breathy*) *ainsi que des consonnes initiales ayant un pic spectral et un centre de gravité situés en moyenne dans des fréquences plus basses*¹⁷⁶.

3.1.2 Consonnes initiales isolées

Les résultats détaillés en fonction du type de consonne initiale présenté aux auditeurs francophones sont visibles dans le tableau 35, ci-après.

¹⁷³ Voir section 3.1.1.1 du chapitre 2.

¹⁷⁴ Voir section 3.1.3.1 du chapitre 2.

¹⁷⁵ Voir section 3.1.2.2 du chapitre 2.

¹⁷⁶ Voir section 3.1.1.1 du chapitre 2.

Tableau 35 – Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de consonnes initiales présentées dans l'expérience d'identification du genre menée sur les francophones.

Type de stimulus	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
[t]	183	43	74	117	61,00	2,82
[k]	186	44	70	114	62,00	3,01
Occ. sourdes	369	87	144	231	61,50	2,92
[s]	227	7	66	73	75,67	4,46
[ʃ]	192	74	34	108	64,00	3,94
Fric. sourdes	419	81	100	181	69,83	4,20
Toutes cons. sourdes	788	168	244	412	65,67	3,56
[d]	279	18	3	21	93,00	4,10
[g]	277	17	6	23	92,33	4,05
Occ. voisées	556	35	9	44	92,67	4,08
[z]	296	4	0	4	98,67	5,64
[ʒ]	282	17	1	18	94,00	5,25
Fric. voisées	578	21	1	22	96,33	5,45
Toutes cons. voisées	1134	56	10	66	94,50	4,76
Toutes cons.	1922	224	254	478	80,08	4,16

On constate que les pourcentages d'identifications réussies, tout comme les degrés de certitude, sont nettement plus élevés sur les consonnes voisées (96,33 % et 4,76) que sur les consonnes sourdes (65,67 % et 3,56). Parmi les segments voisés, le score maximal est atteint sur la consonne [z], avec 98,67 % d'identifications correctes et un degré de certitude moyen de 5,64 / 7. Pour les consonnes non-voisées, c'est le [s] qui permet d'obtenir les résultats les plus élevés : 75,67 % d'identifications réussies et un degré de certitude moyen de 4,46.

Les meilleurs scores pour les consonnes voisées et les consonnes sourdes sont donc atteints sur des fricatives. L'analyse acoustique du chapitre précédent montre que *la durée de ces dernières est nettement supérieure à celle des occlusives*¹⁷⁷. Par ailleurs, les consonnes de type [s], qui obtiennent le meilleur score des sons non-voisés, *ont leur pic spectral et leur centre de gravité dans des fréquences nettement plus élevées pour les voix*

¹⁷⁷ Voir section 3.1.1.4 du chapitre 2.

de femmes que pour les voix d'hommes¹⁷⁸. Enfin, les consonnes [z], obtenant le meilleur taux d'identification du genre toutes consonnes confondues, *présentent des différences femmes-hommes extrêmement élevées au niveau du centre de gravité spectral*, qui se situe en moyenne dans des fréquences plus de deux fois supérieures chez les locutrices.

Le tableau 36, ci-après, contient les résultats équivalents pour l'expérience menée sur les anglophones américains.

Tableau 36 – Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de consonnes initiales présentées dans l'expérience d'identification du genre menée sur les anglophones.

<i>Type de stimulus</i>	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
[t ^h]	254	19	27	46	84,67	3,48
[k ^h]	219	26	55	81	73,00	2,94
Occ. aspirées	473	45	82	127	78,83	3,21
[d]	198	34	68	102	66,00	1,56
[g]	194	38	68	106	64,67	1,53
Occ. non-aspirées	392	72	136	208	65,33	1,55
[s]	248	3	49	52	82,67	3,86
[ʃ]	264	11	25	36	88,00	3,75
Fric. sourdes	512	14	74	88	85,33	3,81
Toutes cons. sourdes	1377	131	292	423	76,50	2,85
[z]	297	3	0	3	99,00	5,87
[ʒ]	299	1	0	1	99,67	6,14
Toutes cons. (fric.) voisées	596	4	0	4	99,33	6,00
Toutes cons.	1973	135	292	427	82,21	3,64

¹⁷⁸ Voir section 3.1.1.1 du chapitre 2.

Chez les anglophones, les résultats sont également bien meilleurs sur les consonnes voisées (99,33 % d'identifications réussies avec un degré de certitude de 6 / 7) que sur les consonnes sourdes (76,5 % et 2,85). En termes de pourcentage d'identifications correctes, les résultats obtenus sur les consonnes sourdes sont supérieurs à ceux observés chez les francophones, en particulier sur les fricatives sourdes, pour lesquelles plus 85 % des jugements sont corrects. Parmi ces consonnes sourdes, le score le plus élevé est atteint sur la fricative post-alvéolaire [ʃ] avec 88 % de bonnes réponses (pour un degré de certitude moyen de 3,75). Concernant les segments voisés, c'est sa contrepartie [ʒ] qui obtient les meilleurs résultats (99,67 % et 6,14 / 7).

Comme chez les auditeurs francophones, les meilleurs taux d'identifications correctes pour les consonnes sourdes sont atteints sur les fricatives. Ici encore, on constate que *leur durée moyenne est sensiblement supérieure à celle des occlusives*¹⁷⁹. D'autre part, les réalisations de la fricative [ʃ], qui obtiennent les meilleurs scores des consonnes sourdes, *présentent une très forte différence femmes-hommes au plan du pic spectral et du centre de gravité*¹⁸⁰. Notons enfin que les stimuli de type [ʒ], qui ont également entraîné des scores d'identifications très élevés, *présentent une différence inter-genres particulièrement importante sur leur pic spectral*.

3.1.3 Voyelles initiales isolées

Les résultats détaillés pour les auditeurs francophones sur les voyelles initiales isolées sont présentés dans le tableau 37, ci-après.

¹⁷⁹ Voir section 3.1.1.4 du chapitre 2.

¹⁸⁰ Voir section 3.1.1.1 du chapitre 2.

Tableau 37 – Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de voyelles initiales présentées dans l'expérience d'identification du genre menée sur les francophones.

Type de stimulus	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
[i]	100	0	0	0	100,00	6,74
[a]	100	0	0	0	100,00	6,60
[u]	94	2	4	6	94,00	6,28
Toutes voyelles	294	2	4	6	98,00	6,54

Les pourcentages d'identifications correctes tout comme les degrés de certitude sont très élevés pour les trois voyelles testées. Sur le [i] et le [a], le score d'identifications réussies atteint 100 %, avec un degré de certitude sensiblement plus élevé pour le [i] (6,74) que pour le [a] (6,60). Seule la voyelle [u] a entraîné quelques réponses erronées, faisant descendre le pourcentage d'identifications correctes à 94 %. Le degré de certitude moyen sur cette voyelle est également plus faible, tout en restant à un niveau élevé : il atteint 6,28 / 7.

En se référant à l'analyse acoustique du chapitre 2, il apparaît que chez les francophones, la voyelle [u] est celle qui présente globalement le moins de différences formantiques hommes-femmes¹⁸¹. Cela est particulièrement vrai pour le deuxième formant, sur lequel les différences inter-genres sont très faibles en comparaison des autres voyelles.

Les résultats correspondants pour l'expérience menée sur les auditeurs anglophones américains figurent dans le tableau 38.

¹⁸¹ Voir section 3.1.2.1 du chapitre 2.

Tableau 38 – Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de voyelles initiales présentées dans l'expérience d'identification du genre menée sur les anglophones.

Type de stimulus	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
[i:]	99	1	0	1	99,00	6,53
[æ]	99	1	0	1	99,00	6,34
[u:]	95	5	0	5	95,00	6,13
Toutes voyelles	293	7	0	7	97,67	6,33

Chez les auditeurs anglophones américains, les pourcentages d'identifications correctes et les degrés de certitude sont également très élevés pour les trois voyelles. Une seule erreur d'identification a été commise sur les voyelles [i:] et [æ]. La voyelle [u:] a entraîné quelques erreurs d'identifications de plus, avec un score de 95 % de bonnes réponses. Concernant les degrés de certitude moyens, la hiérarchie est identique à celle observée chez les francophones : c'est la voyelle fermée antérieure qui présente le score le plus élevé (6,53), devant la voyelle ouverte (6,34) et la voyelle fermée postérieure (6,13).

En recoupant ces résultats avec l'analyse acoustique réalisée sur les stimuli, on remarque que *la voyelle [u:], malgré d'importantes différences inter-genres, présente une forte variabilité sur le plan du deuxième formant*¹⁸².

3.1.4 Combinaisons « consonne + voyelle »

Les résultats détaillés pour les différentes combinaisons « consonne + voyelle » présentées aux auditeurs francophones sont visibles dans le tableau 39, ci-après.

¹⁸² Voir section 3.1.2.1 du chapitre 2.

Tableau 39 – Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de combinaisons « consonne + voyelle » présentées dans l'expérience d'identification du genre par la voix menée sur les francophones.

Type de stimulus	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
C sourde + [i]	399	1	0	1	99,75	6,73
C voisée + [i]	399	0	1	1	99,75	6,74
Toutes cons. + [i]	798	1	1	2	99,75	6,74
C sourde + [a]	393	5	2	7	98,25	6,64
C voisée + [a]	397	2	1	3	99,25	6,69
Toutes cons. + [a]	790	7	3	10	98,75	6,67
C sourde + [u]	382	14	4	18	95,50	6,45
C voisée + [u]	388	12	0	12	97,00	6,61
Toutes cons. + [u]	770	26	4	30	96,25	6,53
Toutes combinaisons C + V	2358	34	8	42	98,25	6,64

Les scores d'identifications du genre réussies tout comme les degrés de certitude atteignent ici encore des niveaux extrêmement élevés, avec 98,25 % de réponses correctes et un degré de certitude moyen de 6,64 / 7, toutes combinaisons confondues. On remarque néanmoins quelques disparités en fonction de la voyelle qui figure après la consonne. Ainsi, les combinaisons *C + [i]* entraînent le meilleur pourcentage d'identifications réussies, avec 99,25 %, et le degré de certitude moyen le plus haut (6,74). Viennent ensuite les combinaisons *C + [a]*, avec respectivement 98,75 % et 6,67, puis les combinaisons *C + [u]*, qui présentent le plus faible score d'identifications réussies pour cette catégories de stimuli (96,25 %), et le degré de certitude moyen le moins élevé (6,53).

Notons que le type de combinaison qui entraîne le plus d'identifications erronées et le degré de certitude le plus bas contient la voyelle [u]. Cette dernière est, comme nous

l'avons vu précédemment, *celle qui présente les plus faibles différences formantiques femmes-hommes, en particulier sur le deuxième formant*¹⁸³.

Le tableau 40, ci-après, contient les résultats équivalents pour l'expérience menée sur les auditeurs anglophones.

Tableau 40 – Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de combinaisons « consonne + voyelle » présentées dans l'expérience d'identification du genre par la voix menée sur les anglophones.

<i>Type de stimulus</i>	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
C sourde + [i:]	591	9	0	9	98,50	6,36
C voisée + [i:]	200	0	0	0	100,00	6,49
Toutes cons + [i:]	791	9	0	9	98,88	6,43
C sourde + [æ]	589	9	2	11	98,17	6,43
C voisée + [æ]	195	5	0	5	97,50	6,39
Toutes cons. + [æ]	784	14	2	16	98,00	6,41
C sourde + [u:]	592	8	0	8	98,67	6,33
C voisée + [u:]	197	3	0	3	98,50	6,48
Toutes cons. + [u:]	789	11	0	11	98,63	6,41
Toutes combinaisons C + V	2364	34	2	36	98,50	6,41

A l'instar des francophones, les scores obtenus pour les auditeurs anglophones sont eux aussi particulièrement élevés pour les stimuli de type C + V. Globalement, le pourcentage d'identifications réussies atteint 98,5 %, avec un degré de certitude de 6,41. Les disparités entre les différentes combinaisons semblent ici moins fortes que chez les auditeurs francophones. Ainsi, les combinaisons C + [i:] ont donné 98,88 % de réponses correctes, contre 98,63 % pour les combinaisons C + [u:] et 98 % pour celles de type

¹⁸³ Voir section 3.1.2.1 du chapitre 2.

C + [æ]. Les degrés de certitude pour ces trois catégories sont eux aussi très proches, allant de 6,41 à 6,43.

On constate qu’au niveau acoustique, *les variations inter-voyelles des différences femmes-hommes semblent également moins fortes que chez les locuteurs francophones*¹⁸⁴.

3.1.5 Mots dissyllabiques

Les résultats détaillés pour les auditeurs francophones sur mots dissyllabiques sont présentés dans le tableau 41, ci-après.

Tableau 41 – *Pourcentage d’identifications réussies et degré de certitude moyen pour les différents types de mot dissyllabique présentés dans l’expérience d’identification du genre par la voix menée sur les francophones.*

<i>Type de stimulus</i>	N de réponses correctes	N de réponses incorrectes			Pourcentage d’identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
[i] + [pi]	100	0	0	0	100,00	6,93
[a] + [pi]	100	0	0	0	100,00	6,93
[u] + [pi]	100	0	0	0	100,00	6,87
C + [i] + [pi]	799	1	0	1	99,88	6,92
C + [a] + [pi]	800	0	0	0	100,00	6,93
C + [u] + [pi]	799	0	1	1	99,88	6,91
Tous mots	2698	1	1	2	99,93	6,92

Sur les mots dissyllabiques présentés dans leur intégralité, le pourcentage d’identifications du genre correctes frôle les 100 % chez les auditeurs francophones : seules deux réponses erronées ont été émises sur un total de 2700. Le degré de certitude moyen est également proche du maximum, avec une moyenne de 6,92 / 7. Ce dernier est légèrement plus faible sur les items contenant la voyelle [u], avec 6,87 sur la combinaison [u] + [pi], et 6,91 sur les combinaisons de type C + [u] + [pi], ce qui concorde avec les résultats observés sur les stimuli plus courts (voyelle isolée et C + V) contenant cette même voyelle [u].

¹⁸⁴ Voir section 3.1.2.1 du chapitre 2.

Les résultats détaillés sur les mots dissyllabiques pour l'expérience menée sur les auditeurs anglophones américains figurent dans le tableau 42.

Tableau 42 – *Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de mot dissyllabique présentés dans l'expérience d'identification du genre par la voix menée sur les anglophones.*

<i>Type de stimulus</i>	N de réponses correctes	N de réponses incorrectes			Pourcentage d'identifications réussies	Degré de certitude moyen
		Sur voix F	Sur voix H	Total		
[i:] + [pi]	100	0	0	0	100,00	6,52
[æ] + [pi]	99	1	0	1	99,00	6,55
[u:] + [pi]	100	0	0	0	100,00	6,45
C + [i:] + [pi]	797	3	0	3	99,63	6,51
C + [æ] + [pi]	790	10	0	10	98,75	6,53
C + [u:] + [pi]	798	2	0	2	99,75	6,60
Tous mots	2684	16	0	16	99,41	6,54

Pour les auditeurs anglophones, le pourcentage d'identifications correctes est également au plus haut (99,41 %) lorsque les mots dissyllabiques sont présentés dans leur intégralité. Il en va de même pour le degré de certitude moyen, qui atteint 6,54 / 7. Notons que le score le plus faible est obtenu sur la combinaison *C + [æ] + [pi]*, avec 98,75 %, soit 10 erreurs d'identifications sur les 16 commises au total sur l'ensemble des mots dissyllabiques : cela rejoint les résultats observés sur les stimuli de type *C + V*.

Les résultats bruts ont maintenant été présentés. La section suivante est dédiée à leur analyse statistique.

3.2 Analyse statistique

Pour vérifier si les différences observées dans la section précédente sont significatives, et afin de mettre en relation ces données avec l'analyse acoustique du chapitre 2, j'ai procédé à divers tests statistiques à l'aide du logiciel *StatView*¹⁸⁵.

3.2.1 Tous stimuli

J'ai conduit une ANOVA à un facteur (« type de stimulus ») sur les pourcentages d'identifications correctes obtenus par les auditeurs francophones lors de l'expérience d'identification du genre par la voix. Le graphique correspondant à cette analyse est présenté ci-après (figure 59). Dans tous les graphiques de cette section, les pourcentages d'identifications correctes sont présentés avec un seuil minimum fixé à 50 %, qui correspond au seuil de chance (question impliquant un choix forcé entre deux items).

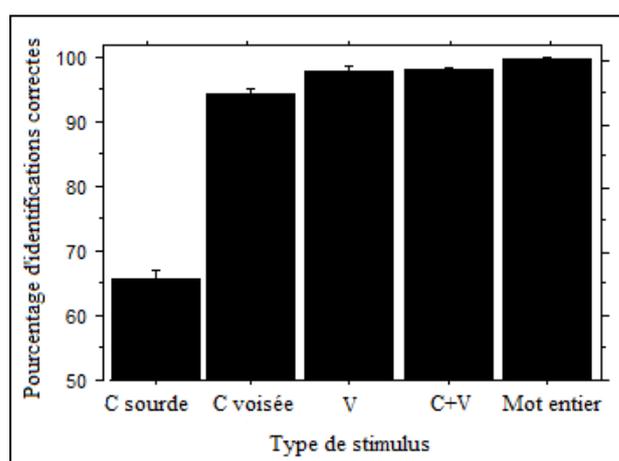


Figure 59 – Pourcentage d'identifications du genre correctes sur les cinq grandes catégories de stimuli pour les auditeurs francophones, avec les barres d'erreur.

Le résultat de cette analyse fait état d'un effet global très significatif du facteur « type de stimulus » sur le pourcentage d'identifications du genre correctes ($F(4,7795)=562,081$; $p<0,0001$). Le test PLSD de Fisher pour les différents types de stimuli pris deux à deux indique que *le taux de réponses correctes est significativement plus élevé sur les consonnes voisées que sur les consonnes sourdes* ($p<0,0001$), *plus élevé sur les voyelles isolées que sur les consonnes voisées* ($p<0,02$), *similaire sur les voyelles*

¹⁸⁵ *StatView* version 5.0.

isolées et les combinaisons « consonne + voyelle » ($p > 0,80$) et significativement plus élevé sur les mots entiers que sur les combinaisons C + V ($p < 0,01$). Par ailleurs, le test-T univarié indique que les résultats obtenus par les auditeurs francophones sur les différents types de stimuli sont tous significativement supérieurs au seuil de chance ($p < 0,0001$ dans tous les cas).

Une ANOVA du même type a été conduite sur les degrés de certitude émis par les auditeurs francophones. Sans surprise, l'effet global du facteur « type de stimulus » est ici encore fortement significatif ($F(4,7795)=1890,48$; $p < 0,0001$). Le test PLSD de Fisher révèle des tendances identiques à celles obtenues pour le pourcentage d'identifications réussies. Ainsi, *le degré de certitude moyen des auditeurs francophones est significativement plus haut sur les consonnes voisées que sur les consonnes sourdes ($p < 0,0001$), sur les voyelles isolées que sur les consonnes voisées ($p < 0,0001$), similaire sur les voyelles isolées et les combinaisons « consonne + voyelle » ($p > 0,1$) et significativement plus haut sur les mots entiers par rapport aux combinaisons C + V ($p < 0,0001$).*

Des tests statistiques similaires ont été effectués sur les résultats des auditeurs anglophones, à commencer par une ANOVA à un facteur (« type de stimulus ») portant sur les pourcentages d'identifications correctes. La figure 60, ci-après, illustre cette analyse.

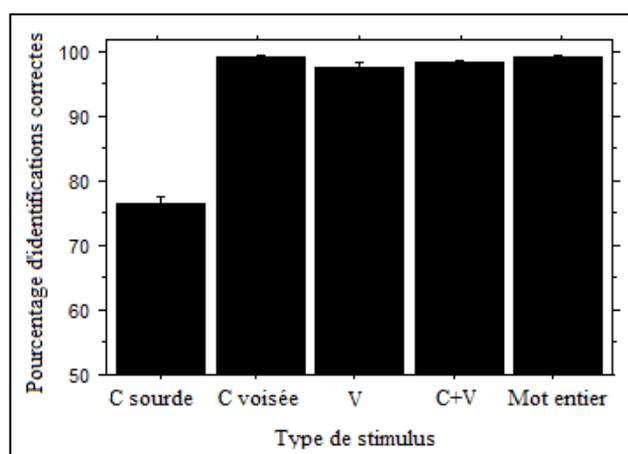


Figure 60 – Pourcentage d'identifications du genre correctes sur les cinq grandes catégories de stimuli pour les auditeurs anglophones, avec les barres d'erreur.

Cette analyse met en évidence un effet global fortement significatif du facteur « type de stimulus » sur le pourcentage d'identifications correctes des auditeurs

anglophones américains ($F(4,7795)=353,35$; $p<0,0001$). Comme chez les auditeurs francophones, le test PLSD de Fisher indique que *le pourcentage d'identifications réussies est significativement plus élevé sur les consonnes voisées que sur les consonnes sourdes* ($p<0,0001$). Cependant, le pourcentage étant déjà extrêmement élevé sur les consonnes voisées, il n'existe ici *aucune différence significative entre les scores obtenus sur les consonnes voisées et les voyelles isolées* ($p>0,20$), *les voyelles isolées et les séquences C + V* ($p>0,5$), *les séquences C + V et les mots entiers* ($p>0,1$). D'autre part, le test-T univarié indique, à l'instar des francophones, que *les résultats obtenus sur les différents types de stimuli sont tous significativement supérieurs au seuil de chance* ($p<0,0001$ dans tous les cas).

Une ANOVA à un facteur (« type de stimulus ») a également été conduite sur les degrés de certitude émis par les auditeurs anglophones. Il existe bien un effet global très significatif de ce facteur ($F(4,7795)=2148,26$; $p<0,0001$). Le PLSD de Fisher révèle que *le degré de certitude moyen des auditeurs anglophones est significativement plus élevé sur les consonnes voisées que sur les consonnes sourdes* ($p<0,0001$), *sur les voyelles isolées que sur les consonnes voisées* ($p<0,01$), *similaire sur les voyelles isolées et les combinaisons « consonne + voyelle »* ($p>0,5$) *et significativement plus élevé sur les mots entiers par rapport aux combinaisons C + V* ($p<0,001$).

Je me suis, d'autre part, intéressé aux différentes voix utilisées comme stimuli. D'après les résultats bruts¹⁸⁶, certaines voix ont semblé poser plus de problèmes que d'autres aux auditeurs. Afin de vérifier si ces tendances sont significatives, j'ai procédé à une ANOVA à un facteur (« voix ») sur le pourcentage d'identifications réussies chez les auditeurs francophones pour l'ensemble de l'expérience. Le graphique résultant de ce test statistique apparaît en figure 61.

¹⁸⁶ Voir section 3.1.1.

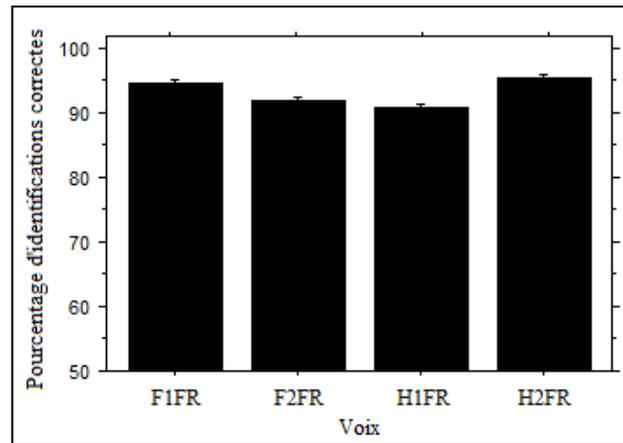


Figure 61 – Pourcentage d'identifications du genre correctes pour les auditeurs francophones en fonction du locuteur produisant les stimuli, avec les barres d'erreur.

Le résultat de l'ANOVA est le suivant : pour l'expérience conduite sur les auditeurs francophones, il existe bien un effet global significatif du facteur « voix » sur le pourcentage d'identifications du genre correctes ($F(3,7796)=15,419$; $p<0,0001$). L'examen des résultats du PLSD de Fisher permet de confirmer les observations faites dans la sous-section précédente. *Les stimuli produits par la voix F2FR ont globalement entraîné un pourcentage de bonnes réponses significativement plus faible que ceux produits par F1FR* ($p<0,001$), de même, *les items produits par le locuteur H1FR obtiennent un score moyen significativement plus bas que ceux de H2FR* ($p<0,0001$). Comme nous l'avons noté précédemment, rappelons que les fréquences de résonance des consonnes initiales de H1FR sont en moyenne nettement plus élevées que celles de H2FR¹⁸⁷, tandis que les valeurs formantiques de F2FR sont globalement plus basses que celles de F1FR¹⁸⁸, en particulier sur la voyelle [u].

Une analyse similaire a été effectuée sur les degrés de certitude émis par les auditeurs francophones. L'effet global du facteur « voix » est là encore fortement significatif ($F(3,7796)=6,77$; $p<0,001$). Le test PLSD de Fisher révèle des tendances identiques à celles obtenues pour le pourcentage d'identifications réussies sur les locuteurs féminins : *le degré de certitude moyen est significativement plus bas sur les stimuli produits par la locutrice F2FR que sur ceux prononcés par F1FR* ($p<0,01$). En

¹⁸⁷ Voir section 3.1.1.1 du chapitre 2.

¹⁸⁸ Voir section 3.1.2.1 du chapitre 2.

revanche, même si le degré de certitude est globalement plus faible pour H1FR que pour H2FR, la différence n'est pas significative ($p > 0,05$).

Des tests statistiques identiques ont été réalisés sur les résultats obtenus par les auditeurs anglophones américains. En premier lieu, une ANOVA à un facteur (« voix ») portant sur les pourcentages d'identifications correctes a été conduite. La figure 62, ci-dessous, illustre cette analyse.

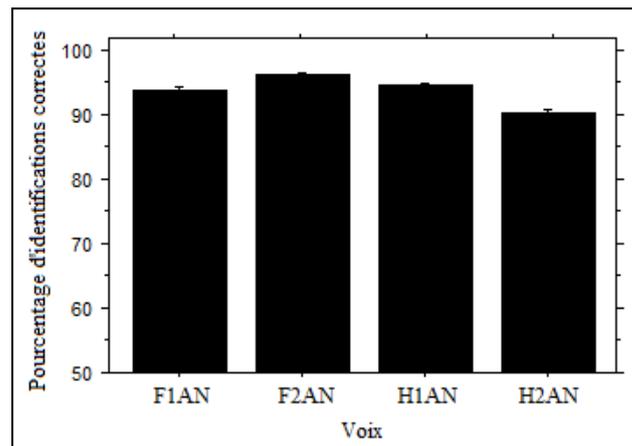


Figure 62 – Pourcentage d'identifications du genre correctes pour les auditeurs anglophones en fonction du locuteur produisant les stimuli, avec les barres d'erreur.

L'ANOVA indique qu'il existe un effet global très significatif du facteur « voix » sur le pourcentage de réponses correctes obtenu par les auditeurs anglophones ($F(3,7796)=20,428$; $p < 0,0001$). A l'instar des francophones, le résultat du test PLSD de Fisher confirme les tendances observées sur les résultats bruts : *les stimuli provenant de la locutrice F1AN ont globalement entraîné un pourcentage d'identifications correctes significativement plus bas que ceux produits par F2AN* ($p < 0,01$) ; *les items produits par le locuteur H2AN obtiennent quant à eux un score moyen significativement plus faible que ceux de H1AN* ($p < 0,0001$). Pour rappel, les consonnes sourdes produites par H1AN, ont leur pic spectral et leur centre de gravité dans des fréquences globalement plus élevées que celles produites par H2AN¹⁸⁹, tandis que la locutrice F1AN présente un F0

¹⁸⁹ Voir section 3.1.1.1 du chapitre 2.

globalement plus bas que F2AN¹⁹⁰ ainsi qu'un pic spectral et un centre de gravité situés en moyenne dans des fréquences plus basses sur les consonnes initiales.

Une analyse similaire a été conduite sur le degré de certitude des auditeurs anglophones. L'effet global du facteur « voix » est à nouveau fortement significatif ($F(3,7796)=25,05$; $p<0,001$). Le test PLSD de Fisher indique que *le degré de certitude moyen est significativement plus bas sur les stimuli produits par la locutrice F1AN que sur ceux prononcés par F2AN* ($p<0,0001$), ce qui va dans le sens des résultats obtenus précédemment sur le pourcentage de bonnes réponses. A l'inverse, *il n'existe pas de différence significative entre les degrés de certitude exprimés sur les stimuli provenant du locuteur H1AN et sur ceux produits par H2AN* ($p>0,9$).

3.2.2 Consonnes initiales isolées

Afin d'analyser plus finement les résultats sur les consonnes initiales présentées isolément, j'ai réalisé une ANOVA à un facteur (« consonne ») sur le pourcentage d'identifications du genre correctes chez les auditeurs francophones. Le graphique correspondant à cette analyse est présenté ci-dessous (figure 63).

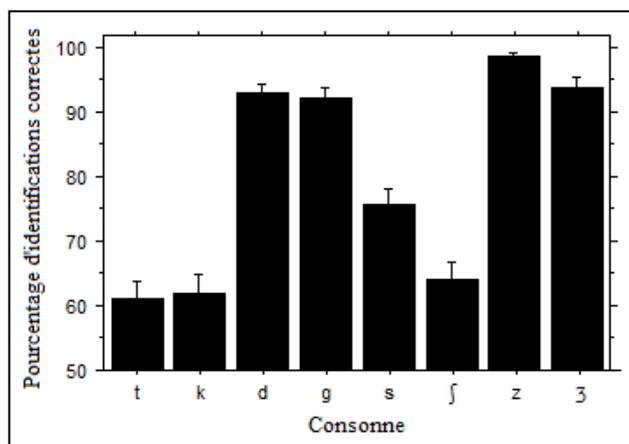


Figure 63 – Pourcentage d'identifications du genre correctes pour les auditeurs francophones en fonction de la consonne présentée isolément, avec les barres d'erreur.

L'ANOVA met en évidence un effet global fortement significatif du facteur « consonne » sur le pourcentage d'identifications correctes obtenu par les auditeurs francophones ($F(7,2392)=76,365$; $p<0,0001$). Un examen des résultats du PLSD de

¹⁹⁰ Voir section 3.1.3.1 du chapitre 2.

Fisher pour les différentes consonnes prises deux à deux indique que *toutes les consonnes voisées obtiennent des scores significativement supérieurs à leurs contreparties sourdes* ($p < 0,0001$ pour [t] / [d], [k] / [g], [s] / [z] et [ʃ] / [ʒ]). *Parmi les consonnes non-voisées, la consonne [s] obtient le meilleur score*, ce dernier étant significativement plus haut que pour [ʃ] ($p < 0,001$), [k] ($p < 0,0001$) et [t] ($p < 0,0001$). *Sur les consonnes voisées, le pourcentage est significativement plus élevé sur la fricative [z] que sur l'occlusive [g]* ($p < 0,05$), mais cette différence n'atteint pas le seuil de significativité lors d'une comparaison avec les scores de [d] ($p > 0,05$) et de [ʒ] ($p > 0,1$). D'autre part, le test-T univarié indique que *les scores obtenus par les auditeurs francophones sur chacune des consonnes sont tous significativement supérieurs au seuil de chance* ($p < 0,0001$ dans tous les cas).

Une ANOVA du même type a été effectuée sur les degrés de certitude émis par les auditeurs francophones. Sans surprise, l'effet global du facteur « consonne » est ici encore fortement significatif ($F(7,2392)=57,04$; $p < 0,0001$). Le test PLSD de Fisher confirme, voire amplifie, les tendances observées sur les pourcentages de bonnes réponses. Ainsi, *chacune des consonnes voisées obtient un degré de certitude significativement supérieur à sa contrepartie sourde* ($p < 0,0001$ pour [t] / [d], [k] / [g], [s] / [z] et [ʃ] / [ʒ]). *La consonne [s] obtient le degré de certitude le plus élevé de toutes les consonnes sourdes*, il est significativement plus élevé que pour [ʃ] ($p < 0,01$), [k] ($p < 0,0001$) et [t] ($p < 0,0001$). *Parmi les consonnes voisées, le degré de certitude est significativement plus élevé sur la fricative [z] que sur les trois autres consonnes du même type*, à savoir [ʒ] ($p < 0,02$), [g] ($p < 0,0001$) et [d] ($p < 0,0001$).

On constate donc que les scores les plus élevés pour les consonnes voisées et sourdes sont atteints sur des fricatives. L'analyse acoustique du chapitre précédent montre que la durée de ces dernières est sensiblement supérieure à celle des occlusives¹⁹¹. J'ai par conséquent effectué un test de Pearson pour vérifier s'il existe une corrélation entre la durée du stimulus (pour les consonnes initiales isolées) et le pourcentage moyen de bonnes réponses correspondant. Le résultat de cette analyse montre qu'il n'existe pas de corrélation significative entre ces deux variables : $r(96)=-0,056$; $z=-0,543$ avec $p > 0,5$. Cependant, le même test conduit sur le degré de certitude moyen révèle une corrélation

¹⁹¹ Voir section 3.1.1.4 du chapitre 2.

positive significative avec la durée des stimuli : $r(96)=0,236$; $z=2,318$ avec $p<0,05$. Cela signifie que plus la consonne présentée était longue, plus les auditeurs francophones ont eu tendance à être sûrs de leurs jugements.

J'ai par ailleurs souhaité tester les éventuelles corrélations entre le F0 moyen des consonnes voisées et les scores obtenus, en séparant les items de type « voix de femme » et « voix d'homme ». Pour les voix de femmes, ces tests de Pearson font état de corrélations positives très faibles et non-significatives entre le F0 de la consonne et le pourcentage d'identifications correctes ($r(24)=0,141$; $z=0,652$ et $p>0,5$) ainsi qu'entre le F0 et le degré de certitude moyen ($r(24)=0,147$; $z=0,679$ et $p>0,4$)¹⁹². Sur les voix d'hommes, des corrélations négatives non-significatives sont observées entre le F0 et le pourcentage de bonnes réponses ($r(24)=-0,183$; $z=-0,85$ et $p>0,3$) et entre le F0 et le degré de certitude moyen ($r(24)=-0,061$; $z=-0,281$ et $p>0,7$)¹⁹³. L'absence de significativité s'explique en partie par le fait que les valeurs de F0, pour les hommes comme pour les femmes, sont ici toutes très caractéristiques de leur genre (F0 élevé pour les femmes, F0 bas pour les hommes¹⁹⁴).

D'autre part, notons que les consonnes [s] et [z], qui obtiennent respectivement le meilleur score des sons non-voisés et voisés, présentent de très fortes différences inter-genres sur le plan du centre de gravité¹⁹⁵. Il n'est cependant pas possible de chercher ici d'éventuelles corrélations entre le centre de gravité et les scores obtenus, car le centre de gravité est entièrement dépendant du type de consonne. Il conviendrait donc de mener des tests distincts pour chaque consonne, en plus de devoir séparer les voix d'hommes et de femmes : les effectifs pour chaque test seraient par conséquent trop faibles pour obtenir des résultats intéressants et fiables.

¹⁹² Les mêmes tests ont été conduits sur les seules fricatives voisées, afin de permettre une comparaison plus pertinente avec les auditeurs anglophones (dont le corpus de stimuli ne comporte pas d'occlusives voisées). Les résultats obtenus sont très similaires : $r(12)=0,04$; $z=0,119$ et $p>0,9$ pour le pourcentage de bonnes réponses, $r(12)=-0,107$; $z=-0,324$ et $p>0,7$ pour le degré de certitude moyen.

¹⁹³ Là encore, les résultats sont très proches si l'on ne prend en compte que les fricatives voisées, afin de permettre une comparaison avec les auditeurs anglophones : $r(12)=-0,377$; $z=-1,190$ et $p>0,2$ pour le pourcentage d'identifications correctes, $r(12)=-0,013$; $z=-0,039$ et $p>0,9$ pour le degré de certitude moyen.

¹⁹⁴ Voir section 3.1.1.3 du chapitre 2.

¹⁹⁵ Voir section 3.1.1.1 du chapitre 2.

Des analyses statistiques similaires ont été réalisées sur les résultats des auditeurs anglophones américains, à commencer par une ANOVA à un facteur (« consonne ») portant sur les pourcentages d'identifications correctes. La figure 64, ci-dessous, illustre cette analyse.

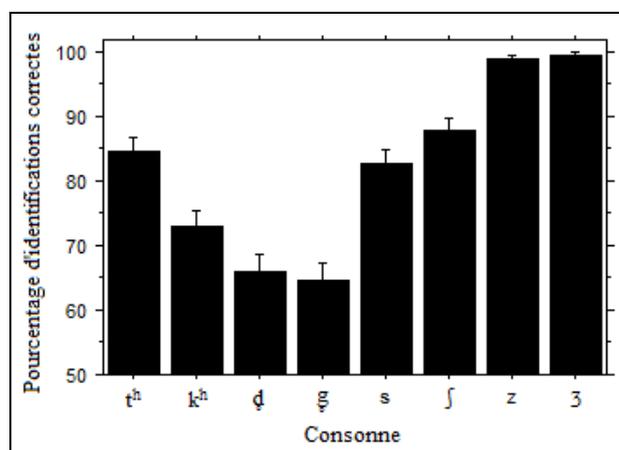


Figure 64 – Pourcentage d'identifications du genre correctes pour les auditeurs anglophones en fonction de la consonne présentée isolément, avec les barres d'erreur.

Le résultat de l'ANOVA est le suivant : pour l'expérience conduite sur les auditeurs anglophones, il existe également un effet global significatif du facteur « consonne » sur le pourcentage d'identifications du genre réussies ($F(7,2392)=42,04$; $p<0,0001$). L'examen des résultats du PLSD de Fisher permet de confirmer les observations faites à partir des résultats bruts. Comme chez les auditeurs francophones, *les consonnes voisées obtiennent des scores significativement plus élevés que leurs contreparties sourdes* ($p<0,0001$ pour [s] / [z] et pour [j] / [ʒ]). De même, *les scores des occlusives aspirées sont significativement supérieurs à ceux de leurs contreparties non-aspirées* ($p<0,0001$ pour [t^h] / [d] ; $p<0,01$ pour [k^h] / [g]). *C'est la fricative [j] qui obtient les meilleurs scores des consonnes sourdes, avec une différence significative par rapport aux consonnes [k^h], [d] et [g] ($p<0,0001$), mais la différence n'atteint pas le seuil de significativité avec [s] et [t^h] ($p>0,05$; $p>0,2$).* *Le pourcentage d'identifications correctes le plus élevé sur les items voisés est atteint sur [ʒ] mais la différence avec la fricative alvéolaire [z] n'est pas significative ($p>0,8$).* Par ailleurs, le test-T univarié indique que *les scores obtenus par les auditeurs anglophones sur chacune des consonnes sont significativement supérieurs au seuil de chance* ($p<0,0001$ dans tous les cas).

La même ANOVA à un facteur a été conduite sur le degré de certitude des auditeurs anglophones américains. Cette analyse révèle une nouvelle fois un effet global très significatif du facteur « consonne » ($F(7,2392)=244,435$; $p<0,0001$). Les résultats du PLSD de Fisher pour les différentes consonnes prises deux à deux vont dans le sens de ceux obtenus avec les pourcentages d'identifications correctes. *Les consonnes voisées et aspirées obtiennent un degré de certitude moyen significativement plus élevé que leurs contreparties sourdes ou non-aspirées* ($p<0,0001$ pour [s] / [z], [ʃ] / [ʒ], [t^h] / [d] et [k^h] / [g]). *Parmi les consonnes sourdes, les fricatives [s] et [ʃ] recueillent un degré de certitude significativement plus élevé que les autres consonnes du même type* ($p<0,05$ avec [t^h] ; $p<0,0001$ dans les autres cas) et la différence entre ces deux items est non-significative ($p>0,5$). *La consonne voisée post-alvéolaire [ʒ] recueille un degré de certitude moyen sensiblement plus élevé que [z],* mais cette différence demeure juste au-dessus du seuil de significativité ($p>0,05$).

A l'instar des auditeurs francophones, les meilleurs scores pour les consonnes sourdes sont atteints sur les fricatives. Ici encore, on constate que leur durée moyenne est sensiblement supérieure à celle des occlusives¹⁹⁶. J'ai donc conduit un test de Pearson pour vérifier s'il existe une corrélation entre la durée des consonnes et le pourcentage de réponses correctes moyen : il existe bien une corrélation positive très significative entre ces deux variables ($r(96)=0,4$; $z=4,091$ et $p<0,0001$). Cette corrélation est également significative avec les degrés de certitude moyens ($r(96)=0,435$; $z=4,496$ et $p<0,0001$). En d'autres termes, *plus la consonne présentée était longue, plus les auditeurs anglophones américains ont eu tendance à correctement identifier le genre du locuteur et à être sûrs de leurs choix.*

J'ai d'autre part souhaité chercher d'éventuelles corrélations entre le F0 moyen des consonnes voisées et les scores obtenus, en séparant les items de type « voix de femme » et « voix d'homme ». *Sur les voix de femmes, ces tests de Pearson mettent en évidence une forte corrélation positive mais non significative entre le F0 de la consonne et le pourcentage d'identifications correctes* ($r(12)=0,537$; $z=1,801$ et $p>0,05$) *ainsi qu'une forte corrélation positive et significative entre le F0 et le degré de certitude moyen* ($r(12)=0,793$; $z=3,237$ et $p<0,01$). Pour les voix d'hommes, les consonnes voisées ayant toutes obtenus 100 % de bonnes réponses, le test a uniquement été effectué avec les

¹⁹⁶ Voir section 3.1.1.4 du chapitre 2.

degrés de certitude moyens. Ce test de Pearson conduit sur les items « voix d'hommes » révèle une corrélation négative forte et significative entre le F0 moyen et le degré de certitude moyen ($r(12)=-0,625$; $z=-2,197$ et $p<0,05$). Cela signifie que plus le F0 d'une consonne produite par une femme était élevé, plus les auditeurs anglophones ont eu tendance à correctement identifier le genre du locuteur et à être sûrs de leurs choix, et plus le F0 d'une consonne produite par un homme était bas, plus les auditeurs ont eu tendance à être sûrs de leurs choix. Il est important de noter que cette tendance n'a pas été observée chez les auditeurs francophones.

On remarque également que les réalisations des fricatives [ʃ] et [ʒ], qui obtiennent des scores très élevés, présentent une très forte différence femmes-hommes au plan du pic spectral et du centre de gravité¹⁹⁷. Comme chez les auditeurs francophones, il n'est pour autant par possible de tester ici les corrélations entre ces paramètres acoustiques et les scores obtenus, en raison de la nécessaire séparation des résultats par consonne et par genre et donc de la faiblesse des effectifs.

3.2.3 Voyelles initiales isolées

J'ai conduit une ANOVA à un facteur (« voyelle ») sur les pourcentages d'identifications correctes obtenus par les auditeurs francophones sur les items de type « voyelle initiale isolée ». Le graphique résultant de ce test statistique apparaît ci-dessous en figure 65.

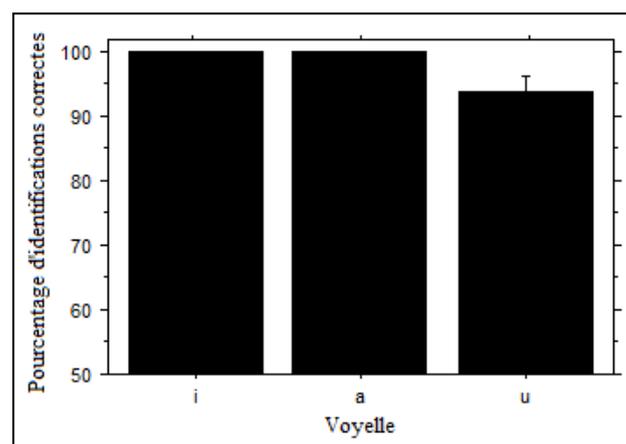


Figure 65 – Pourcentage d'identifications du genre correctes pour les auditeurs francophones en fonction de la voyelle présentée isolément, avec les barres d'erreur.

¹⁹⁷ Voir section 3.1.1.1 du chapitre 2.

Malgré des scores très élevés sur les trois voyelles, l'ANOVA met en évidence un effet significatif du facteur « voyelle » sur le pourcentage de réponses correctes chez les auditeurs francophones ($F(2,297)=6,319$; $p<0,01$). Le test PLSD de Fisher révèle que cet effet global est dû à la voyelle [u], dont le score est significativement inférieur à ceux des voyelles [i] et [a] ($p<0,01$ dans les deux cas), ces deux dernières obtenant chacune 100 % de bonnes réponses. J'ai également procédé à une ANOVA à un facteur (« voyelle ») sur les degrés de certitude émis par les auditeurs francophones : il existe bien un effet global significatif de ce facteur ($F(2,297)=5,37$; $p<0,01$). Le résultat du PLSD de Fisher va dans le sens des observations faites précédemment : le degré de certitude des auditeurs est en effet significativement plus bas sur la voyelle [u] que sur [i] ($p<0,01$) et sur [a] ($p<0,05$).

Rappelons que chez les francophones, la voyelle [u] est celle qui présente globalement le moins de différences formantiques hommes-femmes, en particulier sur le deuxième formant¹⁹⁸. Il n'est pas possible de tester les éventuelles corrélations entre les scores obtenus (pourcentage moyen de bonnes réponses et degré de certitude moyen) et la valeur des différents formants pour les seules voyelles initiales isolées en raison de la faiblesse des effectifs ; ces analyses (tests de Pearson) seront donc effectuées dans la sous-section suivante, consacrée aux items « consonne + voyelle ». Il en va de même pour les corrélations entre les scores, la durée et le F0 moyen.

Des tests statistiques identiques ont été effectués sur les résultats obtenus par les auditeurs anglophones américains. En premier lieu, une ANOVA à un facteur (« voyelle ») portant sur les pourcentages d'identifications correctes a été réalisée. La figure 66, ci-après, illustre cette analyse.

¹⁹⁸ Voir section 3.1.2.1 du chapitre 2.

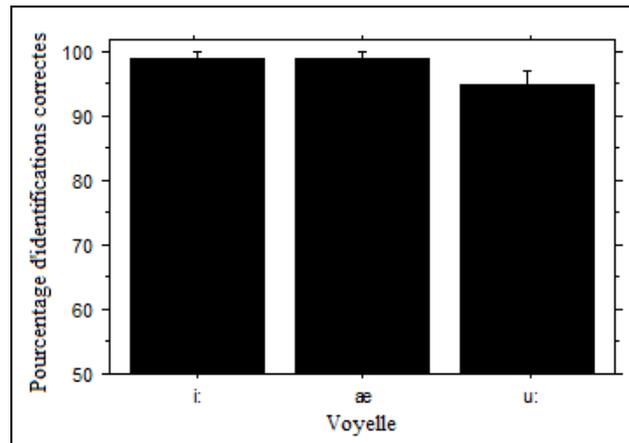


Figure 66 – Pourcentage d'identifications du genre correctes pour les auditeurs anglophones en fonction de la voyelle présentée isolément, avec les barres d'erreur.

Contrairement aux auditeurs francophones, l'ANOVA révèle ici l'absence d'effet global significatif du facteur « voyelle » sur le pourcentage d'identifications du genre réussies par les auditeurs anglophones américains ($F(2,297)=2,354$; $p>0,05$). Par conséquent, le PLSD de Fisher ne décèle aucune différence significative pour les différences voyelles prises deux à deux. Une même ANOVA à un facteur a été conduite sur le degré de certitude des auditeurs anglophones : l'analyse indique qu'il n'existe pas non plus d'effet global significatif du facteur « voyelle » sur le degré de certitude exprimé par les auditeurs ($F(2,297)=2,347$; $p>0,05$). Sans surprise, le test PLSD de Fisher ne fait donc état d'aucune différence significative entre les différentes paires de voyelles.

Les résultats obtenus par les auditeurs anglophones sont donc équivalents sur les trois types de voyelles présentées isolément. Cela pourrait s'expliquer par les différences formantiques inter-genres assez fortes qu'il existe sur les trois voyelles¹⁹⁹, ou par les différences inter-genres sur le plan du F_0 moyen, qui pourraient à elles seules suffire aux auditeurs anglophones pour identifier avec certitude le genre du locuteur. Comme pour les auditeurs francophones, le faible nombre de stimuli de type « voyelle initiale isolée » ne permet pas ici de procéder à des tests de Pearson : ils seront menés dans la sous-section suivante, consacrée aux combinaisons $C + V$.

¹⁹⁹ Voir section 3.1.2.1 du chapitre 2.

3.2.4 Combinaisons « consonne + voyelle »

Afin d'analyser plus en détail les résultats des auditeurs francophones sur les combinaisons de type $C + V$, j'ai effectué une ANOVA à un facteur (« type de stimulus ») sur le pourcentage d'identifications du genre correctes. Le graphique correspondant à cette analyse est présenté ci-dessous (figure 67).

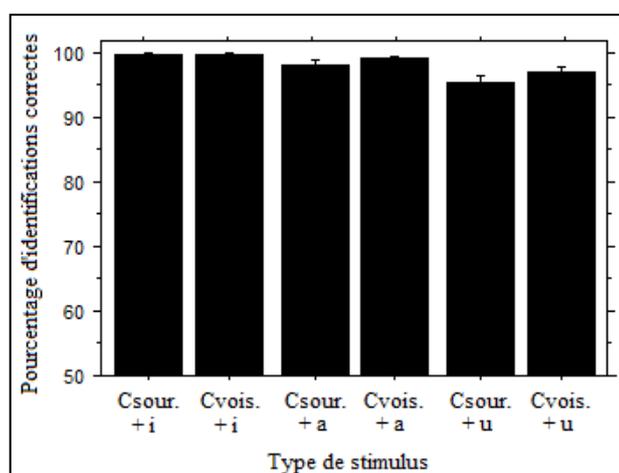


Figure 67 – Pourcentage d'identifications du genre correctes obtenu par les auditeurs francophones en fonction du type de combinaison « $C + V$ », avec les barres d'erreur.

Le résultat de l'ANOVA fait état d'un effet global fortement significatif du facteur « type de stimulus » sur le pourcentage d'identifications réussies chez les auditeurs francophones ($F(5,2394)=6,885$; $p<0,0001$). Le test PLSD de Fisher indique l'absence de différences significatives entre les combinaisons « consonne sourde + [i] » et « consonne voisée + [i] » (scores identiques), « consonne sourde + [a] » et « consonne voisée + [a] » ($p>0,2$), « consonne sourde + [u] » et « consonne voisée + [u] » ($p>0,1$). *La présence d'une consonne voisée plutôt que d'une consonne sourde en initiale n'a donc pas entraîné de pourcentages de réponses correctes significativement plus élevés chez les francophones.* En revanche, il apparaît que *les combinaisons $C + [u]$ ont obtenu des scores significativement plus faibles que les combinaisons du même type contenant une autre voyelle, à savoir [i] ($p<0,0001$ pour les combinaisons avec consonne sourde, $p<0,01$ pour celles avec consonne voisée) et [a] ($p<0,01$ pour les combinaisons avec consonne sourde, $p<0,02$ pour celles avec consonne voisée).*

Une ANOVA du même type a été conduite sur les degrés de certitude émis par les auditeurs francophones. De nouveau, l'effet global du facteur « type de stimulus » est

fortement significatif ($F(5,2394)=5,68$; $p<0,0001$). Selon les résultats du PLSD de Fisher, *la présence d'une consonne voisée plutôt qu'une consonne sourde avant la voyelle [u] fait monter de manière significative le degré de certitude des auditeurs francophones* ($p<0,02$), cependant *cela n'est pas le cas pour les combinaisons contenant la voyelle [i]* ($p>0,8$) *ou la voyelle [a]* ($p>0,5$). A l'instar des observations faites pour les pourcentages d'identifications correctes, on note que *les combinaisons de type C + [u] ont obtenu des degrés de certitude moyens significativement plus faibles que des combinaisons du même type contenant une autre voyelle* : c'est le cas avec [i], tant avec une consonnes sourde en initiale ($p<0,0001$) qu'une consonne voisée ($p<0,05$), et avec la voyelle [a], uniquement avec une consonne sourde en initiale ($p<0,01$ mais $p>0,2$ pour les combinaisons avec consonne voisée).

On remarque que *les types de combinaison qui obtiennent le plus de d'identifications erronées et le degré de certitude le plus bas contiennent la voyelle [u], qui présente les plus faibles différences formantiques inter-genres*²⁰⁰. Un grand nombre de tests de Pearson ont été menés sur les degrés de certitude moyens²⁰¹ obtenus pour les stimuli de cette catégorie (combinaisons C + V), qui ont été mis en parallèle avec les valeurs formantiques des voyelles, la différence H1-H2, la durée des stimuli et le F0 moyen. Pour plus de clarté, les résultats de ces tests sont présentés ci-après sous forme de tableau (tableau 43).

²⁰⁰ Voir section 3.1.2.1 du chapitre 2.

²⁰¹ Les pourcentages de bonnes réponses étant trop homogènes pour cette catégorie de stimuli (100 % pour presque tous les items) n'ont pas été pris en compte pour les tests de Pearson.

Tableau 43 – Résultats des tests de Pearson (corrélation) conduits sur les données relatives aux degrés de certitude moyens des auditeurs francophones pour les stimuli de type C + V. Le symbole « * » indique une corrélation significative entre les deux variables (avec un seuil $p < 0,05$).

Résultats des tests de Pearson pour les auditeurs francophones sur les séquences C + V		Variable Y : Degré de certitude moyen
Variable X	F1 (Hz) voyelle [i] voix F	$r(16)=0,062$; $z=-0,222$; $p>0,8$
	F2 (Hz) voyelle [i] voix F	$r(16)=0,262$; $z=0,966$; $p>0,3$
	F3 (Hz) voyelle [i] voix F	$r(16)=0,076$; $z=0,274$; $p>0,7$
	F1 (Hz) voyelle [i] voix H	$r(16)=0,167$; $z=0,608$; $p>0,5$
	F2 (Hz) voyelle [i] voix H	$r(16)=-0,569$; $z=-2,331$; $p<0,02^*$
	F3 (Hz) voyelle [i] voix H	$r(16)=-0,335$; $z=-1,258$; $p>0,2$
	F1 (Hz) voyelle [a] voix F	$r(16)=-0,271$; $z=-1,003$; $p>0,3$
	F2 (Hz) voyelle [a] voix F	$r(16)=-0,121$; $z=-0,439$; $p>0,6$
	F3 (Hz) voyelle [a] voix F	$r(16)=-0,228$; $z=-0,878$; $p>0,3$
	F1 (Hz) voyelle [a] voix H	$r(16)=0,195$; $z=0,714$; $p>0,4$
	F2 (Hz) voyelle [a] voix H	$r(16)=-0,491$; $z=-1,899$; $p<0,05^*$
	F3 (Hz) voyelle [a] voix H	$r(16)=-0,371$; $z=-1,407$; $p>0,1$
	F1 (Hz) voyelle [u] voix F	$r(16)=-0,634$; $z=2,695$; $p<0,01^*$
	F2 (Hz) voyelle [u] voix F	$r(16)=0,728$; $z=3,331$; $p<0,001^*$
	F3 (Hz) voyelle [u] voix F	$r(16)=0,759$; $z=3,582$; $p<0,001^*$
	F1 (Hz) voyelle [u] voix H	$r(16)=-0,426$; $z=-1,642$; $p>0,1$
	F2 (Hz) voyelle [u] voix H	$r(16)=-0,497$; $z=-1,927$; $p<0,05^*$
	F3 (Hz) voyelle [u] voix H	$r(16)=-0,403$; $z=-1,541$; $p>0,1$
	H1-H2 (dB) voyelle [a] voix F	$r(16)=-0,28$; $z=-1,036$; $p>0,3$
	H1-H2 (dB) voyelle [a] voix H	$r(16)=0,073$; $z=0,263$; $p>0,7$
F0 moyen (Hz) tous stimuli voix F	$r(48)=-0,187$; $z=-1,269$; $p>0,2$	
F0 moyen (Hz) tous stimuli voix H	$r(48)=-0,426$; $z=-3,053$; $p<0,01^*$	
Durée (ms) tous stimuli	$r(96)=-0,04$; $z=-0,387$; $p>0,6$	

Ces résultats révèlent plusieurs éléments intéressants. Il apparaît tout d'abord que *la fréquence des formants vocaliques a très probablement joué un rôle important chez les auditeurs francophones pour cette tâche d'identification du genre par la voix*. En effet, plusieurs corrélations significatives apparaissent entre la fréquence des formants et le degré de certitude des auditeurs. Ces corrélations sont positives pour les voix de femmes (plus la fréquence du formant est élevée, plus le degré de certitude tend à être haut) et négatives pour les voix d'hommes (plus la fréquence est basse, plus le degré de certitude tend à diminuer). Elles portent plus souvent sur F2 (4 corrélations significatives) que sur F1 et F3 (1 corrélation significative chacun) : *la fréquence du deuxième formant pourrait donc jouer un rôle plus fort que celle de F1 et F3 chez les auditeurs francophones*. D'autre part, ces corrélations significatives sont plus fréquemment obtenues pour les voix

d'hommes que pour les voix de femmes, exception faite de la voyelle [u] : cela peut s'expliquer par la faiblesse des différences intra-locuteur et intra-genre chez les femmes francophones sur les valeurs formantiques de [i] et de [a]²⁰².

Concernant les autres paramètres testés, on constate que *la durée des stimuli n'a apparemment pas influencé les jugements des auditeurs sur ces séquences C + V. De même, aucune tendance claire ne se dégage pour la différence H1-H2, qui fournit une indication du type de phonation. Ce paramètre ne semble pas avoir joué un rôle chez les auditeurs francophones. En revanche, la fréquence fondamentale moyenne a probablement bien influencé le jugement des auditeurs francophones, tout du moins pour les voix d'hommes : plus FO était bas, plus les auditeurs ont eu tendance à être sûrs de leur choix.*

Des analyses statistiques similaires ont été réalisées sur les résultats des auditeurs anglophones américains, en commençant par une ANOVA à un facteur (« type de stimulus ») portant sur les pourcentages d'identifications du genre correctes. La figure 68, ci-dessous, illustre cette analyse.

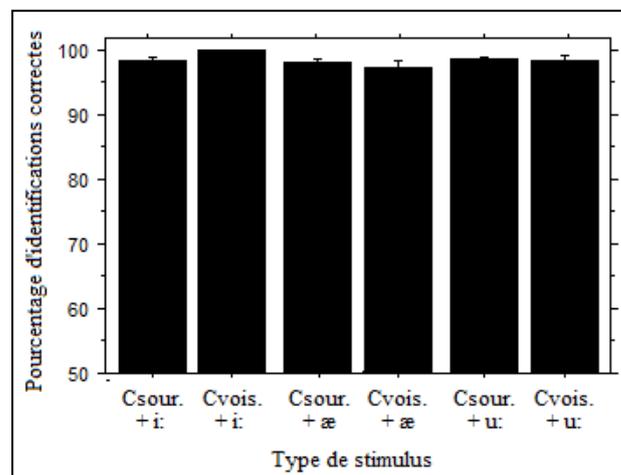


Figure 68 – Pourcentage d'identifications du genre correctes obtenu par les auditeurs anglophones en fonction du type de combinaison « C + V », avec les barres d'erreur.

Le résultat de l'analyse indique qu'à l'inverse des francophones, *il n'existe pas ici d'effet global significatif du facteur « type de stimulus » sur le pourcentage de bonnes réponses obtenu par les auditeurs anglophones pour les items de type C + V*

²⁰² Voir section 3.1.2.1 du chapitre 2.

($F(5,2394)=0,992$; $p>0,4$). Le PLSD de Fisher ne décèle par conséquent aucune différence significative entre les catégories prises deux à deux. Une même ANOVA à un facteur a été effectuée *sur les degrés de certitude exprimés par les anglophones*. Il en ressort une absence d'effet global significatif du facteur « type de stimulus » ($F(5,2394)=1,23$; $p>0,2$). Fort logiquement, le PLSD ne fait état d'aucun écart significatif entre les différents types de stimuli pris deux à deux.

Ces résultats semblent aller dans le même sens que ceux observés sur les voyelles isolées : *contrairement aux francophones, les scores des anglophones sont de nouveau équivalents d'un type de stimulus à l'autre*. J'ai procédé, sur les données des auditeurs anglophones, aux mêmes tests de Pearson que pour les francophones. Les résultats de ces tests apparaissent dans le tableau 44 ci-dessous.

Tableau 44 – Résultats des tests de Pearson (corrélation) conduits sur les données relatives aux degrés de certitude moyens des auditeurs anglophones pour les stimuli de type C + V. Le symbole « * » indique une corrélation significative entre les deux variables (avec un seuil $p<0,05$).

Résultats des tests de Pearson pour les auditeurs anglophones sur les séquences C + V		Variable Y : Degré de certitude moyen
Variable X	F1 (Hz) voyelle [i:] voix F	$r(16)=0,372$; $z=1,41$; $p>0,1$
	F2 (Hz) voyelle [i:] voix F	$r(16)=-0,164$; $z=-0,596$; $p>0,5$
	F3 (Hz) voyelle [i:] voix F	$r(16)=-0,328$; $z=-1,228$; $p>0,2$
	F1 (Hz) voyelle [i:] voix H	$r(16)=-0,82$; $z=-0,296$; $p>0,7$
	F2 (Hz) voyelle [i:] voix H	$r(16)=-0,138$; $z=-0,501$; $p>0,6$
	F3 (Hz) voyelle [i:] voix H	$r(16)=-0,138$; $z=-0,501$; $p>0,6$
	F1 (Hz) voyelle [æ] voix F	$r(16)=0,266$; $z=0,984$; $p>0,3$
	F2 (Hz) voyelle [æ] voix F	$r(16)=-0,221$; $z=-0,811$; $p>0,4$
	F3 (Hz) voyelle [æ] voix F	$r(16)=0,261$; $z=0,962$; $p>0,3$
	F1 (Hz) voyelle [æ] voix H	$r(16)=-0,369$; $z=-1,398$; $p>0,1$
	F2 (Hz) voyelle [æ] voix H	$r(16)=-0,058$; $z=-0,208$; $p>0,8$
	F3 (Hz) voyelle [æ] voix H	$r(16)=-0,094$; $z=-0,339$; $p>0,7$
	F1 (Hz) voyelle [u:] voix F	$r(16)=0,220$; $z=0,807$; $p>0,4$
	F2 (Hz) voyelle [u:] voix F	$r(16)=0,649$; $z=2,788$; $p<0,01^*$
	F3 (Hz) voyelle [u:] voix F	$r(16)=0,454$; $z=1,767$; $p>0,05$
	F1 (Hz) voyelle [u:] voix H	$r(16)=-0,397$; $z=-1,514$; $p>0,1$
	F2 (Hz) voyelle [u:] voix H	$r(16)=-0,019$; $z=-0,068$; $p>0,9$
	F3 (Hz) voyelle [u:] voix H	$r(16)=-0,111$; $z=-0,402$; $p>0,6$
	H1-H2 (dB) voyelle [æ] voix F	$r(16)=0,264$; $z=0,974$; $p>0,3$
	H1-H2 (dB) voyelle [æ] voix H	$r(16)=-0,497$; $z=-1,967$; $p<0,05^*$
F0 moyen (Hz) tous stimuli voix F	$r(48)=0,005$; $z=0,036$; $p>0,9$	
F0 moyen (Hz) tous stimuli voix H	$r(48)=-0,341$; $z=-2,38$; $p<0,02^*$	
Durée (ms) tous stimuli	$r(96)=-0,174$; $z=-1,696$; $p>0,05$	

Contrairement aux auditeurs francophones, les tests de Pearson sont peu concluants sur les valeurs des formants vocaliques : seule une corrélation (positive) significative a été obtenue sur le F2 de la voyelle [u:], pour les voix de femmes. Cela suggère que *la valeur des formants vocaliques pourrait jouer un moindre rôle chez les auditeurs anglophones*. Concernant *la durée des stimuli*, il apparaît que *cette dernière ne semble pas jouer de rôle significatif sur les séquences de type C + V*.

En revanche, deux autres paramètres acoustiques obtiennent des résultats concluants. A l'inverse des francophones, la différence H1-H2 est corrélée significativement au degré de certitude des auditeurs pour les voix d'hommes : plus elle est basse (i.e. plus le QOG²⁰³ est faible), plus les auditeurs tendent à être sûrs de leurs choix. Une tendance opposée se dessine pour les voix de femmes, mais la corrélation n'atteint pas le seuil de significativité. Ces éléments suggèrent que *le type de phonation joue effectivement un rôle dans l'identification du genre par la voix chez les auditeurs anglophones américains*. De même, *la fréquence fondamentale moyenne semble également avoir influencé de manière significative le degré de certitude des auditeurs*. En effet, une corrélation négative significative apparaît pour les voix d'hommes. La corrélation n'est en revanche pas significative avec le F0 moyen des voix de femmes, très probablement à cause de la faible variation intra-locuteur et intra-genre sur ce paramètre²⁰⁴.

3.2.5 Mots dissyllabiques

J'ai conduit une ANOVA à un facteur (« type de mot ») sur les pourcentages d'identifications correctes obtenus par les auditeurs francophones sur les items de type « mots dissyllabiques ». Le résultat de l'ANOVA est particulièrement clair : *il n'existe aucun effet significatif de ce facteur sur le pourcentage de bonnes réponses chez les francophones* ($F(5,2694)=0,27$; $p>0,9$). Logiquement, le test PLSD de Fisher de décèle aucune différence significative entre les types de mot pris deux à deux. J'ai également procédé à une ANOVA du même type sur les degrés de certitude émis par les auditeurs francophones. Le graphique résultant de cette analyse statistique apparaît ci-après en figure 69.

²⁰³ « Quotient d'ouverture glottique » : voir section 2.4 du chapitre 2.

²⁰⁴ Voir section 3.1.2.3 du chapitre 2.

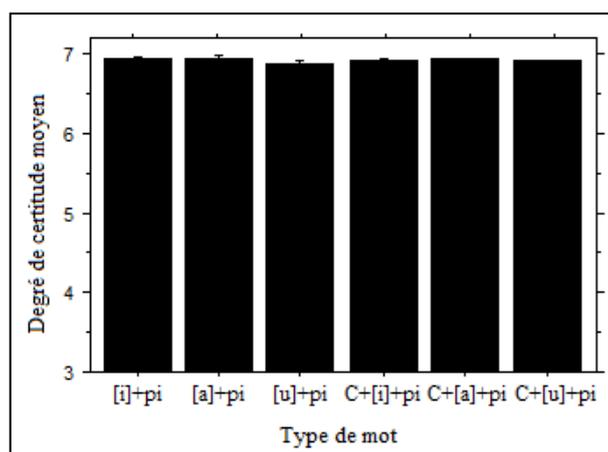


Figure 69 – Degré de certitude moyen (échelle de 0 à 7) exprimé par les auditeurs francophones en fonction du type de mot dissyllabique, avec les barres d'erreur.

Tout comme pour le pourcentage de réponses correctes, *il n'existe pas d'effet significatif du facteur « type de mot » sur le degré de certitude des auditeurs francophones* ($F(5,2694)=0,625$; $p>0,6$). Le PLSD révèle de nouveau l'absence de différence significative entre les paires de types de mot.

Les résultats des auditeurs francophones sur les mots dissyllabiques sont donc équivalents d'un type de mot à l'autre. J'ai tout de même souhaité mettre en parallèle les degrés de certitude moyens avec les données acoustiques à travers des tests de Pearson. Je n'ai pas jugé pertinent de mener ici des corrélations avec les valeurs formantiques de la voyelle en syllabe 1, en raison de la présence d'une deuxième voyelle : ces tests concernent donc uniquement la différence $H1-H2$ ²⁰⁵, le $F0$ moyen, la durée des stimuli (testée ici à la fois de manière globale et en séparant les voix d'hommes et de femmes) ainsi que la plage de variation de $F0$ ²⁰⁶. Les résultats de ces différents tests sont présentés dans le tableau 45 ci-après.

²⁰⁵ Bien que réalisée sur la voyelle en syllabe 1, cette mesure peut donner une indication de la qualité de voix sur l'ensemble du mot dissyllabique.

²⁰⁶ Exprimée en demi-ton et non en hertz, afin de rendre compte de variation effectivement perçue par l'auditeur.

Tableau 45 – Résultats des tests de Pearson conduits sur les données relatives aux degrés de certitude moyens des auditeurs francophones pour les stimuli de type « mots entiers ». Le symbole « * » indique une corrélation significative entre les deux variables (seuil $p < 0,05$).

Résultats des tests de Pearson pour les auditeurs francophones sur les mots dissyllabiques		Variable Y : Degré de certitude moyen
Variable X	H1-H2 (dB) voyelle [a] voix F	$r(18)=-0,148 ; z=-0,577 ; p>0,5$
	H1-H2 (dB) voyelle [a] voix H	$r(18)=-0,061 ; z=-0,235 ; p>0,8$
	F0 moyen (Hz) tous stimuli voix F	$r(54)=0,273 ; z=1,997 ; p<0,05^*$
	F0 moyen (Hz) tous stimuli voix H	$r(54)=0,007 ; z=0,052 ; p>0,9$
	Durée (ms) tous stimuli	$r(108)=-0,083 ; z=-0,852 ; p>0,3$
	Durée (ms) tous stimuli voix F	$r(54)=0,478 ; z=3,712 ; p<0,001^*$
	Durée (ms) tous stimuli voix H	$r(54)=-0,087 ; z=-0,624 ; p>0,5$
	Plage de variation (dt) tous stimuli F	$r(54)=0,212 ; z=1,536 ; p>0,1$
	Plage de variation (dt) tous stimuli H	$r(54)=-0,127 ; z=-0,915 ; p>0,3$

Les résultats de ces tests vont dans le sens de ceux obtenus pour les séquences C + V. On constate tout d'abord que *la qualité de voix, dont la mesure H1-H2 donne une indication, ne semble pas avoir joué un rôle chez les francophones sur les items de type « mots dissyllabiques »*. Il en va de même pour la durée des stimuli toutes voix confondues, pour laquelle on note une absence totale de corrélation avec le degré de certitude moyen exprimé par les auditeurs. On constate cependant *une corrélation positive significative entre la durée des mots et le degré de certitude associé aux seules voix de femmes : une vitesse d'élocution plus lente pourrait donc rendre les auditeurs plus sûrs de leur jugement pour ce type de voix*. Une faible corrélation négative apparaît entre la durée de l'énoncé et le degré de certitude exprimé pour les voix d'hommes, mais cette dernière n'est pas significative.

La fréquence fondamentale moyenne semble avoir sensiblement influencé le jugement des auditeurs. On note en effet une corrélation positive et significative avec le degré de certitude exprimé sur les voix de femmes : plus le F0 moyen était élevé, plus les auditeurs ont eu tendance à être sûrs de leurs choix. Cependant, cette corrélation est relativement faible et on ne retrouve pas sa contrepartie pour les voix d'hommes : cela laisse penser que *l'effet du F0 moyen a été relativement modéré*. Il est donc probable que les fréquences de résonance et en particulier les fréquences formantiques, non testées ici, ont joué un rôle prépondérant chez les auditeurs francophones à l'échelle du mot, comme cela était déjà le cas sur les séquences C + V.

On remarque enfin *une tendance intéressante concernant la plage de variation de F0* : bien que ces dernières ne soient pas significatives, il existe une corrélation positive avec les degrés de certitude pour les voix de femmes et une corrélation négative pour les voix d'hommes. Autrement dit, *plus la plage de variation est élevée pour les voix de femmes et plus les auditeurs francophones tendraient à être sûrs de leurs choix, et inversement pour les voix d'hommes*. La non-significativité de ces deux corrélations pourrait s'expliquer par le type de stimuli utilisé ici, à savoir les mots dissyllabiques : une même expérience conduite avec des énoncés plus longs (phrases) permettrait probablement d'obtenir des résultats significatifs pour ce paramètre acoustique.

Des tests statistiques identiques ont été réalisés sur les résultats obtenus par les auditeurs anglophones américains. En premier lieu, une ANOVA à un facteur (« type de mot ») portant sur les pourcentages d'identifications correctes a été effectuée. A l'instar des francophones, *l'analyse ne détecte ici aucun effet global significatif de ce facteur* ($F(5,2694)=1,92$; $p>0,05$). Le PLSD fait état de l'absence de différence significative entre les types de mot pris deux à deux. Une même ANOVA à un facteur a été conduite sur le degré de certitude des auditeurs anglophones. La figure 70, ci-dessous, illustre cette analyse.

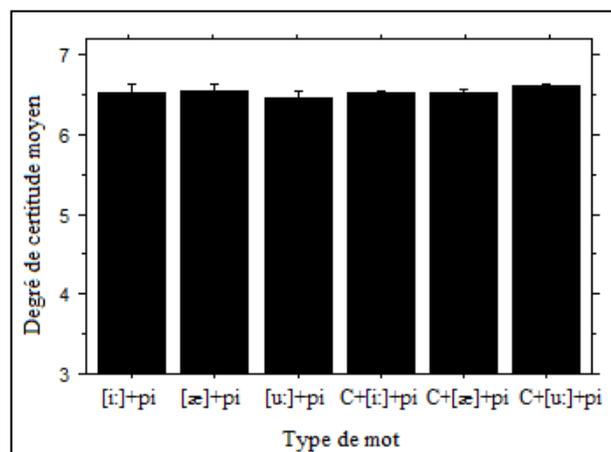


Figure 70 – Degré de certitude moyen (échelle de 0 à 7) exprimé par les auditeurs anglophones en fonction du type de mot dissyllabique, avec les barres d'erreur.

L'analyse indique qu'il n'existe par non plus d'effet global significatif du facteur « type de mot » sur les degrés de certitude émis par les auditeurs anglophones américains ($F(5,2694)=1,128$; $p>0,3$). Sans surprise, le test PLSD de Fisher ne décèle aucune différence significative entre les types de mot pris deux à deux.

Tout comme pour les auditeurs francophones, *les scores des auditeurs anglophones américains sur les mots dissyllabiques sont donc équivalents d'un type de mot à l'autre*. Les mêmes tests de Pearson que pour les francophones ont été effectués. Les résultats de ces tests figurent dans le tableau 46, ci-dessous.

Tableau 46 – Résultats des tests de Pearson conduits sur les données relatives aux degrés de certitude moyens des auditeurs anglophones pour les stimuli de type « mots entiers ». Le symbole « * » indique une corrélation significative entre les deux variables (seuil $p < 0,05$).

Résultats des tests de Pearson pour les auditeurs anglophones sur les mots dissyllabiques		Variable Y : Degré de certitude moyen
Variable X	H1-H2 (dB) voyelle [æ] voix F	$r(18)=0,340$; $z=1,374$; $p>0,1$
	H1-H2 (dB) voyelle [æ] voix H	$r(18)=-0,531$; $z=-2,291$; $p<0,05^*$
	F0 moyen (Hz) tous stimuli voix F	$r(54)=0,587$; $z=4,812$; $p<0,0001^*$
	F0 moyen (Hz) tous stimuli voix H	$r(54)=-0,137$; $z=-0,985$; $p>0,3$
	Durée (ms) tous stimuli toutes voix	$r(108)=0,011$; $z=0,118$; $p>0,9$
	Durée (ms) tous stimuli voix F	$r(54)=0,523$; $z=4,143$; $p<0,0001^*$
	Durée (ms) tous stimuli voix H	$r(54)=-0,379$; $z=-2,849$; $p<0,01^*$
	Plage de variation (dt) tous stimuli F	$r(54)=-0,139$; $z=-0,997$; $p>0,3$
	Plage de variation (dt) tous stimuli H	$r(54)=0,028$; $z=-0,198$; $p>0,8$

Chez les auditeurs anglophones, les tendances observées précédemment sur les stimuli de type C + V se trouvent globalement confirmées par les tests conduits sur les mots dissyllabiques. Ainsi, *la durée des stimuli toutes voix confondues semble ne pas avoir influencé du tout le degré de certitude des auditeurs anglophones*. Cependant, on remarque une tendance particulièrement intéressante lorsque les voix de femmes et d'hommes sont considérées séparément : il existe une corrélation positive significative entre la durée des mots et le degré de certitude exprimé sur les voix de femmes, et une corrélation négative significative pour le degré de certitude associé aux voix d'hommes. Cela suggère qu'*une vitesse d'élocution lente serait considérée par l'auditeur comme caractéristique des voix de femmes, et une vitesse d'élocution rapide comme caractéristique des voix d'hommes*. D'autre part, on constate que *contrairement aux auditeurs francophones, aucune tendance ne se dégage pour la plage de variation de F0 : ce paramètre n'a très probablement pas eu d'influence sur les réponses des auditeurs anglophones*.

La fréquence fondamentale moyenne semble quant à elle avoir joué un rôle déterminant pour les auditeurs anglophones américains. Il existe en effet une forte corrélation positive avec les degrés de certitude exprimés sur les voix de femmes : plus le

F0 moyen était élevé, plus les auditeurs ont eu tendance à être sûrs d'eux. On retrouve une tendance opposée sur les voix d'hommes, même si cette corrélation négative n'atteint pas le seuil de significativité. *Contrairement aux auditeurs francophones, la différence H1-H2, qui donne une estimation du type de phonation, a semble-t-il influencé elle aussi les réponses des auditeurs anglophones.* Le résultat du test de Pearson fait état d'une forte corrélation négative avec les degrés de certitude exprimés sur les voix d'hommes : plus la différence H1-H2 est basse (i.e. plus le QOG est faible), plus les auditeurs sont sûrs de leurs choix. Une corrélation positive apparaît sur les voix de femmes pour ce même paramètre, mais cette dernière n'est pas significative, probablement en raison de la moindre variabilité intra-locuteur et intra-genre.

4. DISCUSSION - CONCLUSION PARTIELLE

L'expérience d'identification du genre par la voix présentée dans ce chapitre se distinguait des études antérieures par plusieurs aspects. Elle a tout d'abord été menée conjointement sur des auditeurs anglophones américains et sur des francophones parisiens, avec des stimuli produits par des locuteurs de la langue correspondante. D'autre part, la technique du dévoilement progressif des stimuli (*gating*) a été utilisée. Enfin, les stimuli, extraits de voix naturelles non resynthétisées, avaient précédemment fait l'objet d'une analyse acoustique multidimensionnelle dans le chapitre 2. Ces particularités méthodologiques ont permis non seulement de confirmer des résultats obtenus précédemment, mais également de faire émerger des éléments nouveaux, donnant lieu à une meilleure compréhension du processus d'identification du genre par la voix.

Sur les segments de très courte durée de type « consonne isolée sourde », les pourcentages d'identifications correctes atteignent en moyenne 66 % chez les francophones et 76 % chez les anglophones, des scores déjà très largement supérieurs au seuil de chance (50 %). Ces résultats semblent valider l'**hypothèse 4**, selon laquelle l'auditeur est capable d'identifier le genre du locuteur à partir de segments très courts. Dans le détail, on constate que les meilleures performances sont observées sur les fricatives sourdes, avec des scores qui atteignent 70 % chez les francophones et 85 % chez les anglophones américains.

Ces chiffres sont assez proches ceux obtenus par Whiteside (1998b), à partir de segments extraits de phrases, et légèrement inférieurs à ceux de l'étude conduite par Schwartz (1968), sur des fricatives produites isolément et longues d'une seconde. Cela s'explique vraisemblablement par la longueur des stimuli utilisés : la mise en relation avec l'analyse acoustique des stimuli²⁰⁷ a en effet révélé l'existence de corrélations significatives entre les scores des auditeurs et la durée des stimuli, à l'échelle des consonnes isolées. Comme cela était attendu, l'analyse acoustique des stimuli indique également que les auditeurs, anglophones comme francophones, se sont probablement basés sur la position des zones de bruit (centre de gravité, pic spectral) pour identifier le genre de l'auditeur sur ce type d'item. Notons par ailleurs que la non-connaissance préalable par l'auditeur du type de phone présenté a semble-t-il affecté les résultats des

²⁰⁷ Effectuée dans le chapitre 2.

auditeurs francophones pour les fricatives sourdes : lorsqu'un phone de type [s] a été produit par un homme avec une zone de bruit particulièrement basse, il a parfois pu être pris pour un [ʃ] produit par une femme et inversement pour un [ʃ] prononcé par une femme avec une zone de bruit très élevée. Ce phénomène, décelé grâce aux degrés de certitude élevés associés à certaines réponses erronées, semble être apparu moins fréquemment chez les anglophones, ce qui expliquerait pourquoi leur pourcentage de bonnes réponses est sensiblement plus élevé que celui des francophones sur les fricatives sourdes. Les scores plus élevés obtenus par les auditeurs anglophones sur les occlusives sourdes sont quant à eux probablement liés à la durée de ces items, en moyenne supérieure sur les stimuli anglais, notamment grâce aux occlusives aspirées²⁰⁸.

Lorsque la consonne présentée était de type voisée, les scores des auditeurs ont fortement et significativement augmenté dans les deux langues, ce qui suggère que la fréquence fondamentale moyenne joue un rôle dans l'identification du genre par la voix tant chez les francophones que chez les anglophones américains. On observe néanmoins ici les premières traces de variations inter-langues. Chez les anglophones, le score frôle les 100 %, quand il n'atteint « que » 94,5 % chez les francophones. D'autre part, des corrélations fortes et significatives entre le F0 moyen des stimuli et le degré de certitude des anglophones ont été observées (corrélations positives sur les voix de femmes, négatives sur les voix d'hommes), ce qui n'a pas été le cas pour les auditeurs francophones. Ces deux éléments semblent indiquer que le F0 moyen est un indice acoustique plus déterminant chez les anglophones américains que chez les francophones lors d'une tâche d'identification du genre par la voix.

Sur les voyelles isolées, les scores des auditeurs anglophones sont comparables à ceux obtenus sur les consonnes voisées, ils progressent en revanche nettement chez les auditeurs francophones, tant sur le plan du pourcentage de bonnes réponses (98 %) que sur le degré de certitude. Cela laisse penser que la position des formants vocaliques jouerait un rôle important dans l'identification du genre par la voix chez les francophones, mais pas nécessairement chez les anglophones.

Le passage à des séquences plus longues de type *C + V* ou mot dissyllabique a entraîné des pourcentages de bonnes réponses et des degrés de certitude extrêmement

²⁰⁸ Voir section 3.1.1.4 du chapitre 2.

élevés dans les deux langues. La conduite de divers tests de Pearson sur ces stimuli a permis de mettre en relation les résultats de l'expérience avec différents paramètres acoustiques. De manière générale, les deux tendances évoquées ci-dessus ont été largement confirmées : la fréquence fondamentale moyenne semble avoir influencé plus fortement les jugements des auditeurs anglophones américains que ceux des francophones, la fréquence des formants vocaliques aurait à l'inverse joué un rôle plus important chez les francophones que chez les anglophones.

Ces données vont clairement dans le sens des résultats obtenus par Pépiot (2010, 2011) dans une précédente expérience d'identification du genre par la voix réalisée conjointement sur des francophones et des anglophones américains, avec des stimuli resynthétisés. Il est ainsi possible d'affirmer que les résultats, en apparence contradictoires, d'études menées exclusivement sur des anglophones américains ou sur des francophones, sont en réalité parfaitement compatibles. 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). Le moindre rôle joué par la position des formants vocaliques chez les anglophones américains pourrait s'expliquer en partie par l'importance des variations régionales sur ce paramètre (voir Clopper, 2004).

Outre ces deux principaux paramètres acoustiques, il est apparu que la qualité de voix²⁰⁹ (modale, craquée, soufflée), à travers la mesure H1-H2, a elle aussi joué un rôle dans l'identification du genre, mais uniquement auprès des auditeurs anglophones américains. Plus H1 était fort par rapport à H2 (voix de type soufflée ou *breathy*), plus les auditeurs ont été sûrs de leur catégorisation « voix de femme » ; plus H1 était faible (voix craquée ou *creaky*), plus les auditeurs ont été sûrs de leur catégorisation « voix d'homme ». Bien que des différences inter-genres sur ce paramètre aient été observées à plusieurs reprises chez les locuteurs anglophones américains (Henton & Bladon, 1988 ; Klatt & Klatt, 1990 ; Henton, 1992), aucune étude expérimentale n'avait à ma connaissance formellement démontré l'influence de ce paramètre dans l'identification du genre par la voix.

²⁰⁹ Employé ici comme synonyme de « type de phonation ».

Une tendance intéressante, bien que non-significative, est apparue sur le plan de la plage de variation de F0. Les résultats obtenus par les auditeurs francophones semblent indiquer qu'une plage de variation large aiderait à l'identification des voix de femmes, alors qu'une plage de variation étroite permettrait de reconnaître plus aisément les voix d'hommes. Cette tendance n'a pas été observée pour les anglophones américains, ce qui concorde avec les résultats antérieurs de Pépiot (2010, 2011) qui mettaient en évidence un rôle de l'intonation dans l'identification du genre par la voix chez les auditeurs francophone, et l'absence d'influence de ce paramètre acoustique chez les auditeurs anglophones.

On remarque par ailleurs qu'à partir des séquences *C + V*, la durée des stimuli n'a plus été corrélée aux scores des auditeurs, quelle que soit leur langue. On peut donc penser que dans une tâche d'identification du genre, hormis pour les phones isolés, la durée de l'input ne semble pas être déterminante : c'est la quantité d'informations (i.e. d'indices acoustiques) qui semble importer. Cependant, à l'échelle de mots dissyllabiques, il est apparu que la durée de l'énoncé, et donc la vitesse d'élocution, pouvait être en elle-même un indice acoustique permettant d'identifier le genre par la voix. Chez les anglophones américains comme chez les francophones, plus les mots produits par les femmes étaient longs, plus les auditeurs ont été sûrs de leur choix. Sur les voix d'hommes, une corrélation inverse a été observée, mais uniquement pour les anglophones : plus le mot était court, plus ces derniers ont eu tendance à exprimer un degré de certitude élevé. Tout comme pour le type de phonation, des différences acoustiques inter-genres sur le plan de la durée des énoncés ont déjà été constatées (Byrd, 1994 ; Simpson, 1998), mais je n'ai connaissance d'aucune étude expérimentale ayant mis en évidence (ou même testé) le rôle perceptif joué par ce paramètre.

A partir des résultats de cette expérience et en tenant compte des études antérieures, un tableau de synthèse a été élaboré (tableau 47, ci-après). Il fait état du degré d'influence estimé des différents paramètres acoustiques pouvant jouer un rôle dans l'identification du genre par la voix à partir de mots dissyllabiques, selon la langue maternelle des auditeurs (francophones ou anglophones américains).

Tableau 47 – Degré d'influence estimé des paramètres acoustiques pouvant jouer un rôle dans l'identification du genre par la voix à partir de mots dissyllabique, pour les auditeurs francophones et anglophones américains. Le symbole « \emptyset » indique l'absence d'influence, puis échelle graduée allant de « * » (très faible influence) à « ***** » (très forte influence).

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

Les données figurant dans ce tableau sont à prendre à titre indicatif. On sait en effet que l'influence des paramètres acoustiques dans ce type de tâche est relative : par exemple, lorsqu'une voix présente un F0 moyen à un niveau intermédiaire (i.e. ambigu), le rôle joué par les autres paramètres se trouve automatiquement renforcé (Coleman, 1976 ; Pépiot, 2010, 2011). D'autre part, la présente étude ne prétend pas être exhaustive et comporte certaines limites. Seules quatre voix différentes (deux femmes, deux hommes) ont été utilisées dans chaque langue pour obtenir les stimuli. De plus, en raison des critères de sélection (même tranche d'âge, même origine géographique, non-fumeurs, etc.), les différentes voix d'hommes et de femmes sont relativement homogènes, ce qui tend à limiter l'ampleur des corrélations entre les réponses des auditeurs et les divers paramètres acoustiques.

Néanmoins, au regard des importantes variations inter-langues observées entre les auditeurs francophones et anglophones américains à travers cette étude, il est possible d'affirmer que l'**hypothèse 5**, selon laquelle le processus d'identification du genre par la voix diffère en fonction de la langue maternelle de l'auditeur, a été largement validée. Cela suggère que les auditeurs possèdent des représentations mentales socialement construites de ce que sont les « voix de femmes » et les « voix d'hommes ».

On remarque par ailleurs que les résultats de l'expérience sont parfaitement cohérents avec l'analyse acoustique réalisée dans le chapitre 2 : tous les paramètres acoustiques ayant joué un rôle dans l'identification du genre par la voix présentaient bien des différences inter-genres significatives. Nous avons donc pour l'heure exploré expérimentalement deux aspects relatifs aux voix de femmes et d'hommes : les

différences acoustiques et l'identification du genre par la voix. Reste maintenant un dernier point à examiner : les conséquences des différences acoustiques inter-genres sur le traitement de la parole par l'auditeur. Cette question est abordée dans le chapitre 4, à travers une expérience de détection de mots.

CHAPITRE 4

SUR LES TEMPS DE TRAITEMENT DES VOIX DE FEMMES ET D'HOMMES

RESUME

Cette étude a pour objet les temps de traitement des voix de femmes et d'hommes. Plusieurs auteurs ont mis en évidence la difficulté accrue de l'identification des voyelles lorsque ces dernières sont produites avec un F0 élevé (Ryalls & Lieberman, 1982). Cela a-t-il des conséquences sur le traitement des mots ? Les voix de femmes sont-elles traitées plus lentement que les voix d'hommes ? Une expérience de détection de mots a été réalisée auprès d'auditeurs francophones, afin de tester le temps de réponse des participants en fonction du genre du locuteur ayant produit le mot-cible. Les résultats suggèrent que les voix d'hommes et de femmes sont traitées par l'auditeur à vitesse équivalente, mais néanmoins comme deux entités différentes. D'autre part, aucune corrélation significative entre la fréquence fondamentale moyenne du mot-cible et le temps de réponse des auditeurs n'a pu être mise en évidence.

ABSTRACT

This study deals with processing time of female and male speech. Several authors showed that vowel identification was more difficult on voices with a high F0 (Ryalls & Lieberman, 1982). Does this have consequences on word processing? Are female voices processed more slowly than male ones? A word spotting experiment was conducted with French native speakers in order to test the participants' response times, depending on whether the target word is produced by a male or a female voice. Results suggest that these two types of voice are processed equally fast, even though they seem processed as two different entities. Moreover, no significant correlation was found between mean F0 of the target word and response time.

MOTS-CLES : voix de femmes, voix d'hommes, fréquence fondamentale, temps de traitement, détection de mots.

KEYWORDS: female voices, male voices, processing time, fundamental frequency, word spotting.

Ce quatrième chapitre est consacré à l'expérience de détection de mots ayant pour but de mesurer les temps de traitement des voix d'hommes et de femmes²¹⁰. Une première section présente la méthode, la seconde expose les résultats, tout d'abord de manière brute, puis à travers leur analyse statistique.

1. METHODE

Dans cette section sont décrits de manière détaillée le choix du matériau linguistique et des locuteurs, la procédure d'enregistrement, l'obtention des stimuli, le choix des participants et la procédure expérimentale.

1.1 Matériau linguistique

La détection de mots est un paradigme fréquemment utilisé pour les expériences de psycholinguistique (voir notamment une étude pionnière de Marslen-Wilson & Tyler, publiée en 1980). Ce dispositif expérimental se caractérise par la présentation de *séries de mots* (ou de phrases) de longueur variable et se terminant par un *mot-cible*, préalablement communiqué au participant. La tâche de ce dernier est d'appuyer sur un bouton dès qu'il perçoit le mot-cible.

Afin de mener à bien cette expérience, du matériau linguistique en langue française était nécessaire. Le choix des mots a été réalisé sur la base de plusieurs critères : une longueur équivalente (dissyllabiques), une fréquence d'occurrence relativement élevée (figurant tous dans les 1500 mots les plus fréquents de la langue française²¹¹), un contenu émotionnel le plus neutre possible.

Au total 61 mots ont été sélectionnés, soit un mot-cible et 60 autres mots (dont sept réservés aux items d'entraînement)²¹². Voici la liste de ces mots :

- **Mot-cible** : *étage*. Ce dernier a été choisi en raison de sa voyelle initiale [e], présentant d'importantes différences formantiques entre hommes et femmes, et

²¹⁰ Une partie des résultats de cette expérience ont également été publiés dans Pépiot (2012).

²¹¹ Sur la base de données mises à disposition par le Ministère de l'Éducation Nationale, de la Jeunesse et de la Vie Associative : <http://eduscol.education.fr/cid47916/liste-des-mots-classee-par-frequence-decroissante.html>.

²¹² Pour plus de détails sur leur utilisation dans le dispositif expérimental, voir section 1.6.

donc susceptible de maximiser les effets recherchés dans cette expérience (différences de temps de traitement entre voix de femmes et voix d'hommes).

- **Mots (non cible) :** *abeille, action, ami, après, argent, article, avance, cadeau, cerveau, chanter, chemin, cheveu, compter, creuser, debout, depuis, dimanche, facile, façon, falloir, glaçon, humain, immense, instant, maison, marché, mari, marteau, matin, moitié, moment, moyen, objet, offrir, pareil, parent, partie, passer, photo, pourquoi, pouvoir, quitter, savoir, sentir, sérieux, service, silence, tableau, tirer, tourner, utile, visage, voiture.* Mots réservés aux items d'entraînement : *choisir, heureux, jamais, million, oser, oubli, plusieurs.*

1.2 Choix des locuteurs pour les enregistrements

Pour ces enregistrements, j'ai fait appel à huit locuteurs francophones, âgés de 20 à 34 ans. Ces derniers sont tous locuteurs du français dit *parisien*²¹³ ou *standard*, non-fumeurs et ne présentant pas de trouble de la parole. Voici les autres informations pertinentes les concernant :

- **Locutrice numéro 1 (F1) :** 22 ans, francophone native, étudiante en sciences du langage niveau L3, F0 moyen : 228 Hz²¹⁴.
- **Locutrice numéro 2 (F2) :** 26 ans, francophone native, étudiante en sciences du langage niveau M2, F0 moyen : 192 Hz.
- **Locutrice numéro 3 (F3) :** 24 ans, francophone native, étudiante en sciences du langage niveau M2, F0 moyen : 190 Hz.
- **Locutrice numéro 4 (F4) :** 21 ans, francophone native, étudiante en information-communication niveau L3, F0 moyen : 195 Hz.
- **Locuteur numéro 1 (H1) :** 20 ans, francophone natif, étudiant en information-communication niveau L2, F0 moyen : 116 Hz.

²¹³ Les individus désignés ici comme locuteurs du français *parisien* vivent en Ile-de-France depuis plus de 5 ans et ne présentent pas d'accent régionalement marqué.

²¹⁴ Les F0 moyens ont tous été mesurés avec le logiciel *Praat*, sur l'ensemble des 122 phrases (61*2) lues pour l'expérience.

- **Locuteur numéro 2 (H2)** : 34 ans, francophone natif, étudiant en sciences du langage niveau M2, F0 moyen : 135 Hz.
- **Locuteur numéro 3 (H3)** : 32 ans, francophone natif, doctorant en sciences du langage, F0 moyen : 152 Hz.
- **Locuteur numéro 4 (H4)** : 22 ans, francophone natif, étudiant en information-communication niveau L3, F0 moyen : 127 Hz.

1.3 Procédure d'enregistrement

Afin d'homogénéiser les paramètres prosodiques sur les 61 mots²¹⁵ devant servir pour l'expérience, chacun d'entre eux a été placé dans le contexte suivant : « *Il a dit MOT deux fois* ». L'utilisation de ce type contexte permet notamment de neutraliser l'« effet de liste » (F0 montant sur tous les items d'une liste de mots jusqu'à l'avant-dernier, puis descendant sur le dernier item).

Tout comme lors les enregistrements réalisés pour les expériences décrites dans le chapitre précédent, j'ai utilisé un enregistreur numérique portable *Edirol R-09HR* de marque Roland, comportant un microphone doté d'une bande passante allant de 20 Hz à 40 kHz. Les enregistrements ont été effectués au format Wave²¹⁶, avec la sensibilité du microphone réglée en mode élevé, et sans utiliser les différentes options disponibles, telles que l'« Auto Gain Control²¹⁷ » et le « Low Cut²¹⁸ », qui auraient pu engendrer des distorsions du signal. Pour les précautions d'enregistrement, j'ai là encore tenu compte des conseils dispensés par Philippe Martin (2008) et Jean-Michel Tarrier (2003).

Pour chaque locuteur enregistré, une phase de réglage avait lieu dans un premier temps. Durant cette phase, j'ajustais le volume d'enregistrement de manière adéquate (intensité élevée, mais sans saturation) pendant que le participant parlait avec le microphone positionné à une trentaine de centimètres de sa bouche, mais sans être

²¹⁵ La liste des mots est visible dans la section 1.1.

²¹⁶ En stéréo (pas d'enregistrement direct en mono possible avec l'*Edirol R-09HR*), avec encodage à 44,1 kHz. Ces fichiers ont ensuite été convertis en mono, en préservant la fréquence d'échantillonnage.

²¹⁷ Fonctionnalité permettant l'ajustement automatique du volume d'enregistrement.

²¹⁸ Suppression des basses fréquences (utilisé notamment pour limiter les bruits de vent lors d'enregistrements en extérieur).

enregistré. Par la suite, les locuteurs étaient invités à lire les phrases à une vitesse « normale, c'est-à-dire pas trop rapide » et comme si ces dernières étaient des « phrases isolées ». Chacune des phrases devait être lue deux fois, afin de me laisser la possibilité ultérieure de choisir l'occurrence la plus satisfaisante acoustiquement. La distance entre la bouche et le microphone était maintenue constante, toujours aux environs de trente centimètres.

Afin de limiter au maximum le bruit de fond, les enregistrements ont été réalisés dans une pièce isolée phoniquement (chambre anéchoïque), à l'écart des bruits de circulation, sans appareil électrique bruyant en fonctionnement (ordinateurs, ventilateurs, néons...). L'enregistreur était posé sur un socle, lui-même posé sur un tapis en mousse afin de l'isoler au mieux de toutes vibrations. De plus, je me suis assuré que les participants ne généraient pas de bruits parasites (coups sur la table, le sol, ou même froissement de la feuille sur laquelle était écrite la phrase à lire).

Voyons maintenant en détail les opérations effectuées sur les enregistrements « bruts » afin d'obtenir les stimuli de l'expérience.

1.4 Obtention des stimuli

Comme je l'ai mentionné dans la section précédente, les mots isolés destinés à être utilisés pour l'expérience de détection de mots ont été enregistrés en contexte (« *Il a dit MOT deux fois* »). Il m'a fallu dans un premier temps extraire chaque mot de son contexte, puis constituer un unique fichier son par *série* de stimuli²¹⁹.

1.4.1 Extraction des stimuli

L'extraction des stimuli de leur contexte « *Il a dit MOT deux fois* » a été réalisée dans le logiciel *Praat*²²⁰. Cette opération a été effectuée manuellement : pour ce faire, je me suis appuyé à la fois sur le spectrogramme, l'onde sonore et la courbe de F0 afin de déterminer avec précision les frontières de mots. Une fois ces frontières identifiées, une première sélection du mot a été effectuée. Une illustration de cette opération est visible

²¹⁹ Cette opération était nécessaire pour mettre au point l'expérience dans le logiciel *Perceval 3.0.5.0*. Voir section 1.6 pour plus de détails.

²²⁰ *Praat 5.1.44*. Voir Boersma & Weenink (2010).

sur la figure 71 ci-après. Dans un deuxième temps, les bornes gauche et droite de la sélection ont été décalées au passage par zéro le plus proche (opération automatisée par la création d'un script) afin d'éviter les *clics* parasites.

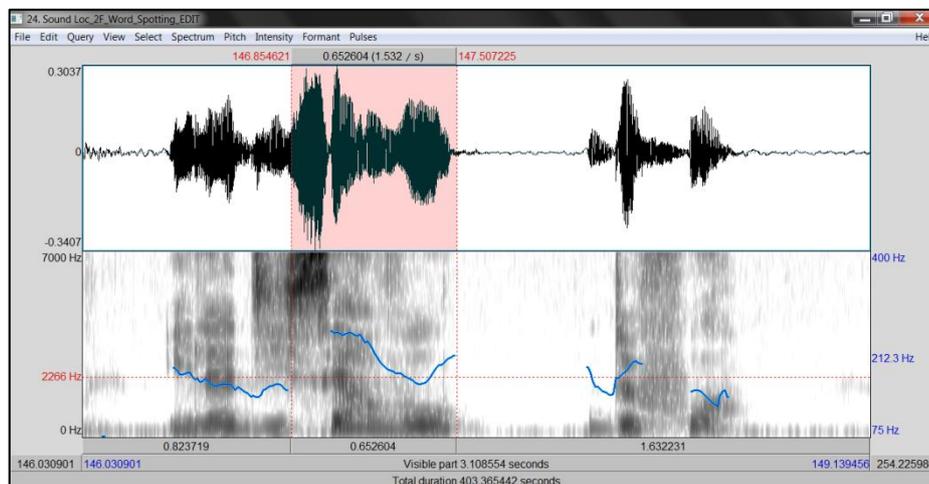


Figure 71 – Extraction du mot cerveau de la phrase il a dit « cerveau » deux fois prononcée par la locutrice F2, à l'aide du logiciel Praat. La partie supérieure représente l'onde sonore ; la partie inférieure le spectrogramme et la courbe de F0 (en bleu). La zone sélectionnée figure en rose sur l'onde sonore.

Pour finir, chacun des mots a été exporté dans un fichier au format Wave (en gardant la fréquence d'échantillonnage à 44 kHz).

1.4.2 Création d'un fichier son par série de stimuli

L'utilisation du logiciel *Perceval* pour la passation de l'expérience m'a contraint à créer un seul fichier son par série de stimuli. Cette opération a également été réalisée dans le logiciel *Praat*, à l'aide d'un script que j'ai programmé à cet effet. Ce dernier m'a permis de compiler les fichiers sons contenant les mots d'une série dans un seul fichier son, tout en insérant un silence d'une durée de 600 ms entre chaque mot²²¹. Les séries de stimuli sont librement téléchargeables en ligne²²².

²²¹ Voir section 1.6.

²²² Voir annexe 1.

Au total, 70 fichiers sons²²³ correspondant aux 70 séries de mots destinées à l'expérience de détection de mots ont ainsi été obtenus.

1.5 Participants

Au total, 25 auditeurs (8 hommes et 17 femmes) ont pris part à cette expérience de détection de mots. Ces participants sont tous des locuteurs francophones natifs vivant en région parisienne, ne présentant pas de trouble du langage ou de l'audition, et âgés de 18 à 65 ans. La moyenne d'âge est de 27,6 ans (36,1 ans pour les hommes, 23,6 ans pour les femmes). L'âge, le genre et la profession de chacun des participants sont visibles en annexe 4.

1.6 Procédure expérimentale

La mise en place d'une expérience de détection de mots nécessite la maîtrise de plusieurs variables inhérentes à ce paradigme, et la neutralisation de divers biais. Quatre conditions expérimentales doivent ici être utilisées pour tester les hypothèses :

- **Condition A** (*homogène*) : contexte *voix d'hommes* avant mot-cible expérimental *voix d'homme*.
- **Condition B** (*homogène*) : contexte *voix de femmes* avant mot-cible expérimental *voix de femme*.
- **Condition C** (*non-homogène*) : contexte *voix de femmes* avant mot-cible expérimental *voix d'homme*.
- **Condition D** (*non-homogène*) : contexte *voix d'hommes* avant mot-cible expérimental *voix de femme*.

Afin de maximiser son effet, le *contexte* s'étend sur une série dite *pré-expérimentale*, longue alternativement de trois à quatre items (mot-cible inclus) et sur les mots non-cibles de la série expérimentale, c'est-à-dire quatre items, soit un total de sept à huit mots précédant directement le mot-cible expérimental. Des séries de distracteurs ont également été utilisées et un schéma de base est ainsi répété : deux séries de distracteurs,

²²³ Format Wave avec fréquence d'échantillonnage de 44 kHz.

une série pré-expérimentale, une série expérimentale. La longueur des séries de distracteurs varie de deux à sept items, mot-cible inclus. Seuls les temps de réponse portant *sur les mots-cibles expérimentaux* seront effectivement pris en compte.

L'expérience se divise en quatre *blocs* qui comportent à chaque fois les quatre conditions expérimentales dans un ordre différent :

- **Bloc 1** : (2 séries de distracteurs), Condition A, (2 séries de distracteurs), Condition B, (2 séries de distracteurs), Condition C, (2 séries de distracteurs), Condition D.
- **Bloc 2** : (2 séries de distracteurs), Condition B, (2 séries de distracteurs), Condition A, (2 séries de distracteurs), Condition D, (2 séries de distracteurs), Condition C.
- **Bloc 3** : (2 séries de distracteurs), Condition D, (2 séries de distracteurs), Condition C, (2 séries de distracteurs), Condition B, (2 séries de distracteurs), Condition A.
- **Bloc 4** : (2 séries de distracteurs), Condition C, (2 séries de distracteurs), Condition D, (2 séries de distracteurs), Condition A, (2 séries de distracteurs), Condition B.

Ainsi chaque condition, qui contient deux séries de mots (*pré-expérimentale* et *expérimentale*), est testée quatre fois dans l'expérience, en occupant chacune des positions (1, 2, 3 ou 4) possibles à l'intérieur des blocs.

La longueur des séries (en termes de nombre d'items) est maintenue constante dans chaque bloc, tout en suivant un schéma suffisamment aléatoire pour ne pas être mémorisé par le participant : 2, 6, 3, 5, 7, 3, 4, 5, 2, 7, 3, 5, 6, 2, 4, 5 (mots-cibles compris). Un même mot-cible, le mot *étage*, a été utilisé pour toute l'expérience, afin de limiter les divers biais qu'aurait pu induire l'utilisation de mots-cibles variés. En revanche, un autre élément a dû par conséquent être pris en compte : à force de détecter ce même mot-cible, il est probable que les auditeurs améliorent globalement leur temps de réponse au fur et à mesure qu'ils avancent dans l'expérience. Pour neutraliser ce biais, une moitié d'auditeurs s'est vu présenter les blocs dans l'ordre 1, 2, 3, 4 et l'autre moitié dans l'ordre 3, 4, 1, 2. De plus, une vérification statistique sera réalisée *a posteriori* sur ce point.

Le mot *étage* a été choisi comme mot-cible en raison de la voyelle /e/ placée en initiale, qui présente généralement d'importantes différences formantiques femmes-hommes. Cela peut potentiellement permettre de maximiser les différences de temps de traitement pour ce mot selon qu'il est produit par une voix d'homme ou par une voix de femme.

Les 53 mots non-cibles²²⁴ apparaissent une fois par bloc, toujours dans un ordre différent. Une attention toute particulière a été portée au choix des mots précédant directement le mot-cible expérimental : aucun d'entre eux ne présente de voyelle en initiale de mot, ni d'autres similarités phonétiques fortes avec le mot-cible *étage* ; de même, ils n'appartiennent pas à un champ sémantique trop proche. De plus, sur l'ensemble de l'expérience, ces mots n'apparaissent jamais plus d'une fois dans cette position. Quant aux séries d'entraînement, sept mots spécifiques ont été utilisés²²⁵, chacun d'entre eux apparaissant trois fois.

Le répartition des voix²²⁶ pour les différents mots s'est faite en appliquant plusieurs règles, établies en vue de limiter de façon optimale les différents biais possibles :

- Concernant les **mots-cibles expérimentaux**, c'est la même voix d'homme (conditions 1 et 3) et la même voix de femme (conditions 2 et 4) qui sont présentées à l'intérieur d'un même bloc (H1 et F1 pour le bloc 1, H2 et F2 pour le bloc 2, etc.). Ces voix n'apparaissent sur aucun autre mot-cible dans le même bloc (séries de distracteurs ou pré-expérimentales), ni sur un mot non-cible des séries expérimentales où elles figurent en mot-cible. En revanche, elles apparaissent sur d'autres mots non-cibles, au même titre que les autres voix, et ce au minimum une fois avant d'apparaître en position expérimentale (la découverte d'une nouvelle voix pourrait modifier les temps de réponse des auditeurs), et une fois en série pré-expérimentale pour les conditions homogènes (conditions 1 et 3).

²²⁴ Liste des mots visible en section 1.1.

²²⁵ Voir section 1.1.

²²⁶ Pour rappel, 8 locuteurs (4 hommes et 4 femmes) ont été enregistrés sur l'ensemble des mots : H1, H2, H3, H4, F1, F2, F3, F4.

- **De manière générale**, aucune voix n'apparaît plus de deux fois dans une même série, ni sur deux mots consécutifs. D'autre part, ne figurent pas plus de deux mots-cibles d'affilée et pas plus de trois mots consécutifs avec une voix du même genre (H ou F)²²⁷. Au total, chaque voix apparaît deux fois par bloc en position de mot-cible (jamais sur deux mots-cibles consécutifs), et huit ou neuf fois par bloc, tous items confondus. Dans les séries d'entraînement (ou *warm-up*), chaque voix apparaît trois ou quatre fois. Le tableau 48 ci-dessous décrit la répartition des différentes voix d'hommes et de femmes utilisées pour l'expérience.

Tableau 48 – Répartition des stimuli de type « voix d'hommes » et « voix de femmes » dans l'expérience de détection de mots.

	Voix d'hommes				Voix de femmes			
	H1	H2	H3	H4	F1	F2	F3	F4
<i>Warm-up</i>	3	4	3	4	3	3	3	4
Bloc 1	8	9	9	8	9	8	9	9
Bloc 2	9	8	8	9	9	9	8	9
Bloc 3	9	8	9	9	8	9	9	8
Bloc 4	9	9	9	8	8	9	9	8
Total	38	38	38	38	37	38	38	38
	188				187			

Le plan d'expérience, construit selon les règles énoncées plus haut, est visible ci-dessous (tableau 49). L'ensemble des séries de mots de l'expérience (les séries d'entraînement ainsi que celles qui forment les quatre blocs), sont ici présentées en détail. Le code du locuteur produisant chaque mot (F1, H1, F2, H2, etc.) est également indiqué.

Tableau 49 – Ordre de présentation et nature des stimuli de l'expérience de détection de mots. La colonne de gauche indique le type de série. Le début des différents « blocs » est mentionné sur une ligne dédiée. Les mots-cibles de chaque série ainsi que tous les items des séries pré-expérimentales et expérimentales apparaissent en gras. Les mots-cibles expérimentaux sont en rouge.

<i>Warm up 1</i>	oser H1	jamais F2	étage F1				
<i>Warm up 2</i>	oubli F4	heureux H2	million F3	étage H1			
<i>Warm up 3</i>	plusieurs H4	étage H3					
<i>Warm up 4</i>	jamais F4	choisir F2	oser H3	plusieurs H4	heureux H2	million F1	étage F3
<i>Warm up 5</i>	choisir F4	oubli H3	jamais H4	oser F1	étage H2		
<i>Warm up 6</i>	heureux H1	million H2	choisir F3	oubli H4	plusieurs F2	étage F4	

²²⁷ A l'exception des séries pré-expérimentales et expérimentales homogènes (conditions 1 et 3).

<i>Bloc 1</i>							
Distracteur 1	silence F1	étage H3					
Distracteur 2	pourquoi H2	tourner H4	compter F4	instant F3	abeille H1	étage F2	
Pré-exp. Cond. A	tirer H4	moment H1	étage H2				
Exp. Cond. A	maison H3	facile H2	cadeau H4	chemin H3	étage H1		
Distracteur 1	voiture F4	service F3	dimanche H2	avance F1	glaçon H4	objet H3	étage F4
Distracteur 2	cerveau F3	partie F2	étage H4				
Pré-exp. Cond. B	visage F4	marché F1	article F2	étage F3			
Exp. Cond. B	moyen F2	utile F3	moitié F4	façon F2	étage F1		
Distracteur 1	humain H2	étage H3					
Distracteur 2	quitter H3	sentir F2	matin F1	argent H2	cheveu F4	pareil H1	étage F3
Pré-exp. Cond. C	debout F1	passer F3	étage F4				
Exp. Cond. C	depuis F2	action F4	savoir F3	creuser F1	étage H1		
Distracteur 1	tableau H3	parent F1	après H1	sérieux F3	immense F4	étage H2	
Distracteur 2	offrir H4	étage F2					
Pré-exp. Cond. D	ami H1	pouvoir H2	mari H3	étage H4			
Exp. Cond. D	chanter H3	photo H4	falloir H1	marteau H2	étage F1		
<i>Bloc 2</i>							
Distracteur 1	article F2	étage H1					
Distracteur 2	utile H3	cerveau F1	offrir H1	quitter H4	maison F4	étage H3	
Pré-exp. Cond. B	ami F3	moment F2	étage F1				
Exp. Cond. B	sentir F4	matin F3	falloir F1	compter F4	étage F2		
Distracteur 1	façon H2	chemin H3	silence F3	tourner H1	action F2	debout F1	étage H4
Distracteur 2	creuser H1	passer F4	étage F3				
Pré-exp. Cond. A	cheveu H4	pareil H2	immense H3	étage H1			
Exp. Cond. A	pourquoi H4	marché H1	service H3	voiture H4	étage H2		
Distracteur 1	dimanche F1	étage F4					
Distracteur 2	tirer H2	savoir F3	argent F1	instant H1	abeille F2	sérieux H4	étage F3
Pré-exp. Cond. D	objet H2	mari H3	étage H4				
Exp. Cond. D	après H1	partie H2	glaçon H3	chanter H4	étage F2		

Distracteur 1	marteau F1	facile F2	moitié H1	parent F4	depuis H2	étage H3	
Distracteur 2	photo H4	étage F4					
Pré-exp. Cond. C	cadeau F2	humain F3	tableau F4	étage F1			
Exp. Cond. C	moyen F3	avance F2	visage F1	pouvoir F4	étage H2		
Bloc 3							
Distracteur 1	sérieux F1	étage H1					
Distracteur 2	argent F2	action H4	pareil F3	matin F4	façon H3	étage F1	
Pré-exp. Cond. D	partie H1	objet H4	étage H2				
Exp. Cond. D	tourner H3	article H1	silence H2	sentir H4	étage F3		
Distracteur 1	maison F2	cheveu F1	cadeau H3	savoir H1	abeille F4	compter H2	étage H4
Distracteur 2	cerveau F3	ami H2	étage F2				
Pré-exp. Cond. C	falloir F1	pourquoi F3	passer F2	étage F4			
Exp. Cond. C	avance F3	moitié F1	chanter F4	moment F2	étage H3		
Distracteur 1	photo H1	étage F2					
Distracteur 2	après H3	parent F4	tirer H2	facile H1	quitter F3	utile H3	étage H4
Pré-exp. Cond. B	visage F2	depuis F3	étage F1				
Exp. Cond. B	instant F4	chemin F2	moyen F1	dimanche F4	étage F3		
Distracteur 1	glaçon H2	immense H4	offrir F2	humain H3	marteau F1	étage H1	
Distracteur 2	pouvoir F3	étage F4					
Pré-exp. Cond. A	tableau H4	voiture H3	debout H1	étage H2			
Exp. Cond. A	marché H4	creuser H1	service H2	mari H4	étage H3		
Bloc 4							
Distracteur 1	après H3	étage H2					
Distracteur 2	instant F4	maison H1	debout F2	cadeau H4	falloir H3	étage F3	
Pré-exp. Cond. C	facile F1	cerveau F3	étage F2				
Exp. Cond. C	visage F4	creuser F3	offrir F1	moyen F2	étage H4		
Distracteur 1	moment H2	humain F4	mari H1	action H4	marché F2	objet F3	étage H1
Distracteur 2	pareil H3	dimanche F1	étage F3				
Pré-exp. Cond. D	quitter H1	pourquoi H2	utile H4	étage H3			

Exp. Cond. D	chemin H2	article H1	sérieux H3	partie H4	étage F4		
Distracteur 1	avance F1	étage H2					
Distracteur 2	chanter H3	voiture H1	silence F2	passer F4	service H4	façon F3	étage F1
Pré-exp. Cond. A	marteau H4	abeille H3	étage H1				
Exp. Cond. A	compter H2	tableau H1	cheveu H3	tirer H2	étage H4		
Distracteur 1	moitié F1	parent H2	pouvoir F4	sentir F3	ami H1	étage F2	
Distracteur 2	immense H2	étage H3					
Pré-exp. Cond. B	depuis F3	argent F2	matin F4	étage F1			
Exp. Cond. B	savoir F2	glaçon F3	tourner F1	photo F2	étage F4		

L'expérience a été réalisée à l'aide d'un ordinateur portable, du logiciel *Perceval 3.0.5.0*²²⁸ et de la *Button Box*, périphérique mis au point par le Laboratoire Parole et Langage de l'Université de Provence (voir figure 72, ci-après). L'utilisation d'un tel boîtier a l'avantage de permettre une mesure fiable et précise des temps de réponse des participants, à l'inverse, par exemple, d'un clavier d'ordinateur. La mise en place de cette expérience dans *Perceval* a par ailleurs nécessité la programmation d'un script dédié²²⁹. Les passations se sont déroulées dans une pièce calme (salle de l'Université Paris 8 réservée à cet effet).

²²⁸ Voir Ghio et al. (2003) et André et al. (2003). Le logiciel est téléchargeable gratuitement à l'adresse suivante : <http://aune.lpl.univ-aix.fr/~lpldev/perceval/percevaldownload.html>.

²²⁹ Le script ainsi qu'un lien pour le télécharger directement au format adapté au logiciel *Perceval* sont visibles en annexe 7.



Figure 72 – Une participante à l'expérience de détection de mots, photographiée au début de l'expérience, lors de l'affichage de la consigne écrite à l'écran. La Button Box est visible dans la partie centrale basse de la photographie.

Le déroulement de l'expérience a été rigoureusement identique pour chaque participant. Tout d'abord, plusieurs informations sur l'auditeur étaient recueillies et rentrées dans le logiciel : son âge, son sexe, ses initiales, sa ou ses langue(s) maternelle(s) et sa profession. C'est également à ce moment que j'attribuais l'ordre de présentation des stimuli : blocs 1, 2, 3, 4 pour la moitié des participants (« groupe 1 ») ; blocs 3, 4, 1, 2 pour l'autre moitié (« groupe 2 »)²³⁰. La consigne orale suivante était ensuite communiquée au participant : « Durant cette expérience, vous allez entendre des mots isolés. A chaque fois que vous entendrez le mot *étage*, il faudra appuyer le plus rapidement possible sur le bouton bleu du boîtier. Vous pouvez tenir le boîtier de la manière qui vous convient le mieux : essayez simplement de garder la même position et d'utiliser la même main et le même doigt durant toute l'expérience ». Une fois installé devant l'écran d'ordinateur et équipé d'un casque audio²³¹, le participant était à nouveau invité, par consigne écrite affichée à l'écran, à appuyer sur le bouton bleu de la *button box* le plus rapidement possible dès qu'il entendrait le mot *étage* (voir figure 72).

²³⁰ Treize participants ont été placés dans le « groupe 1 » et 12 dans le « groupe 2 ». Voir annexe 4 pour la composition détaillée des deux groupes.

²³¹ Casque fermé de marque *Sony*, disposant d'une bande passante de 5 Hz à 28 kHz.

Dans un premier temps, six séries de mots d'entraînement étaient présentées. L'expérience était ensuite automatiquement suspendue : je m'assurais à cet instant que le participant avait bien compris le principe de l'expérience, et qu'il n'avait aucune question. Ce dernier était alors prévenu qu'il ne serait plus possible d'arrêter l'expérience avant son terme, c'est-à-dire 8 minutes plus tard. Durant toute la durée de cette expérience, aucun stimulus visuel n'était présenté à l'écran. Les stimuli audio ont été présentés avec un intervalle inter-stimulus²³² de 600 ms, ce même délai étant également appliqué entre la fin d'une série et le début de la suivante. Le choix de cet intervalle relativement court a été effectué dans le but de maintenir éveillée l'attention du sujet tout au long de l'expérience tout en limitant la durée totale de celle-ci.

D'autre part, je me suis systématiquement assuré de fermer tous les programmes et tâches de fond autres que *Perceval* avant le lancement de l'expérience, dans le but d'éviter tout traitement parallèle qui aurait pu perturber le recueil des temps de réponse par le logiciel.

La méthodologie a maintenant été exposée. Dans la section suivante, je décris brièvement l'analyse des données.

²³² Délai entre la fin d'un mot et le début du suivant.

2. ANALYSE DES DONNEES

A la fin de chaque passation, les temps de réponse (exprimés en millisecondes) pour l'ensemble des séries de stimuli étaient automatiquement inscrits par *Perceval* dans un fichier au format compatible avec le logiciel *Excel*.

Ces temps de réponse affichés correspondent au délai entre le début de la présentation de la série de stimuli et l'appui sur le bouton par le participant. Afin d'obtenir les données pertinentes, c'est-à-dire le temps de réponse du sujet à partir du début du mot-cible, il m'a donc fallu effectuer systématiquement le calcul suivant : « *Temps de réponse affiché par Perceval* » - « *Délai entre le début de la série et le début du mot-cible* ».

Au total, 64 temps de réponse par participant ont été collectés, correspondant à tous les mots-cibles de l'expérience (hors séries d'entraînement), qu'ils apparaissent dans des séries de distracteurs, pré-expérimentales ou expérimentales. *Seuls les temps de réponse correspondant aux mots-cibles des séries expérimentales ont été effectivement retenus*. D'autre part, parmi ces derniers, aucune mesure pouvant être considérée comme « absurde²³³ » n'a été observée : tous les temps de réponse ont donc été conservés et pris en compte pour les résultats. Seize mesures ont ainsi été relevées par participant (4 pour chaque condition expérimentale). Pour l'ensemble des 25 participants, cela correspond à un total de 400 mesures, soit 100 temps de réponse pour chacune des quatre conditions expérimentales.

Les résultats bruts ainsi que leur analyse statistique sont présentés dans la section suivante.

²³³ Temps de réponse excessivement court (en dessous d'un seuil raisonnable fixé à 200 ms) ou excessivement long (supérieur à 1000 ms).

3. RESULTATS

Dans un premier temps, les résultats de l'expérience de détection de mots sont présentés de manière brute et synthétique²³⁴ pour les hommes, les femmes, puis pour l'ensemble des auditeurs. Une deuxième partie est consacrée à l'analyse statistique de ces résultats.

3.1 Résultats bruts

Les temps de réponse²³⁵ obtenus pour la reconnaissance du mot-cible *étage* dans les quatre conditions expérimentales pour les 17 auditrices sont visibles dans le tableau 50 ci-dessous.

Tableau 50 – Nombre de réponses recueillies, temps de réponse moyens et écarts-types (exprimés en millisecondes) pour les 17 auditrices dans chacune des quatre conditions expérimentales, avec le mot-cible étage.

	Nombre de réponses recueillies	Temps de réponse moyen (ms)	Ecart-type (ms)
Condition A Mot-cible H avec contexte H	68	510	102
Condition B Mot-cible F avec contexte F	68	508	103
Condition C Mot-cible H avec contexte F	68	485	88
Condition D Mot-cible F avec contexte H	68	489	103
Toutes conditions	272	498	99

Les temps de réponse obtenus pour la reconnaissance du mot-cible dans les quatre conditions expérimentales pour les huit auditeurs masculins sont présentés dans le tableau 51 ci-après.

²³⁴ Le fichier contenant les résultats détaillés, participant par participant, est téléchargeable en ligne : voir annexe 9.

²³⁵ Tous les temps de réponse donnés dans cette section sont exprimés en millisecondes et arrondis à l'unité.

Tableau 51 – Nombre de réponses recueillies, temps de réponse moyens et écarts-types (exprimés en millisecondes) pour les huit auditeurs masculins dans chacune des quatre conditions expérimentales, avec le mot-cible étage.

	Nombre de réponses recueillies	Temps de réponse moyen (ms)	Ecart-type (ms)
Condition A Mot-cible H avec contexte H	32	482	103
Condition B Mot-cible F avec contexte F	32	467	112
Condition C Mot-cible H avec contexte F	32	464	106
Condition D Mot-cible F avec contexte H	32	443	104
Toutes conditions	128	465	106

On constate que de manière générale, les temps de réponse des participants masculins et féminins semblent relativement proches. Il en va de même pour les différences relatives entre les quatre conditions expérimentales, qui paraissent similaires pour les deux genres.

Voyons maintenant les résultats dans leur ensemble : le tableau 52, ci-après, montre les temps de réponse obtenus pour la reconnaissance du mot-cible *étage* dans les quatre conditions expérimentales pour les 25 participants (8 hommes et 17 femmes). Le nombre de réponses recueillies pour chaque condition ainsi que les écarts-types sont également précisés.

Tableau 52 – Nombre de réponses recueillies, temps de réponse moyens et écarts-types (exprimés en millisecondes) pour l'ensemble des auditeurs dans chacune des quatre conditions expérimentales, avec le mot-cible étage.

	Nombre de réponses recueillies	Temps de réponse moyen (ms)	Ecart-type (ms)
Condition A Mot-cible H avec contexte H	100	502	103
Condition B Mot-cible F avec contexte F	100	495	107
Condition C Mot-cible H avec contexte F	100	478	94
Condition D Mot-cible F avec contexte H	100	474	105
Toutes conditions	400	487	103

On remarque que *les temps de réponse moyens sont relativement équivalents* entre les conditions A et B d'une part, et entre les conditions B et C d'autre part, c'est-à-dire *entre les mots-cibles produits par des hommes et ceux produits par des femmes, en contexte équivalent*. Une différence assez importante apparaît en revanche entre les conditions A et C, ainsi que B et D (temps de réponse plus courts dans les conditions C et D), laissant supposer *un possible effet du contexte (homogène ou non-homogène avec le mot-cible) sur les temps de traitement des mots-cibles*. D'autre part, notons que les temps de réponse des auditeurs masculins sont globalement inférieurs à ceux des femmes, mais que les différences relatives entre les conditions expérimentales ne semblent pas varier de manière importante entre ces deux groupes.

Les résultats bruts ont maintenant été présentés. Ces derniers ont fait l'objet de différents tests statistiques, dont les résultats sont décrits dans la section suivante.

3.2 Analyse statistique

Afin de vérifier diverses interactions possibles entre les facteurs et d'établir si les différences décrites dans la section précédente sont significatives, j'ai procédé à une analyse statistique des résultats à l'aide du logiciel *StatView*²³⁶.

3.2.1 Différences de temps de réponse entre auditeurs féminins et masculins

Dans un premier temps, j'ai souhaité vérifier si le genre des auditeurs avait joué de manière significative sur les temps de réponse de manière absolue, c'est-à-dire sur l'ensemble de l'expérience, mais surtout relativement aux différentes conditions expérimentales. Pour cela, j'ai réalisé un test statistique de type ANOVA à deux facteurs, « genre des auditeurs » et « condition expérimentale », sur les temps de réponse des auditeurs. Le graphique des interactions correspondant à cette analyse est présenté ci-après (figure 73)²³⁷.

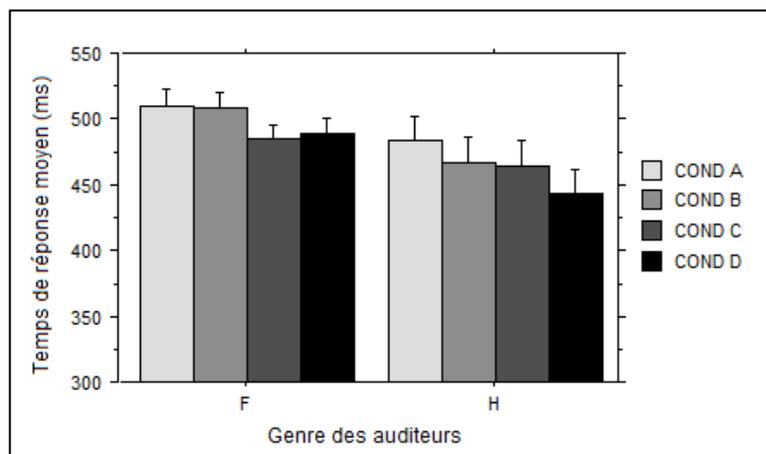


Figure 73 – Graphique des interactions représentant les temps de réponse moyens (exprimés en millisecondes) en fonction du genre des auditeurs (femme ou homme) et des conditions expérimentales (conditions A, B, C et D), avec les barres d'erreur.

²³⁶ *StatView* version 5.0.

²³⁷ La plupart des graphiques de cette section 3.2 où figurent des temps de réponse moyens seront présentés avec une limite inférieure fixée à 300 ms.

Les résultats de ce test montrent qu'il existe un effet global significatif du facteur « genre des auditeurs » sur les temps de réponse des auditeurs, avec $F(1,392)=9,277$ et $p<0,01$. Ainsi, les auditeurs masculins ont été significativement plus rapides que les auditrices pour reconnaître le mot-cible de manière globale, c'est-à-dire pour l'ensemble des conditions.

En revanche, et c'est là un élément important au regard des hypothèses, il n'existe aucune interaction entre les facteurs « genre des auditeurs » et « condition expérimentale » : $F(3,392)=0,299$ avec $p>0,80$. Cela signifie que *les différences relatives de temps de réponse entre les quatre conditions expérimentales (A, B, C, D) n'ont pas varié significativement en fonction du genre des auditeurs : l'analyse des temps de réponse en fonction des conditions expérimentales²³⁸ pourra donc être effectuée sur l'ensemble des auditeurs, indépendamment de leur genre*. De même, les tests statistiques présentés dans les sous-sections suivantes porteront sur les temps de réponse de l'ensemble des auditeurs (femmes et hommes confondus).

3.2.2 Différences de temps de réponse en fonction du locuteur produisant le mot-cible

Avant de tester les effets des différentes conditions expérimentales sur les temps de réponse, j'ai souhaité m'assurer qu'il n'existait pas un éventuel biais : si les temps de réponse varient en fonction du genre de la personne ayant produit le mot (H ou F), il est possible que cela provienne simplement du fait que ces voix sont des voix de personnes différentes (indépendamment de leur genre). Pour ce faire, j'ai effectué une ANOVA à un facteur, « locuteur ayant produit le mot-cible »²³⁹, sur les temps de réponse des auditeurs, toutes conditions confondues. Le graphique correspondant à ce test statistique est visible ci-après (figure 74).

²³⁸ Voir section 3.2.5.

²³⁹ Voir section 1.2 pour plus de détails sur les locuteurs.

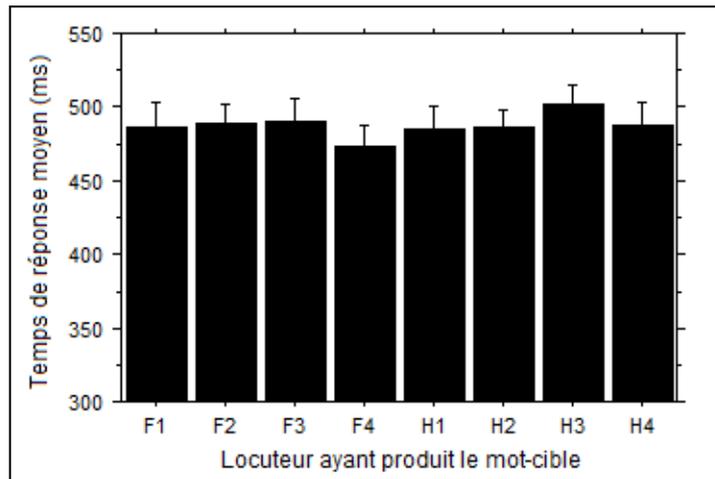


Figure 74 – Graphique représentant les temps de réponse moyens des auditeurs (exprimés en millisecondes) en fonction du locuteur ayant produit le mot-cible, avec les barres d'erreur.

Les résultats de ce test sont très clairs : il n'existe aucun effet significatif du facteur « locuteur ayant produit le mot-cible » sur les temps de réponse des auditeurs, avec $F(7,392)=0,288$ et $p>0,95$. En d'autres termes, *indépendamment du genre du locuteur, il n'y a pas de différences significatives de temps de réponse en fonction de la voix qui a produit le mot-cible.*

Un autre biais potentiel existe : le mot-cible étant de longueur sensiblement différente en fonction du locuteur qui l'a produit, il est possible que le temps de réponse des auditeurs soit corrélé positivement à la longueur des stimuli (*i.e.* plus le stimulus est court, plus il serait reconnu rapidement). Pour évaluer cette corrélation et sa significativité, j'ai effectué un test de Pearson sur la durée des stimuli et le temps de réponse moyen des auditeurs. Il en est ressorti une très faible corrélation positive : $r(8)=0,206$. Cette dernière est très largement non significative, avec $z=0,467$ et $p>0,6$. Le tableau 53, ci-après, illustre ce résultat avec la durée du mot-cible *étage* en fonction du locuteur l'ayant prononcé²⁴⁰, ainsi que le temps de réponse moyen des auditeurs pour chacun de ces items (toutes conditions confondues).

²⁴⁰ La durée de ces items a été mesurée à l'aide du logiciel *Praat 5.1.44*.

Tableau 53 – *Durée du mot-cible « étage » en fonction du locuteur l'ayant produit, et temps de réponse moyen des auditeurs pour chacun de ces huit items, toutes conditions confondues.*

Locuteur	Durée du mot-cible « étage » (ms)	Temps de réponse moyen des auditeurs (ms)
F1	510	487
F2	605	489
F3	607	490
F4	540	473
H1	475	485
H2	500	486
H3	542	502
H4	558	488

Au regard des résultats de ce test, il est donc possible d'affirmer que *la longueur du mot-cible étage, qui a varié légèrement en fonction du locuteur l'ayant produit, ne semble pas avoir joué sur les temps de réponse des auditeurs, toutes conditions expérimentales confondues.*

Un autre élément est susceptible d'avoir influencé le temps de réponse des auditeurs : la fréquence fondamentale moyenne, qui indépendamment du genre du locuteur, varie d'un individu à l'autre. Ryalls & Lieberman (1982) et Diehl & al (1996) ont mis en évidence qu'au dessus de 150 Hz, la difficulté d'identification des voyelles²⁴¹ augmentait avec la fréquence fondamentale moyenne. Cette variable pourrait donc avoir des conséquences sur le temps de traitement des mots. Pour vérifier cette hypothèse, j'ai conduit un test de Pearson sur le F0 moyen des mots-cibles et le temps de réponse moyen des auditeurs. Le résultat du test fait état d'une faible corrélation négative entre ces facteurs : $r(8)=-0,369$. Cette corrélation ne s'avère pas significative : $z=-0,866$ et $p>0,3$. La fréquence fondamentale moyenne des différentes occurrences du mot-cible *étage*²⁴² ainsi que le temps de réponse moyen des auditeurs correspondant (toutes conditions confondues), figurent dans le tableau 54, ci-après.

²⁴¹ Il s'agissait dans ces deux études de voyelles synthétisées (voir section 4.3.2 du chapitre 1).

²⁴² La fréquence fondamentale moyenne de ces items a été mesurée avec le logiciel *Praat 5.1.44*.

Tableau 54 – F0 moyen du mot-cible « étage » en fonction du locuteur l'ayant produit, et temps de réponse moyen des auditeurs pour chacun de ces huit items, toutes conditions confondues.

Locuteur	F0 moyen du mot-cible « étage » (Hz)	Temps de réponse moyen des auditeurs (ms)
F1	239	487
F2	187	489
F3	202	490
F4	226	473
H1	127	485
H2	135	486
H3	150	502
H4	143	488

Compte tenu des résultats de ce test de Pearson, il est possible de conclure que la fréquence fondamentale moyenne du mot-cible, sensiblement différente d'une voix à l'autre y compris pour un même genre, ne paraît pas avoir influencé les temps de réponse des auditeurs, toutes conditions expérimentales confondues.

3.2.3 Différences de temps de réponse en fonction du moment de présentation du mot-cible

Comme cela a été mentionné précédemment, l'utilisation d'un même mot-cible tout au long de l'expérience aurait pu entraîner une diminution progressive du temps de réponse des sujets. Un test de Spearman a été conduit sur les temps de réponse des sujets et le moment de présentation de chaque mot-cible expérimental (ordre chronologique). La figure 75, ci-après, illustre cette analyse.

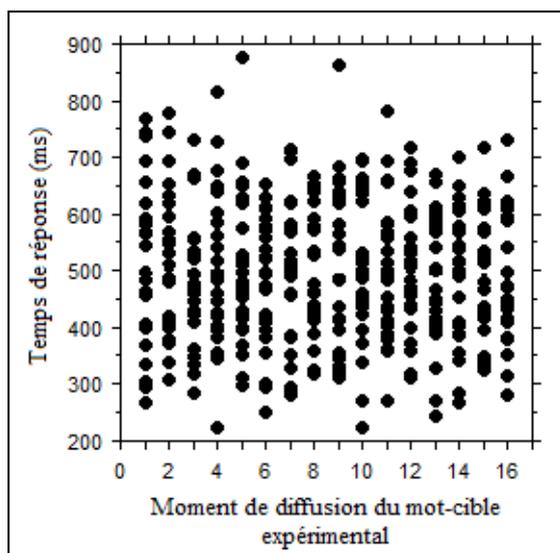


Figure 75 – Nuage de points illustrant l'absence de corrélation entre le temps de réponse des auditeurs et le moment de présentation (ordre chronologique) du mot-cible expérimental.

On observe une absence totale de corrélation entre ces deux variables, avec $\rho=0,001$ et $p>0,95$. *La répétition du mot-cible étage ne semble donc pas avoir entraîné de diminution progressive des temps de réponse des auditeurs.*

3.2.4 Le facteur « sujet »

Afin de choisir le type de test statistique à réaliser pour tester l'influence des conditions expérimentales, une analyse de l'effet du facteur « sujet » a également été effectuée.

Dans un premier temps, j'ai conduit une ANOVA à un facteur (« sujet »), sur les temps de réponse (toutes conditions et tous mots-cibles confondus). Le résultat est très net : il existe un important effet global significatif du facteur « sujet », avec $F(24, 375)=33,52$ et $p<0,0001$. Le graphique correspondant à cette analyse est visible ci-après (figure 76). Le code utilisé pour chaque participant correspond aux initiales, suivies de l'âge, puis du genre (F ou H)²⁴³.

²⁴³ Des informations complémentaires sur les participants sont disponibles en annexe 4.

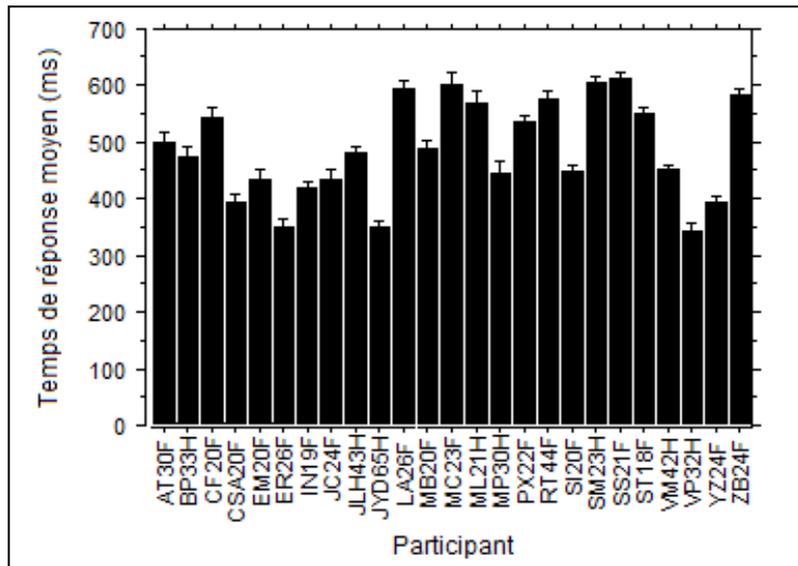


Figure 76 – Graphique représentant les temps de réponse moyens (exprimés en millisecondes) des 25 participants, toutes conditions et tous mots-cibles confondus, avec les barres d'erreur.

Le résultat de ce test signifie que *toutes conditions confondues les auditeurs ont eu des temps de réponse globalement différents, certains auditeurs étant, de manière générale, plus rapides que d'autres.*

De fait, le temps de réponse moyen pour chaque condition expérimentale aura donc une variance très élevée. Afin de limiter cette variance et son effet sur la potentielle significativité des différences inter-conditions, testées dans la section suivante, il serait possible de prendre en compte ce facteur « sujet » dans le calcul, en effectuant une ANOVA à deux facteurs, « sujet » et « conditions expérimentales » sur les temps de réponse des auditeurs.

Mais cela n'est envisageable que s'il n'existe aucune interaction entre ces deux facteurs, c'est-à-dire si les disparités inter-sujets sont simplement globales et que les différences relatives de temps de réponse entre les quatre conditions expérimentales ne varient pas en fonction de ce facteur « sujet ». Après vérification, il n'existe effectivement aucune interaction entre les facteurs « condition expérimentale » et « sujet » : $F(72,300)=0,647$ avec $p>0,95$. *Les différences relatives de temps de réponse entre les quatre conditions expérimentales (A, B, C, D) n'ont donc pas varié significativement en fonction du facteur « sujet ».*

3.2.5 Différences de temps de réponse en fonction des conditions expérimentales

Les possibles biais liés au genre des auditeurs, aux différentes voix, à l'ordre de présentation et au facteur « sujet » étant maintenant écartés, l'effet du facteur « condition expérimentale » sur le temps de réponse moyen des auditeurs a pu être testé. Pour ce faire, j'ai réalisé une ANOVA à deux facteurs, « condition expérimentale » et « sujet »²⁴⁴, sur les temps de réponse des auditeurs. Les temps de réponse des auditeurs en fonction de la condition expérimentale sont visibles sur la figure 77, ci-dessous.

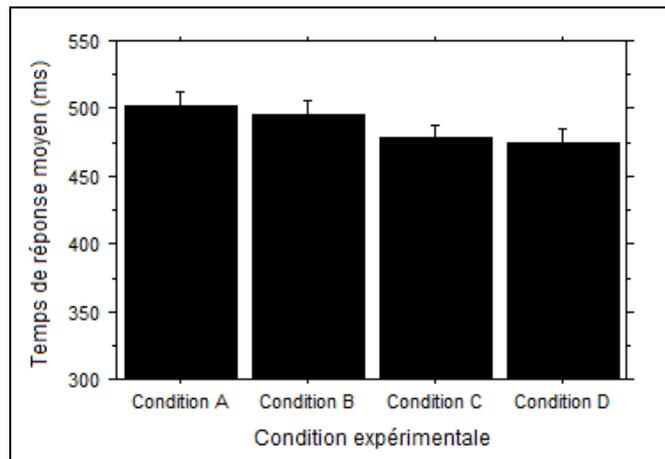


Figure 77 – Graphique représentant les temps de réponse moyens (exprimés en millisecondes) des auditeurs en fonction de la condition expérimentale, avec les barres d'erreur correspondantes.

Le résultat obtenu permet de conclure à l'existence d'un effet global très significatif du facteur « condition expérimentale » : $F(3,300)=4,597$ avec $p<0,01$. *Les temps de réponse des auditeurs ont donc varié significativement en fonction de la condition expérimentale.*

Afin de mieux comprendre ces variations de temps de réponse, j'ai souhaité observer les différences entre chacune des conditions expérimentales afin de savoir lesquelles sont significatives. Pour rappel, les quatre conditions expérimentales sont les suivantes :

²⁴⁴ Ce deuxième facteur a été inclus dans l'analyse afin d'obtenir une variance plus juste. Voir section 3.2.4 pour plus de détails.

- **Condition A** (*homogène*) : contexte *voix d'hommes* avant mot-cible expérimental *voix d'homme*.
- **Condition B** (*homogène*) : contexte *voix de femmes* avant mot-cible expérimental *voix de femme*.
- **Condition C** (*non-homogène*) : contexte *voix de femmes* avant mot-cible expérimental *voix d'homme*.
- **Condition D** (*non-homogène*) : contexte *voix d'hommes* avant mot-cible expérimental *voix de femme*.

Le test PLSD de Fisher montre que la différence est fortement significative entre les conditions A et C ($p < 0,01$), ainsi qu'entre les conditions B et D ($p < 0,02$). *Il existe donc bien un effet du contexte : les temps de réponse pour les voix de femmes, comme pour les voix d'hommes, ont été significativement plus courts dans les conditions « non-homogènes » (C et D), où les mots précédant le mot-cible sont produits par des voix du genre opposé, que dans les conditions « homogènes » (A et B).*

En revanche, aucune différence significative n'existe entre les temps de réponse moyens des conditions A et B ($p > 0,40$), ni entre ceux des conditions C et D ($p > 0,60$). *En contexte équivalent, les mots-cibles produits par des voix de femmes et ceux produits par des voix d'hommes ont donc entraîné des temps de réponse comparables, aucun type de voix n'ayant été traité significativement plus rapidement que l'autre.*

En plus de cette comparaison en contexte équivalent, j'ai souhaité vérifier si globalement, les temps de réponse moyens obtenus pour les voix de femmes (conditions B et D) et pour les voix d'hommes (conditions A et C) ne présentaient pas de différence significative. Pour cela, j'ai donc regroupé les temps de réponse des conditions B et D (voix de femmes), et ceux des conditions A et C (voix d'hommes), et effectué une ANOVA à deux facteurs, « type de voix produisant le mot-cible » et « sujet »²⁴⁵, sur les temps de réponse moyen des auditeurs. Le graphique représentant les temps de réponse moyens des auditeurs en fonction du type de voix produisant le mot-cible est visible ci-après (figure 78).

²⁴⁵ Facteur inclus ici encore dans le but d'obtenir une variance plus juste : voir section 3.2.3.

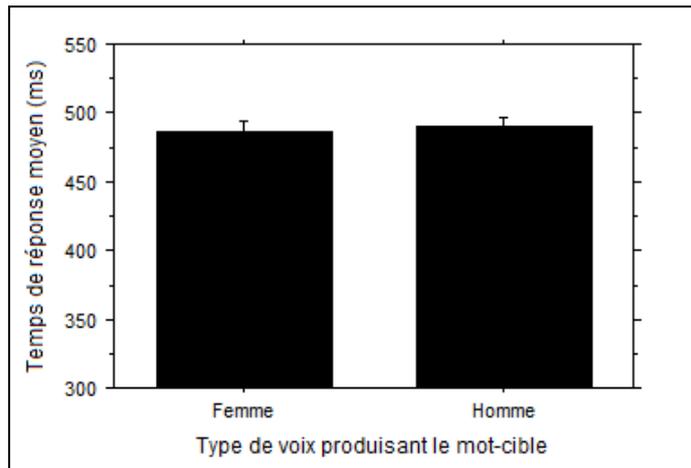


Figure 78 – Graphique représentant les temps de réponse moyens (exprimés en millisecondes) des auditeurs, en fonction du type de voix (femme ou homme) produisant le mot-cible, avec les barres d’erreur.

Les résultats de ce test montrent l’absence d’un effet significatif du facteur « type de voix produisant le mot-cible » sur les temps de réponse moyens des auditeurs : $F(1,350)=0,852$ avec $p>0,30$. *Il n’existe donc pas de différence significative entre les temps de réponse des auditeurs pour les mots-cibles produits par des femmes (conditions expérimentales B et D), et pour ceux produits par des hommes (conditions expérimentales A et C) : les auditeurs semblent bien avoir traité à vitesse équivalente les deux types de voix.*

4. DISCUSSION - CONCLUSION PARTIELLE

Cette expérience de détection de mots a permis d'obtenir un certain nombre de résultats intéressants. Tout d'abord, l'**hypothèse 7** a été confirmée : aussi bien en contexte équivalent que tous contextes confondus, les temps de traitement des mots produits par des voix d'hommes et de ceux produits par des voix de femmes ne présentent aucune différence significative. Ainsi, *les voix de femmes et les voix d'hommes semblent être traitées à la même vitesse par les auditeurs sur des mots isolés*. De plus, cette affirmation s'applique *quel que soit le genre de l'auditeur*. Par ailleurs, aucune corrélation significative n'est apparue entre la fréquence fondamentale moyenne du mot-cible²⁴⁶ et le temps de réponse moyen des auditeurs, ce qui suggère qu'*indépendamment du genre du locuteur, un mot produit avec une fréquence fondamentale élevée ne serait pas plus long à reconnaître par les auditeurs qu'un mot produit avec une fréquence fondamentale basse*.

Ces données sont à mettre en perspective avec des recherches antérieures. Ryalls & Lieberman (1982) et Diehl et al. (1996) avaient mis en évidence un lien entre fréquence fondamentale et difficulté d'identification des voyelles. Les voyelles seraient ainsi plus difficiles à identifier lorsqu'elles sont produites avec un F0 supérieur à 150 Hz : l'explication avancée par les auteurs étant le nombre restreint d'harmoniques lorsque la fréquence fondamentale est élevée. Cela pourrait suggérer que, plus généralement, les voix de femmes sont plus difficiles à traiter par les auditeurs. Ces éléments apparaissent donc contradictoires avec les résultats obtenus ici. Cependant, les expériences réalisées dans les deux études antérieures avaient pour but de mesurer le pourcentage d'erreur d'identification et non pas le temps de réponse. De plus, contrairement à la présente étude, les stimuli utilisés étaient de type synthétique : la qualité acoustique de la synthèse, en particulier pour les items ayant un F0 élevé, a donc pu jouer sur les réponses des auditeurs. Enfin, l'unité linguistique utilisée (voyelle isolée) peut sembler quelque peu artificielle : en dehors de conditions expérimentales, les auditeurs ont probablement rarement à identifier une unité si petite hors contexte.

²⁴⁶ La fréquence fondamentale moyenne des mots-cibles expérimentaux de cette étude variait de 127 à 239 Hz.

On peut donc parfaitement penser que les auditeurs, qui sont quotidiennement exposés à des voix à F0 élevé ainsi qu'à des voix à F0 bas, ont pu développer des capacités de traitement équivalentes pour ces différents types de voix, tout du moins pour un input d'une taille au moins équivalente à celle d'un mot. Ainsi, même si les voyelles produites avec un F0 élevé sont effectivement plus difficiles à identifier, ce qui reste à démontrer formellement, les auditeurs ont la possibilité de compenser avec les consonnes, dont on sait qu'elles sont particulièrement décisives pour l'accès au lexique (voir notamment Owren & Cardillo, 2006).

Strand (2000) a quant à elle utilisé des mots isolés et mesuré les temps de réponse en fonction du type de voix (homme / femme), comme je l'ai fait dans la présente étude. Néanmoins, le paradigme utilisé n'était pas le même : il s'agissait d'une expérience de répétition de mots, qui implique non seulement une tâche de perception mais également un travail de production. Malgré ces divergences méthodologiques, les résultats obtenus dans l'expérience de Strand sont finalement conformes à ceux obtenus ici : aucune différence significative de temps de réponse n'avait pu être observée entre les voix de femmes et les voix d'hommes. Notons que dans cette précédente étude, seule une voix d'homme et une voix de femme dites « stéréotypiques » avaient été utilisées, il était donc également nécessaire de confirmer ces tendances avec un plus grand nombre de voix, ce qui est désormais chose faite.

La deuxième observation importante concerne l'éventuel effet d'accommodation de l'auditeur au genre du locuteur, évoqué dans la conclusion du chapitre 1. Ce dernier a pu être mis en évidence grâce aux différences significatives de temps de traitement observées entre les conditions A et C d'une part et B et D d'autre part : les conditions *non-homogènes* (C et D) ont entraîné des temps de traitement inférieurs aux conditions *homogènes* (A et B). L'écoute d'un grand nombre de stimuli de type *voix d'homme* avant un mot-cible de type *voix de femme* (ou *vice versa*) a fait baisser le temps de réponse des auditeurs sur le mot-cible par rapport aux conditions dites *homogènes*, dans lesquelles le contexte et le mot-cible étaient de même type (uniquement *voix d'hommes* ou *voix de femmes*).

Ce phénomène s'explique très probablement par un regain d'attention du sujet, causé par un changement de paradigme : ce dernier ne peut être dû qu'au changement du genre du locuteur, étant donné que tous les autres paramètres (locuteur, mot, etc.) varient

de la même manière dans toutes les conditions expérimentales. Ces éléments semblent donc aller dans le sens de l'**hypothèse 6**, selon laquelle les voix de femmes et d'hommes sont considérées comme deux entités distinctes par le cerveau. D'autre part, ces résultats paraissent confirmer ceux obtenus précédemment par Sokhi et al. (2005) et Lattner et al. (2005) : grâce à l'utilisation de l'IRMf, ces auteurs ont pu mettre évidence que les voix de femmes et d'hommes activent de manière différente certaines zones du cerveau des auditeurs.

Les trois études expérimentales conduites dans le cadre de cette thèse, à savoir l'analyse acoustique, les expériences d'identification du genre par la voix et l'expérience de détection de mots ont maintenant été exposées. Les sections suivantes sont consacrées à la conclusion / discussion générale ainsi qu'aux perspectives de recherche et d'applications pratiques.

CONCLUSION – DISCUSSION GENERALE

L'état de la question dressé dans le premier chapitre a permis de dégager plusieurs points faisant l'objet de débats ou ayant été insuffisamment étudiés, dans le domaine des différences entre voix de femmes et voix d'hommes. Un certain nombre de ces éléments ont pu être éclaircis grâce à l'analyse acoustique et aux expériences conduites dans la présente étude.

Tout d'abord, les résultats de l'analyse acoustique multiparamétrique et inter-langues présentée dans le deuxième chapitre ont révélé des différences inter-genres significatives non seulement sur le plan de la fréquence fondamentale et des fréquences de résonance, mais également sur des paramètres tels que la qualité de voix (à travers la mesure H1-H2), le *VOT*, la durée des énoncés ou encore la plage de variation de F0. Si une partie de ces différences ont été observées aussi bien chez les francophones parisiens que chez les anglophones américains du nord-est des Etats-Unis, un certain nombre d'entre elles ne sont apparues que dans l'une des deux langues, ou à des degrés différents selon la langue. Ces importantes variations inter-langues, de même que la nature de certaines différences observées, mettent clairement en évidence la dimension construite des différences inter-genres, qui ne peuvent s'expliquer uniquement par des différences physiologiques et anatomiques.

Les expériences d'identification du genre par la voix décrites dans le troisième chapitre ont permis de mieux comprendre les stratégies utilisées par les auditeurs anglophones américains et francophones parisiens dans ce type de tâche. Il est notamment apparu d'importantes différences inter-langues : certains indices acoustiques tels que la fréquence fondamentale moyenne ont semblé jouer un rôle plus important chez les auditeurs anglophones, quand d'autres paramètres, en particulier la position des formants vocaliques, ont influencé plus fortement les auditeurs francophones. Ces données, qui concordent largement avec celles de l'analyse acoustique, permettent de conclure que les représentations mentales des voix de femmes et d'hommes chez les auditeurs varient sensiblement d'une langue à l'autre.

A partir de ces résultats, un schéma de synthèse a été établi pour tenter de rendre compte des différents processus entrant en jeu dans la construction des différences

acoustiques inter-genres et des représentations mentales des voix de femmes et d'hommes chez les auditeurs. Ce schéma est visible ci-dessous (figure 79).

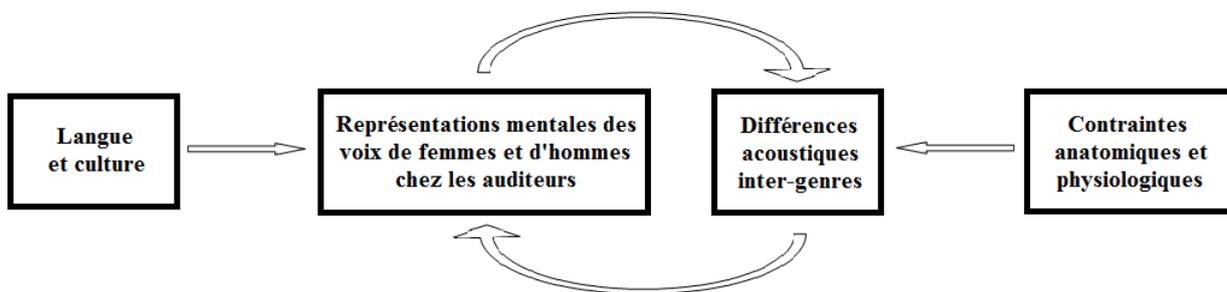


Figure 79 – Schéma représentant les interactions entre les différents facteurs jouant un rôle dans la construction des différences acoustiques inter-genres et de leurs représentations mentales chez les auditeurs.

A l'extrémité droite de ce schéma se trouvent les contraintes anatomiques et physiologiques, en particulier la taille du conduit vocal ainsi que la longueur et l'épaisseur des plis vocaux. Ces contraintes vont exercer une influence directe sur les différences acoustiques inter-genres. A l'autre extrémité du schéma entrent en jeu la langue et la culture qui, comme l'indiquent les variations inter-langues observées dans le chapitre 3, vont influencer sur les représentations mentales des voix de femmes et d'hommes chez les auditeurs. Ces représentations mentales, dépendantes de la langue et de la culture, vont naturellement jouer sur les différences acoustiques inter-genres : « un homme est censé parler de cette manière *donc* je vais parler de cette manière ». De même, les différences acoustiques inter-genres effectives, conditionnées en partie par les contraintes anatomiques et physiologiques, vont déterminer l'input acoustique reçu par les auditeurs et donc influencer également leurs représentations mentales.

On constate par ailleurs que les caractéristiques acoustiques des voix d'hommes, tels qu'un F0 bas, des fréquences de résonance basses ou même une vitesse d'élocution rapide, sont généralement associées à des postures d'autorité, de confiance en soi voire de domination, à l'inverse des traits vocaux typiques des voix de femmes (Apple et al., 1979 ; Fonagy, 1983 ; Ohala, 1984 ; Whiteside, 1995). Ainsi, les différences femmes-hommes sur le plan de la voix peuvent relever de contingences bien plus larges. Se pose notamment la question de la déconstruction des différences acoustiques femmes-hommes, dans une perspective de lutte féministe.

Le quatrième et dernier chapitre de ce travail de recherche avait pour objectif de mesurer les temps de traitement des voix de femmes et d'hommes, à travers une expérience de détection de mots. Il apparaît qu'en dépit des importantes différences acoustiques qu'ils présentent, ces deux types de voix sont traités à vitesse équivalente par l'auditeur, tout du moins au niveau lexical. Ces résultats sont à mettre en relation avec certaines interprétations erronées de l'étude de Sokhi et al. (2005), faites par la presse généraliste anglo-saxonne : « *Men who are accused of never listening by women now have an excuse : women's voices are more difficult for men to listen to than other men's* »²⁴⁷ ; « *Men's brains are not designed to listen to women's voices* »²⁴⁸. Ces considérations machistes ne présentaient bien entendu aucun fondement scientifique, puisque l'étude en question avait simplement mis en évidence que l'écoute des voix de femmes et d'hommes n'activait pas de la même manière certaines zones du cerveau chez les auditeurs masculins (l'expérience n'ayant pas été conduite sur des auditrices). Les résultats obtenus dans l'expérience de détection de mots, qui concordent notamment avec ceux de Strand (2000), permettent de contredire de manière formelle ces allégations.

Ce travail de recherche aura donc permis d'enrichir les connaissances scientifiques sur les différences entre voix de femmes et d'hommes, sur les conséquences psycholinguistiques de ces différences acoustiques, ainsi que sur le processus d'identification du genre par la voix. Il est en outre souhaitable que les résultats de cette étude contribuent, même modestement, au recul de certains préjugés sexistes.

²⁴⁷ <http://www.dailymail.co.uk/health/article-358320/What-dear.html>

²⁴⁸ <http://webcenters.netscape.compuserve.com/men/package.jsp?floc=DC-topb&name=fte/womenspeak/womenspeak>

PERSPECTIVES

Au regard des résultats obtenus dans la présente étude, de nombreuses perspectives se dessinent, tant au plan de la recherche fondamentale que des applications pratiques.

En premier lieu, l'analyse acoustique multiparamétrique présentée dans le deuxième chapitre pourrait être poursuivie sur un plus grand nombre de locuteurs et locutrices anglophones américains du nord-est et francophones parisiens, afin de neutraliser de manière plus sûre les variations interindividuelles. Dans un deuxième temps, elle pourrait être étendue aux productions de locuteurs d'autres langues et dialectes, en suivant très exactement la même méthode (corpus et procédure d'enregistrement similaires). Cela permettrait d'obtenir des données complémentaires sur les variations inter-langues des différences acoustiques femmes-hommes.

Il serait par ailleurs intéressant d'étudier les comportements de locuteurs et locutrices bilingues ou multilingues. On sait notamment que les locutrices bilingues anglais-russe tendent à augmenter leur fréquence fondamentale moyenne lorsqu'elles emploient la langue russe (Altenberg & Ferrand, 2006). Qu'en est-il, par exemple, des sujets bilingues français-anglais américain ? Les hommes utilisent-ils un type de phonation plus *creaky* en anglais ? Les femmes élargissent-elles leur plage de variation en français ? Les différences inter-genres sur le F1 sont-elles plus importantes quand ces locuteurs bilingues utilisent la langue anglaise ? Une analyse acoustique multiparamétrique permettrait là encore de répondre à ces questions.

D'un point de vue perceptif, l'ajustement des frontières phonémiques en fonction du genre du locuteur²⁴⁹ n'a pour l'instant été mis en évidence que chez les locuteurs anglophones américains. Il a été démontré que ces derniers ajustent à la hausse la frontière entre /ʊ/ et /ʌ/ sur le plan du F1 lorsqu'un mot est produit par une femme (Johnson et al., 1999). De même, la frontière perceptive entre les fricatives /s/ et /ʃ/ au niveau de la zone de bruit est elle aussi rehaussée lors de l'écoute de voix de femmes (Strand & Johnson, 1996). Des expériences du même type pourraient être réalisées sur

²⁴⁹ Voir section 4.3.1 du chapitre 1.

des auditeurs francophones, en testant par exemple les frontières perceptives de voyelles se différenciant uniquement par la position des deuxième et troisième formants. A l'instar des études citées précédemment, ces items seraient placés au sein de mots produits par des locutrices et des locuteurs, puis un continuum acoustique serait ensuite créé par resynthèse (e.g. [pɛʁ] => [pœʁ]).

Concernant les applications pratiques, les différences acoustiques femmes-hommes pourraient être exploitées dans le domaine de la didactique des langues. Il serait souhaitable de sensibiliser les apprenants aux différences inter-genres existant dans la langue d'apprentissage. Cette sensibilisation pourrait se traduire par plusieurs mesures concrètes. Tout d'abord, il conviendrait de présenter aux apprenants un input *diversifié*, c'est-à-dire composé de voix de femmes et d'hommes. Selon Kahn (1974), le type d'input (en termes de genre du locuteur) conditionnerait en effet les productions des apprenants : des apprenants de l'arabe, femmes et hommes, ayant été exposés uniquement à la voix de leur professeur de genre masculin, auraient tous calqué leur prononciation sur la sienne, notamment pour ce qui est de la fréquence des formants vocaliques. Une autre application concrète se rapporte aux logiciels tels que SaRP²⁵⁰, qui permettent aux apprenants de comparer la fréquence de leurs formants avec des valeurs de référence présentées dans un triangle vocalique. Compte tenu des larges différences inter-genres existant sur ce paramètre, il conviendrait de proposer pour chaque langue des triangles vocaliques de référence distincts pour les voix femmes et les voix d'hommes.

Les données obtenues dans le présent travail pourraient d'autre part s'avérer particulièrement utiles pour les personnes transgenres. La voix étant un élément constitutif de l'identité de genre, la bonne connaissance des paramètres acoustiques à travailler permettrait d'aider ces individus à être effectivement perçus comme *homme* ou *femme* (Wiltshire, 1995). Au-delà de la fréquence fondamentale moyenne, adaptée souvent spontanément par les personnes transgenres (voir Spencer, 1988), d'autres éléments tels que la fréquence des formants vocaliques, la qualité de voix, la vitesse d'élocution ou encore la plage de variation de F0 pourraient être travaillés. La liste des paramètres acoustiques pertinents devrait idéalement être établie pour chaque langue, en fonction du rôle effectif joué par ces derniers dans l'identification du genre par la voix :

²⁵⁰ Voir Nikolov et al. (2007, 2011) et Ditchева et al. (2009).

concernant l'anglais américain et le français parisien, le tableau proposé en conclusion du chapitre 3 pourrait servir de base. Les transgenres seraient alors formés à l'utilisation de logiciels tels que *Praat*, *Speech Analyzer* et *SaRP*, afin de pouvoir travailler leur voix de manière autonome, ou seraient à défaut accompagnés par des personnes qualifiées (phonéticiens, orthophonistes ou phoniâtres).

De la même manière, les résultats de la présente étude pourraient être exploités dans le champ de la phonétique clinique. La rééducation de patients atteints de certaines pathologies vocales (lésions au niveau des plis vocaux, paralysie récurrentielle, etc.) pourrait ainsi être optimisée, en travaillant divers paramètres acoustiques leur permettant de retrouver une voix plus clairement identifiable comme étant celle d'un homme ou d'une femme.

Enfin, les données obtenues ici sont susceptibles de fournir des indications intéressantes aux chercheurs et ingénieurs travaillant dans le domaine du traitement automatique de la parole. Ces derniers pourraient tirer parti à la fois des résultats de l'analyse acoustique et de ceux des expériences d'identification du genre par la voix. Leur prise en compte permettrait, d'une part, de mieux adapter au genre du locuteur les modèles utilisés pour la reconnaissance vocale et, d'autre part, d'améliorer les techniques d'identification automatique du genre du locuteur.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ABITBOL, J., ABITBOL, P. & ABITBOL, B. (1999). Sex hormones and the female voice. *Journal of Voice*, 13, 424-446.
- ALTENBERG, E. P. & FERRAND, C. T. (2006). Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice*, 20, 89-96.
- AMIR, O. & BIRON-SHENTAL, T. (2004). The impact of hormonal fluctuations on female vocal folds. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 12, 180-184.
- ANDRE, C., GHIO, A., CAVE, C. & TESTON, B. (2003). Perceval: a computer-driven system for experimentation on auditory and visual perception. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelone, 1421-1424.
- ANDREWS, M. & SCHMIDT, C. (1997). Gender presentation: perceptual and acoustical analyses of voice. *Journal of Voice*, 11, 307-313.
- ANTAL, A., VARGA, E., NITSCHKE, M., CHADAIDE, Z., PAULUS, W., KOVACS, G & VIDNYANSZKY, Z. (2004). Direct current stimulation over MT+/V5 modulates motion aftereffect in humans. *Neuroreport*, 15, 2491-2494.
- APPLE, W., STREETER, L. & KRAUSS, R. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37, 715-727.
- ARNOLD, A. (2008). *La prosodie comme espace d'émergence du genre* (119 p.). Mémoire de Master 2 sous la direction de Lucas Greco et Jacqueline Vaissière. Université Paris III.
- ARNOLD, A. (2012). Le rôle de la fréquence fondamentale et des fréquences de résonance dans la perception du genre. *TIPA - Travaux Interdisciplinaires sur la Parole et le Langage*, 28, 1-18.
- ARONOVITCH, C. D. (1976). The voice of personality: stereotyped judgments and their relation to voice quality and sex of speaker. *Journal of Social Psychology*, 99, 207-220.
- AUERBACH, O., HAMMOND, E. & GARFINKEL, L. (1970). Histologic changes in the larynx in relation to smoking habits. *Cancer*, 25, 92-104.
- BACHOROWSKI, J. A. & OWREN, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America*, 106, 1054-1063.

- BARLOW, M. & WAGNER, M. (2000). Perceptions of identity, gender and idiolect in prosodically altered speech using a composite model approach. In *Proceedings of the 8th Australian International Conference on Speech Science and Technology*, Canberra, 330-335.
- BARNEY, A., DE STEFANO, A. & HENRICH, N. (2007). The effect of glottal opening on the acoustic response of the vocal tract. *Acta Acustica united with Acustica*, 93, 1046-1056.
- BEDARD, C. & BELIN, P. (2004). A “voice inversion effect?”. *Brain and Cognition*, 55, 247-249.
- BEAUVOIR, S. (1949). *Le deuxième sexe*. Paris : Gallimard.
- BELLANDESE, M. H. (2009). Fundamental frequency and gender identification in standard esophageal and tracheoesophageal speakers. *Journal of Communication Disorders*, 42, 89-99.
- BENNETT, S. & MONTERO-DIAZ, L. (1982). Children's perception of speaker sex. *Journal of Phonetics*, 10, 113-121.
- BIEMANS, M. (2000). *Gender variation in voice quality*. Utrecht, Pays-Bas : LOT.
- BLADON, R. A., HENTON, C. G. & PICKERING, J. B. (1984). Towards an auditory theory of speaker normalization. *Language Communication*, 4, 59-69.
- BLESS, D. M. & ABBS, J. (1983). *Vocal fold physiology: Contemporary research and clinical issues*. San Diego, Etats-Unis : College-Hill Press.
- BOË, L.-J., CONTINI, M. & RAKOTOFIRINGA, H. (1975). Étude statistique de la fréquence laryngienne. *Phonetica*, 32, 1-23.
- BREND, R. M. (1971). Male-female intonation patterns in American English. In *Proceedings of the 7th International Congress of Phonetic Sciences*, La Haye, 866-870.
- BROWN, W. S., MORRIS, R., HICKS, D. & HOWELL, E. (1993). Phonational profiles of female professional singers and nonsingers. *Journal of Voice*, 7, 219-226.
- BUSBY, P. A. & PLANT, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America*, 97, 2603-2607.
- BUTLER, J. (1990). *Gender Trouble*. New-York, Etats-Unis : Routledge.
- BUTLER, J. (1993). *Bodies that Matter*. New-York, Etats-Unis : Routledge

- BYRD, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, 92, 593-596.
- BYRD, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39-54.
- CALLIOPE (1989). *La parole et son traitement automatique*. Paris : Masson.
- CARAMAZZA, A., YENI-KOMSHIAN, G. H., ZURIF, E. B. & CARBONE, E. (1973). The acquisition of a new phonological contrast: the case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 52, 421-428.
- CHEN, G-T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics*, 2, 159-171.
- CHERNG, C-H., WONG, C-S., HSU, C-H. & HO, S-T. (2002). Airway length in adults: estimation of the optimal endotracheal tube length for orotracheal intubation. *Journal of Clinical Anesthesia*, 14, 271-274.
- CHEVRIE-MULLER, C., SALOMON, D. & FERREY, G. (1971). Contribution à l'établissement de quelques constantes physiologiques de la voix parlée de la femme adolescente, adulte et âgée. *Journal Français d'Oto-Rhino-Laryngologie*, 16, 433-455.
- CHIBA, T. & KAJIYAMA, M. (1941). *The Vowel – Its Nature and Structure*. Tokyo, Japon : Tokyo-Kaiseikan.
- CHILDERS, D. G. & WONG, C. F. (1994). Measuring and modeling vocal source-tract interaction. *IEEE Transactions on Biomedical Engineering*, 41, 663-671.
- CHILDERS, D. G. & WU, K. (1991). Gender recognition from speech: Part II: Fine analysis. *Journal of the Acoustical Society of America*, 90, 1841-56.
- CLEVELAND, T. (1976). The acoustic properties of voice timbre types and the importance of these properties in the determination of voice classification in male singers. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 17, 17-29.
- CLEVELAND, T. (1977). Acoustic properties of voice timbre types and their influence on voice classification. *Journal of the Acoustical Society of America*, 61, 1622-1629.
- CLOPPER, C. G. (2004). *Linguistic experience and the perceptual classification of dialect variation* (228 p.). Thèse de doctorat sous la direction de David Pisoni. Indiana University.
- CLOPPER, C. G., PISONI, D. B., DE JONG, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, 118, 1661-1676.

- CLOPPER, C. G. & SMILJANIC, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, 39, 237-245.
- COADOU, M. (2007). *Qualité de voix et accents régionaux en anglais britannique* (328 p.). Thèse de doctorat sous la direction de Daniel Hirst. Université d'Aix-Marseille I.
- COLEMAN, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research*, 14, 565-577.
- COLEMAN, R. O. (1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research*, 19, 168-180.
- COLEMAN, R. O. (1983). Acoustic correlates of speaker sex identification: implications for the transsexual voice. *The Journal of Sex Research*, 19, 293-295
- DALY, N. & WARREN, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, 5, 85-96.
- DE BOER, B. (2010). Investigating the acoustic effect of the descended larynx with articulatory models. *Journal of Phonetics*, 38, 679-686.
- DEAUX, K & LAFRANCE, M. (1998). Gender. In Gilbert, D. et al. (dir.), *The Handbook of Social Psychology*. New-York, Etats-Unis : Oxford University Press, 788-827.
- DEHAENE-LAMBERTZ, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience*, 12, 449-460.
- DIEHL, R. L., LINDBLOM, B., HOEMEKE, K. A. & FAHEY, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, 24, 187-208.
- DITCHEVA, M. & DOMMERGUES, J.-Y. (2008). Modèles acoustiques des voyelles du français et de l'anglais – voix de femme. *Plovdiv University "Paissii Hilendarski" – Bulgaria, Scientific Works – Philology*, 46, 203-210.
- DITCHEVA, M., DOMMERGUES, J.-Y. & RYST, E. (2009). Modélisation et acquisition des voyelles à l'aide du logiciel SaRP. *Plovdiv University "Paissii Hilendarski" – Bulgaria, Scientific Works – Philology*, 47, 88-95.
- DOLSON, M. (1994). The Pitch of Speech as a Function of Linguistic Community. *Music Perception*, 11, 321-331.
- ECKERT, P. (1989). The whole woman: sex and gender differences in variation. *Language Variation and Change*, 1, 245-267.

- EDELSKY, C. (1979). Question intonation and sex roles. *Language in Society*, 8, 15-32.
- ELERT, C. C. & HAMMARBERG, B. (1991). Regional voice variation in Sweden. In *Proceedings of the 12th International Congress of Phonetic Sciences*, Université de Provence, 418-420.
- ELYAN, O. (1978). Sex differences in speech style. *Women Speaking*, 4, 4-8.
- ERICSDOTTER, C. (2005). *Articulatory-acoustic relationships in Swedish vowel sounds*. Thèse de doctorat sous la direction de Olle Engstrand et Björn Lindblom. Stockholm University.
- ERICSDOTTER, C. & ERICSSON, A. M. (2001). Gender differences in vowel duration in read Swedish: Preliminary results. *Working Papers of the Department of Linguistics, Lund University*, 49, 34-37.
- FANT, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scaling. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 7, 22-30.
- FANT, G. (1970). *Acoustic Theory of Speech Production*, (2e éd.), La Haye, Pays-Bas : Mouton.
- FANT, G. (1975). Non-uniform vowel normalization. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 16, 1-19.
- FANT, G., ISHIZAKA, K., LINDQVIST, J. & SUNDBERG, J. (1972). Subglottal Formants, *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 13, 1-12.
- FAUTH C. (2012). Paralysies récurrentielles et perturbation de l'intelligibilité de la parole et de la classification homme / femme. In *Actes du Colloque International Jeunes Chercheurs 2012*, Strasbourg, (à paraître).
- FERRAGNE, E. & PELLEGRINO, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, 40, 1-34.
- FITCH, W. T., & GIEDD, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106, 1511-1522.
- FITZSIMONS, M., SHEAHAN, N. & STAUNTON, H. (2001). Gender and the integration of acoustic dimensions of prosody: implications for clinical studies. *Brain and Language*, 78, 94-108.
- FONAGY, I. (1983). *La vive voix : essais de psycho-phonétique*. Paris : Payot.
- FOX, R. A. & NISSEN, S. L. (2005). Sex-related acoustic changes in voiceless English fricatives. *Journal of Speech, Language, and Hearing Research*, 48, 753-765.

- FRITZELL, B., HAMMARBERG, B., GAUFFIN, J., KARLSSON, I. & SUNDBERG, J. (1986). Breathiness and insufficient vocal fold closure. *Journal of Phonetics*, 14, 549-553.
- FU, Q-J., CHINCHILLA, S. & GALVIN, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 5, 253-260.
- GENDROT, C. & ADDA-DECKER, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Proceedings of Eurospeech – Lisbon (Portugal)*, 2453-2456.
- GHIO, A., ANDRÉ, C, TESTON, B. & CAVÉ, C. (2003). Perceval : une station automatisée de tests de perception et d'évaluation auditive et visuelle. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence*, 22, 115-133.
- GILBERT, H. R. & WEISMER, G. G. (1974). The effects of smoking on the speaking fundamental frequency of adult women. *Journal of Psycholinguistic Research*, 3, 225-231.
- GOLDSTEIN, U. (1980). *An articulatory model for the vocal tracts of growing children*. Thèse de doctorat sous la direction de Kenneth N. Stevens. Cambridge : Massachusetts Institute of Technology.
- GONZALEZ, J. (2004). Formant frequencies and body size of speaker: a weak relationship in adult humans. *Journal of Phonetics*, 32, 277–287.
- GORDON, M. (2001). Linguistic aspects of voice quality with special reference to Athabaskan. In *Proceedings of the 2001 Athabaskan Languages Conference*, 163-178.
- GORDON, M. & LADEFOGED, P. (2001). Phonation types: A crosslinguistic overview. *Journal of Phonetics*, 29, 383-406.
- GOTTFRIED, T. L. & CHEW, S. L. (1986). Intelligibility of vowels sung by a countertenor. *Journal of the Acoustical Society of America*, 79, 124-130.
- GRADDOL, D. (1986). Discourse specific pitch behaviour. In Johns-Lewis, C. (dir.), *Intonation in discourse*. Londres, Royaume-Uni : Croom Helm.
- GRAY, H. (1918). *Anatomy of the Human Body*. Philadelphie, Etats-Unis : Lea & Febiger.
- GREISBACH, R. (1999). Estimation of speaker height from formant frequencies. *Forensic Linguistics*, 6, 265-277.
- GROSJEAN, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267–283.

- GUGATSCHKA, M., KIESLER, K., OBERMAYER-PIETSCH, B., SCHOEKLER, B., SCHMID, C., GROSELJ-STRELE, A. & FRIEDRICH, G. (2010). Sex hormones and the elderly male voice. *Journal of Voice*, 24, 369-373.
- HAAN, J. (2002). *Speaking of questions. An exploration of Dutch question intonation*. Utrecht, Pays-Bas : LOT.
- HAAN, J. & VAN HEUVEN, V. J. (1999). Male vs. female pitch range in Dutch questions. In *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1581-1584.
- HAMMARBERG, B., FRITZELL, B., GAUFFIN, J. & SUNDBERG, J. (1986). Acoustic and perceptual analysis of vocal dysfunction. *Journal of Phonetics*, 14, 533-547.
- HANSON, H. & CHUANG, E. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, 106, 1064-77.
- HARAWAY, D. (1991). *Simians, Cyborgs, and Women: the Reinvention of Nature*. New-York, Etats-Unis : Routledge.
- HARNSBERGER, J. D., SHRIVASTAV, R., BROWN, W. S., ROTHMAN, H. & HOLLIEN, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice*, 22, 58-69.
- HEFFERNAN, K. (2004). Evidence from HNR that /s/ is a social marker of gender. *Toronto Working Papers in Linguistics*, 23, 71-84.
- HENRICH, N., DOVAL, B. & D'ALESSANDRO, C. (1999). Glottal open quotient estimation using linear prediction. In *Proceedings of the International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, 12-17.
- HENTON, C. G. (1985). *A comparative study of phonetic sex-specific differences across languages* (387 p.). Thèse de doctorat. University of Oxford.
- HENTON, C. G. (1989a). Fact and fiction in the description of female and male pitch. *Language and Communication*, 9, 299-311.
- HENTON, C. G. (1989b). Sociophonetic aspects of creaky voice. *Journal of the Acoustical Society of America*, 86, S26.
- HENTON, C. G. (1992). Sex and speech synthesis: techniques, successes, and challenges. In *Proceedings of the Fourth Australian International Conference on Speech Science and Technology*, Brisbane, 738-743.

- 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*, Stockholm, 420-423.
- HENTON, C. G. & BLADON, R. A. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication*, 5, 221-227.
- HENTON, C. G. & BLADON, R. A. (1986). Creak as a sociophonetic marker. In Hyman, L. & Li, C. N. *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin*, Londres : Routledge, 3-29.
- HERBST, L. (1964). *Untersuchungen zur Indifferenzlage der Sprechstimme. Studien zur Problematik des physiologischen Hauptsprechtonbereichs*. Thèse de doctorat. Université de Halle-Wittenberg.
- HERBST, L. (1969). Die Umfänge der physiologischen Hauptsprechtonbereiche von Frauen und Männern. *Zeitschrift für Phonetik und Sprachliche Kommunikation*, 22, 426-438.
- HILLENBRAND, J., GETTY, L. A., CLARK, M. J. & WHEELER, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099-3111.
- HOLLIEN, H. (1960). Vocal pitch variation related to changes in vocal fold length. *Journal of Speech and Hearing Research*, 3, 150-156.
- HOLLIEN, H. & SHIP, T. (1972). Speaking fundamental frequency and chronological age in males. *Journal of Speech and Hearing Research*, 15, 155-159.
- HOWIE, J. & DELATTRE, P. (1962). An experimental study of the effect of pitch on the intelligibility of vowels. *The National Association of Teachers of Singing Bulletin*, 18, 6-9.
- HUBER, D. (1989). Voice characteristics of female speech and their representation in computer speech synthesis and recognition. In *Proceedings of Eurospeech '89*, Paris, 477-480.
- HWA CHEN, S. (2007). Sex differences in frequency and intensity in reading and voice range profiles for Taiwanese adult speakers. *Folia Phoniatica et Logopaedica*, 59, 1-9.
- INGEMANN, F. (1968). Identification of the speaker's sex from voiceless fricatives. *Journal of the Acoustical Society of America*, 44, 1142-1144.
- JACEWICZ, E., FOX, R. A., O'NEILL, C. & SALMONS, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21, 233-256.

- JACEWICZ, E., FOX, R. A. & SALMONS, J. (2007). Vowel duration in three American English dialects. *American Speech*, 82, 367-385.
- JEANNON, J. P. (1998). Vocim analysis of laryngeal images: is breathiness related to the glottic area? *Clinical Otolaryngology & Allied Sciences*, 23, 351-353.
- JOHANSSON, C., SUNDBERG, J. & WILBRAND, H. (1982). X-ray study of articulation and formant frequencies in two female singers. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 23, 117-134.
- JOHNSON, K. (2005). Speaker normalization in speech perception. In Pisoni, D. & Remez, R. (dir.), *The Handbook of Speech Perception*. Oxford, Royaume-Uni : Blackwell Publishers, 363-389.
- JOHNSON, K. (2006). Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. *Journal of Phonetics*, 34, 485-499.
- JOHNSON, K. & MARTIN, J. (2001). Acoustic vowel reduction in Creek: effects of distinctive length and position in the word. *Phonetica*, 58, 81-102.
- JOHNSON, K., STRAND, E. & D'IMPERIO, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27, 359-384.
- JONGMAN, A., WAYLAND, R., WONG, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108, 1252-1263.
- KAHANE, J. (1978). A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy*, 151, 11-20.
- KAHN, M. (1974). Nonphysiological variation in the pronunciation of Arabic by men and women. *Journal of the Acoustical Society of America*, 55, S43.
- KAMIYAMA, T. (2009). *Apprentissage phonétique des voyelles du français langue étrangère chez des apprenants japonophones* (335 p.). Thèse de doctorat sous la direction de Jacqueline Vaissière. Université Paris III.
- KARLSSON, F., ZETTERHOLM, E. & SULLIVAN, K. (2004). Development of a gender difference in voice onset time. In *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, Sydney, 316-321.
- KARLSSON, I. (1991). Female voices in speech synthesis. *Journal of Phonetics*, 19, 111-120.
- KARLSSON, I. (1992). Evaluations of acoustic differences between male and female voices; a pilot study. *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 1, 19-31.

- KITZING, P. (1979). *Glottografisk frekvensindikering: en undersökningsmetod för mätning av röstläge och röstomfång samt framställning av röstfrekvensdistributionen*. Lund University, Malmö.
- KLATT, D. H. (1986). Detailed spectral analysis of a female voice. *Journal of the Acoustical Society of America*, 80, S97.
- KLATT, D. H. (1987a). Acoustic correlates of breathiness: first harmonic amplitude, turbulence noise, and tracheal coupling. *Journal of the Acoustical Society of America*, 82, S91.
- KLATT, D. H. (1987b). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, 82, 737-793.
- KLATT, D. H. & KLATT, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.
- KOENIG, L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language and Hearing Research*, 43, 1211-1228.
- KRAMER, C. (1977). Perceptions of female and male speech. *Language and Speech*, 20, 151-161.
- LABOV, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205-254.
- LABOV, W., ASH, S. & BOBERG, C. (2006). *The Atlas of North American English: phonetics, phonology and sound change*. Berlin, Allemagne : Mouton de Gruyter.
- LADEFOGED, P. (1967). Stress and respiratory activity. In Ladefoged, P., *Three Areas of Experimental Phonetics*. Londres, GB : Oxford University Press, 1-49.
- LADEFOGED, P. (1971). *Preliminaries to linguistic phonetics*. Chicago, Etats-Unis : University of Chicago Press.
- LAQUEUR, T. (1990). *Making sex: body and gender from the Greeks to Freud*. Cambridge, Etats-Unis : Harvard University Press.
- LASS, N. J., HUGHES, K. R., BOWYER, M. D., WATERS, L. T. & BOURNE, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America*, 59, 675-678.
- LATTNER, S., MEYER, M. E. & FRIEDERICI, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, 24, 11-20.

- LAVIER, J. (1980). *The phonetic description of voice quality*. Cambridge, Royaume-Uni : Cambridge University Press.
- LE BRETON, D. (2011). *Eclats de voix : une anthropologie des voix*. Paris : Editions Métailié.
- LEE, S., POTAMIANOS, A. & NARAYANAN, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105, 1455-1468.
- LIN, C. & WANG, H. (2011). Automatic estimation of voice onset time for word-initial stops by applying random forest to onset detection, *Journal of the Acoustical Society of America*, 130, 514-525.
- LIU, T. C., IRISH, J. C., ADAMS, S. G., DURKIN, L. C. & HUNT, E. J. (1996). Prospective study of patients' subjective responses to botulinum toxin injection for spasmodic dysphonia. *Journal of Otolaryngology*, 25, 66-74.
- LÖWY, I. & ROUCH, H. (2003). Genèse et développement du genre : les sciences et les origines de la distinction entre sexe et genre. *Cahiers du Genre*, 34, 5-16.
- MACDONALD, J., & MCGURK, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, 24, 253-257.
- MARSLEN-WILSON, W. & TYLER L. K (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.
- MARTIN, Ph. (2008). *Phonétique acoustique, introduction à l'analyse acoustique de la parole*. Paris : Armand Colin.
- MARTIN, Pi. (1995). The opposition between /ε/ and /ε:/ in current Quebec French. *La Linguistique*, 31, 33-45.
- MARTIN, Pi. (1998a). Dynamique vocalique en français du Québec. *La Linguistique*, 34, 67-76.
- MARTIN, Pi. (1998b). Les voyelles d'aperture moyenne en français du Québec. *Cahiers de l'ILSL*, 11, 215-242.
- MARTIN, Pi. (2001). Les voyelles nasales en français du Québec. *La Linguistique*, 37, 49-70.
- MATTINGLY, I. G. (1966). Speaker variation and vocal-tract size. *Journal of the Acoustical Society of America*, 39, 1219.

- MAURER, D., COOK, N., LANDIS, T. & D'HEUREUSE, C. (1992). Are measured differences between the formants of men, women and children due to f₀ differences? *Journal of the International Phonetic Association*, 21, 66-79.
- MCCONNELL-GINET, S. (1978). Intonation in a man's world. *Signs*, 3, 541-559.
- MCGOWAN, R. & NITTROUER, S. (1988). Differences in fricative production between children and adults: Evidence from an acoustic analysis of /ʃ/ and /s/. *Journal of the Acoustical Society of America*, 83, 229-236.
- MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- MENDOZA, E., VALENCIA, N., MUÑOZ, J. & TRUJILLO, H. (1996). Differences in voice quality between men and women: use of the Long-Term Average Spectrum (LTAS). *Journal of Voice*, 10, 59-66.
- MILLER, J. D., LEE, S., UCHANSKI, R. M., HEIDBREDE, A. F., RICHMAN, B. B. & TADLOCK, J. (1996). Creation of two children's speech databases. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, 849-852.
- MILLER, J. L. (1994). On the internal structure of phonetic categories: a progress report. *Cognition*, 50, 271-285.
- MOORE, B. C. & GLASBERG, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74, 750-753.
- MORRIS, R. J., BROWN, W. S., HICKS, D. M. & HOWELL, E. (1995). Phonational profiles of male trained singers and nonsingers. *Journal of Voice*, 9, 142-148.
- MORRIS, R. J., MCCREA, C. R., HERRING, K. D. (2008). Voice onset time differences between adult males and females: isolated syllables. *Journal of Phonetics*, 36, 308-317.
- MULLENNIX, J. W., JOHNSON, K. A., TOPCU-DURGUN, M. & FARNSWORTH, L. M. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America*, 98, 3080-3095.
- MUNSON, B., JEFFERSON, S. V. & MCDONALD, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *Journal of the Acoustical Society of America*, 119, 2427-2437.

- NAGLE, K. F., EADIE, T. L., WRIGHT, D. R. & SUMIDA, Y. A. (2012). Effect of fundamental frequency on judgments of electrolaryngeal speech. *American Journal of Speech-Language Pathology*, 21, 154-166.
- NGUYEN, N. (2005). La perception de la parole. In Nguyen, N., Wauquier-Gravelines, S. & Durand, J. (dir.), *Phonologie et phonétique: Forme et substance*. Paris : Hermès, 425-447.
- NIKOLOV, R., DOMMERGUES, J-Y. & RYST, E. (2007). SaRP : outil polyvalent de représentations multi-points et multi-séries des formants vocaliques. *Plovdiv University "Paissii Hilendarski" – Bulgaria, Scientific Works – Philology*, 45, 88-95.
- NISHIZAWA, N., SAWASHIMA, M. & YONEMOTO, K. (1988). Vocal fold length in vocal pitch change. In Fujimura, O., *Vocal Physiology: Voice Production, Mechanisms, and Function*. New York : Raven, 75-82.
- NITTROUER, S. (1995). Children learn separate aspects of speech production at different rates: Evidence from spectral moments. *Journal of the Acoustical Society of America*, 97, 520-530.
- NORDSTROM, P-E. (1977). Female and infant vocal tracts simulated from male area functions. *Journal of Phonetics*, 5, 81-92.
- NOYES, J. M. & FRANKISH, C. R. (1989). A review of speech recognition: applications in the office. *Behaviour and Information Technology*, 8, 475-486.
- OAKLEY, A. (1972). *Sex, Gender and Society*. Londres, Royaume-Uni : Temple Smith.
- OH, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39, 59-67.
- OHALA, J. (1984). An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica*, 41, 1-16.
- OLSEN, C. L. (1981). Sex differences in English intonation observed in female impersonation. *Toronto Papers of the Speech and Voice Society*, 2, 30-49.
- OWREN, M. J. & CARDILLO, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America*, 119, 1727-1739.
- PABON, J. P. & PLOMP, R. (1988). Automatic phonetogram recording supplemented with acoustical voice-quality parameters. *Journal of Speech and Hearing Research*, 31, 710-722.

- PAUSEWANG GELFER, M. & MIKOS, V. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice*, 19, 544-554.
- PAUSEWANG GELFER, M. & RYAN YOUNG, S. (1997). Comparisons of intensity measures and their stability in male and female speakers. *Journal of Voice*, 11, 178-186.
- PEGORARO-KROOK, M. I. (1988). Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia Phoniatica*, 40, 82-90.
- PEPIOT, E. (2009). *The making of French vocalic triangles: the case of a woman's voice versus a man's voice* (50 p.). Mémoire de Master 1 sous la direction de Jean-Yves Dommergues. Université Paris VIII.
- PEPIOT, E. (2010). *Sur l'identification du genre par la voix chez des auditeurs anglophones et francophones* (84 p.). Mémoire de Master 2 sous la direction de Jean-Yves Dommergues. Université Paris VIII.
- PEPIOT, E. (2011). Voix de femmes, voix d'hommes : à propos de l'identification du genre par la voix chez des auditeurs anglophones et francophones. *Plovdiv University "Paissii Hilendarski" – Bulgaria, Scientific Works – Philology*, 49 (à paraître).
- PEPIOT, E. (2012). Les temps de traitement des voix de femmes et d'hommes sont-ils équivalents ? In *Actes des JEP-TALN-RECITAL 2012*, Grenoble, 153-160.
- PEPIOT, E. (2013). Voice, speech and gender: male-female acoustic differences and cross-language variation in English and French speakers. *Actes des Rencontres Jeunes Chercheurs 2011 et 2012 de l'ED 268* (à paraître).
- PERRY, T. L., OHDE, R. N. & ASHMEAD, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, 109, 2988-2998.
- PETERSON, G. E. & BARNEY, H. L. (1952). Control methods used in a study of the identification of vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- PODSEVA, R. J. (2007). Phonation type as a stylistic variable: the use of falsetto in constructing a persona. *Journal of Sociolinguistics*, 11, 478-504.
- RAY, G. B. & ZAHN, C. J. (1990). Regional speech rates in the United States: A preliminary analysis. *Communication Research Reports*, 7, 34-37.

- REPP, B. & LIBERMAN, A. (1987). Phonetic category boundaries are flexible. In Harnad, S. (dir.), *Categorical Perception: The groundwork of cognition*. New York, Etats-Unis : Cambridge University Press, 89-112.
- ROBB, M., GILBERT, H. & LERMAN, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatica et Logopaedica*, 57, 125-133.
- ROSE, P. (1991). How effective are long term mean and standard deviation as normalization parameters for tonal fundamental frequency? *Speech Communication*, 10, 229-247.
- RYALLS, J., ZIPPRER, A. & BALDAUFF, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language and Hearing Research*, 40, 642-645.
- RYALLS, J. H., & LIEBERMAN, P. (1982). Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America*, 72, 1631-1634.
- SACHS, J., LIEBERMAN, P. & ERICKSON, D. (1973). Anatomical and cultural determinants of male and female speech. In Shuy, R. W & Fasold, R. W. (dir.), *Language Attitudes: Current Trends and Prospects*. Washington, Etats-Unis : Georgetown University Press, 74-84.
- SCHMIDT, C. P., PAUSEWANG-GELFER, M., ANDREWS, M. L. (1990). Intensity range as a function of task and training. *Journal of Voice*, 4, 30-36.
- SCHWARTZ, M. F. (1968). Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America*, 43, 1178-1179.
- SCHWARTZ, M. F. & RINE, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America*, 44, 1736-1737.
- SCHWEINBERGER, S. R., CASPER, C., HAUTHAL, N., KAUFMANN, J. M., KAWAHARA, H., KLOTH, N., ROBERTSON, D. M., SIMPSON, A. P. & ZÄSKE, R. (2008). Auditory adaptation in voice perception. *Current Biology*, 18, 684-688.
- SCOTTO DI CARLO, N. & GERMAIN, A. (1985). A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica*, 42, 188-197.
- SHADLE, C. H., BADIN, P. & MOULINIER, A. (1991). Towards the spectral characteristics of fricative consonants. In *Proceedings of the 12th International Congress of Phonetic Sciences*, Université de Provence, 42-45.
- SHRIVASTAV, R. & SAPIENZA, C. M. (2003). Objective measures of breathy voice quality obtained using an auditory model. *Journal of the Acoustical Society of America*, 114, 2217-24.

- SIGMUND, M. (2008). Gender distinction using short segments of speech signal. *International Journal of Computer Science and Network Security*, 8, 159-162.
- SIMPSON, A. P. (1998). *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK), 33.
- SIMPSON, A. P. (2000). Gender-specific differences in the articulatory and acoustic realization of interword vowel sequences in American English. In Hoole, P. et al. (dir.), *5th Seminar on Speech Production: Models and Data*. Kloster Seeon, 209-212.
- SIMPSON, A. P. (2001). Dynamic consequences of differences in male and female vocal tract dimensions. *Journal of the Acoustical Society of America*, 109, 2153-2164.
- SIMPSON, A. P. (2002). Gender-specific articulatory-acoustic relations in vowel sequences. *Journal of Phonetics*, 30, 417-435.
- SIMPSON, A. P. (2003). Possible articulatory reasons for sex-specific differences in vowel duration. In *Proceedings of the 6th International Seminar on Speech Production*, Sydney, 261-266.
- SIMPSON, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3, 621-640.
- SIMPSON, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, 40, 477-490.
- SIMPSON, A. P. & ERICSDOTTER, C. (2003). Sex-specific durational differences in English and Swedish. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 1113-1116.
- SIMPSON, A. P. & ERICSDOTTER, C. (2007). Sex-specific differences in f₀ and vowel space. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 933-936.
- SMITH, B. L. (1978). Effects of place of articulation and vowel environment on voiced stop consonant production. *Glossa*, 12, 163-175.
- SÖDERSTEN, M. & LINDESTAD, P. A. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech and Hearing Research*, 33, 601-611.

- SÖDERSTEN, M. & LINDESTAD, P. A. (1992). A comparison of vocal fold closure in rigid telescopic and flexible fiberoptic laryngostroboscopy. *Acta Otolaryngologica*, 112, 144-150.
- SÖDERSTEN, M., HERTEGARD, S. & HAMMARBERG, B. (1995). Glottal closure, transglottal airflow, and voice quality in healthy middle-aged woman. *Journal of Voice*, 9, 182-197.
- SOKHI, D. S., HUNTER, M. D., WILKINSON, I. D. & WOODRUFF, P. W. (2005). Male and female voices activate distinct regions in the male brain. *NeuroImage*, 27, 572-578.
- SORENSEN D. & HORII, Y. (1982). Cigarette smoking and voice fundamental frequency. *Journal of Communication Disorders*, 15, 125-144.
- SPENCER, L. (1988). Speech characteristics of male-to-female transsexuals: A perceptual and acoustic study. *Folia Phoniatica et Logopaedica*, 40, 31-42.
- STEVENS, K. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, 27, 484-493.
- STEVENS, K. (1999). *Acoustic Phonetics*, Cambridge, Etats-Unis : MIT Press.
- STOICHEFF, M. L. (1981). Speaking fundamental frequency characteristics of nonsmoking female adults. *Journal of Speech and Hearing Research*, 24, 437-441.
- STRAND, E. (2000). *Gender stereotype effects in speech processing* (184 p.). Thèse de doctorat sous la direction de Keith Johnson. The Ohio State University.
- STRAND, E. & JOHNSON, K. (1996) Gradient and visual speaker normalization in the perception of fricatives. In *Natural language processing and speech technology: results of the 3rd KONENS conference*, Bielefeld, (D. Gibbon, dir.), October, 14-26. Berlin, Allemagne : Mouton de Gruyter.
- SWARTZ, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, 75, 983-992.
- SWEETING, P. M. & BAKEN, R. J. (1982). Voice onset time in a normal-aged population. *Journal of Speech and Hearing Research*, 25, 129-134.
- TAKEFUTA, Y., JANCOSSEK, E. G & BRUNT, M. (1972). A statistical analysis of melody curves in the intonation of American English. In *Proceedings of the 7th International Congress of Phonetic Sciences*, Montreal (1971), 1035-1039.
- TARRIER, J-M. (2003). L'enregistrement et la prise de son. In Delais-Roussarie, E. & Durand, J. (dir.), *Corpus et variation en phonologie du français : méthodes et analyses*. Toulouse : Presses Universitaires de France, 187-212.

- TESTON, B. (2004). L'évaluation instrumentale des dysphonies. Etat actuel et perspectives d'évolution. In Giovanni, A. (dir.), *Le bilan d'une dysphonie*. Marseille : Solal, 105-169.
- TITZE, I. R. (1976). On the mechanics of vocal fold vibration. *Journal of the Acoustical Society of America*, 60, 1366-1380.
- TITZE, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 85, 1699-1707.
- TITZE, I. R. & STORY, B. H. (1997). Acoustic interactions of the voice source with the lower vocal tract. *Journal of the Acoustical Society of America*, 101, 2234-2243.
- TORRE, P., & BARLOW, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42, 324-333.
- TRAUNMULLER, H. (1984). Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels. *Speech Communication*, 3, 49-61.
- TRAUNMULLER, H. (1988). Paralinguistic variation and invariance in the characteristic frequencies of vowels, *Phonetica*, 45, 1-29.
- TRAUNMULLER, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88, 97-100.
- TRAUNMULLER, H. & ERIKSSON, A. (1994). The size of F0 excursions in speech production and perception. *Working Papers of the Department of Linguistics and Phonetics of Lund University*, 43, 136-139.
- TRITTIN, P. J. & DE SANTOS Y LLEO, A. (2005). Voice quality analysis of male and female Spanish speakers. *Speech Communication*, 16, 359-368.
- TURNER, R. E., WALTERS, T. C., MONAGHAN, J. J. & PATTERSON R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *Journal of the Acoustical Society of America*, 125, 2374-2386.
- UNGER, R. K. & CRAWFORD, M. (1993). Commentary: sex and gender - The troubled relationship between terms and concepts. *Psychological Science*, 4, 122-124.
- VAISSIERE, J. (2005). Perception of intonation. In Pisoni, D. & Remez, R. (dir.), *The Handbook of Speech Perception*. Oxford, Royaume-Uni : Blackwell Publishers, 236-263.
- VAISSIERE, J. (2006). *La phonétique*. Paris : Presses Universitaires de France.

- VAN BORSEL, J., JANSSENS, J. & DE BODT, M. (2009). Breathiness as a Feminine Voice Characteristic: A Perceptual Approach. *Journal of Voice*, 23(3), 291-294.
- VAN DOMMELEN, W. A. (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, 33, 259-272.
- VERCHERAND, G. (2010). *Production et perception de la parole chuchotée en français : de la consonne à l'énoncé* (260 p.). Thèse de doctorat sous la direction de Philippe Martin. Université Paris VII.
- WADNERKAR, M. B., COWELL, P. E. & WHITESIDE, S. P. (2006). Speech across the menstrual cycle: A replication and extension study. *Neuroscience Letters*, 408, 21-24.
- WAGNER, A. & BRAUN, A. (2003). Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages. In *Proceedings of the 15th International Congress of Phonetic Sciences*, 651-654.
- WARREN, P. & DALY, N. (2000). Sex as a factor in rises in New Zealand English. In Holmes, J. (dir.), *Gendered speech in social context: Perspectives from Gownand Town*. Wellington, Nouvelle-Zélande : Victoria University Press, 99-115.
- WASSINK, A. B. (1999). *A sociophonetic analysis of Jamaican vowels* (367 p.). Thèse de doctorat sous la direction de Patrice Beddor et Lesley Milroy. University of Michigan.
- WHITESIDE, S. P. (1995). Temporal-based speaker sex differences in read speech: A sociophonetic approach. In *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm, 516-519.
- WHITESIDE, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. *Journal of the International Phonetic Association*, 26, 23-40.
- WHITESIDE, S. P. (1998a). Identification of a speaker's sex: a study of vowels. *Perceptual and Motor Skills*, 86, 579-584.
- WHITESIDE, S. P. (1998b). Identification of a speaker's sex: a fricative study. *Perceptual and Motor Skills*, 86, 587-591.
- WHITESIDE, S. P. (1998c). The identification of a speaker's sex from synthesized vowels. *Perceptual and Motor Skills*, 86, 595-600.
- WHITESIDE, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *Journal of the Acoustical Society of America*, 110, 464-478.

- WHITESIDE, S. P., HANSON, A. & COWELL, P. E. (2004a). Hormones and temporal components of speech: sex differences and effects of menstrual cyclicity on speech. *Neuroscience Letters*, 367, 44-47.
- WHITESIDE, S. P., HENRY, L. & DOBBIN, R. (2004b). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America*, 116, 1179-1183.
- WHITESIDE, S. P. & IRVING, C. J. (1997). Speakers' sex differences in voice onset time, some preliminary findings. *Perceptual and Motor Skills*, 85, 459-463.
- WHITESIDE, S. P. & MARSHALL, J. (2001). Developmental trends in Voice Onset Time: some evidence for sex differences. *Phonetica*, 58, 196-210.
- WILTSHIRE, A. (1995). Not by pitch alone: a view of transsexual vocal rehabilitation. *National Student Speech Language Hearing Association Journal*, 22, 53-57.
- WITTIG, M. (1992). *The Straight Mind and Other Essays*. Hemel Hempstead, Royaume-Uni : Harvester Wheatsheaf.
- WOLFE, V. I., RATUSNIK, D. L., SMITH, F. H. & NORTHROP, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorder*, 55, 43-50.
- WOLK, L., ABDELLI-BERUH, N. B. & SLAVIN, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, 26, e111-e116.
- WU, K. & CHILDERS, D. G. (1991). Gender recognition from speech: Part I. Coarse analysis. *Journal of the Acoustical Society of America*, 90, 1828-40.
- YANG, B. (1992). An acoustical study of Korean monophthongs produced by male and female speakers. *Journal of the Acoustical Society of America*, 91, 2280-2283.
- ZWICKER, E. (1961), Subdivision of the audible frequency range into critical bands. *Journal of the Acoustical Society of America*, 33, 248.
- ZHU, X. Y. & CAHILL, L. W. (1992). Automatic gender identification by voice. In *Proceedings of the Fourth Australian International Conference on Speech Science and Technology*, Brisbane, 738-743.

REFERENCES SITOGRAPHIQUES

- BOERSMA P. & WEENINK, D. (2010). *Praat: doing phonetics by computer (Version 5.1.44)*. <<http://www.praat.org/>>
- NIKOLOV, R., DITCHEVA, M., DOMMERGUES, J-Y., HERRY-BENIT, N. & RYST, E. (2011). *SaRP : Speech Analyzer Rapid Plot (2006-2011) Plotting vowels in F2-F1 scatter charts with multiple data sets*. <<http://web.uni-plovdiv.bg/rousni/sarp>>
- TOLONEN, H., KUULASMAA, K. & RUOKOKOSKI, E. (2000). *MONICA Population Survey Data Book*. W.H.O. <<http://www.ktl.fi/publications/monica/surveydb/title.htm>>
- TRAUNMULLER, H. & ERIKSSON, A. (1995). *The frequency range of the voice fundamental in the speech of male and female adults*. Manuscrit non-publié. Stockholm. <http://www.ling.su.se/staff/hartmut/f0_m&f.pdf>
- WELLS, J. (2000). *Computer-coding the IPA: a proposed extension of SAMPA*. <<http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>>

LISTE DES FIGURES

FIGURE 1 - Représentation de l'intérieur du larynx. Les plis vocaux (vocal folds) sont visibles au centre de l'image, de part et d'autre de la trachée. (Gray, 1918).....	- 20 -
FIGURE 2 - Continuum des types de phonation en fonction du degré d'aperture des plis vocaux. (Gordon & Ladefoged, 2001)	- 22 -
FIGURE 3 - Représentation schématique du spectre fréquentiel d'un son produit par une vibration des plis vocaux à 200 Hz, avant le passage dans le conduit vocal. On y voit la fréquence fondamentale (notée F0) et les fréquences harmoniques. (D'après Stevens, 1999).....	- 23 -
FIGURE 4 - Représentation schématique du spectre fréquentiel d'une voyelle centrale, produite avec une fréquence fondamentale à 200 Hz. Les quatre premiers formants (F1,F2, F3, F4) sont indiqués en rouge. (D'après Stevens, 1999).....	- 24 -
FIGURE 5 - Représentation schématique de l'évolution du F0 moyen des femmes (courbe grise) et des hommes (courbe noire) adultes avec l'âge. (D'après Pegoraro-Krook, 1988 et Traummüller & Eriksson, 1995).....	- 26 -
FIGURE 6 - Coupe horizontale du larynx d'un homme (en pointillés) et d'une femme (trait continu), les pourcentages dans la partie droite illustrent la différence de taille des plis vocaux. (Kahane, 1978).....	- 28 -
FIGURE 7 - Représentations de la fréquence fondamentale d'un homme (en rouge) et d'une femme (en vert) pour la phrase allemande « Riechst du nicht die frische Luft? » avec une échelle linéaire en hertz (à gauche) et en demi-tons (à droite). Les lignes verticales au centre indiquent les plages de variations pour chacune des représentations. (Simpson, 2009)	- 31 -
FIGURE 8 - Triangle vocalique moyen pour les femmes (trait continu) et les hommes (en pointillés) locuteurs de l'anglais américain. L'axe vertical indique la fréquence (en Hz) du premier formant (F1), et l'axe horizontal celle du deuxième formant (F2). (Simpson, 2009, d'après les données recueillies par Hillenbrand et al., 1995).....	- 35 -
FIGURE 9 - Différences entre les valeurs moyennes de F1, F2 et F3 chez les femmes et chez les hommes dans 26 langues, exprimées en bark. (Johnson, 2005).....	- 36 -
FIGURE 10 - Triangles vocaliques de l'allemand pour des locuteurs masculins (trait foncé) et féminins (trait clair) réalisés à partir de données recueillies sur 27 femmes et 14 hommes. (Simpson & Ericsson, 2007)	- 38 -
FIGURE 11 - Représentation schématique de la position de la langue dans le conduit vocal lors de l'articulation d'un [a] (en bas à gauche) et d'un [i] (en haut à droite) par un homme (en pointillés) et une femme (trait continu). Les symboles « + » (homme) et « x » (femme) indiquent un point central de la langue pour les deux voyelles. (Simpson, 2009).....	- 39 -
FIGURE 12 - Spectres de la fricative [s] du mot « Steven », prononcée par un homme (à gauche) et une femme (à droite) anglophones obtenus avec le logiciel <i>Speech Analyzer</i> . L'axe vertical représente l'intensité, l'axe horizontal les fréquences.	- 44 -
FIGURE 13 - Pic spectral moyen (exprimé en Hz) pour chacune des fricatives sourdes de l'anglais lorsque produite par des hommes (en gris) et des femmes (en noir). (Fox et Nissen, 2005)	- 45 -
FIGURE 14 - Représentation des cycles d'ouverture-fermeture des plis vocaux avec une voix modale (en haut) et avec une voix soufflée (en bas). Le temps figure sur l'axe horizontal, le flux d'air glottal (Ug) sur l'axe vertical. (D'après Klatt & Klatt, 1990).....	- 47 -
FIGURE 15 - Spectrogrammes de la séquence [sti:] dans le mot « Steven », prononcée par un homme (en haut) et une femme (en bas) anglophones, obtenus avec le logiciel <i>Praat</i> . Les deux cadres rouges représentent les VOT des locuteurs après le relâchement du [t], la courbe bleue est celle du F0.	- 51 -
FIGURE 16 - Durée moyenne (en ms) des phrases produites par les hommes (en haut) et les femmes (en bas) locuteurs de l'anglais américain. (Byrd, 1994)	- 58 -

FIGURE 17 - IRMf montrant sous trois angles différents la zone du cerveau (en jaune) activée plus fortement par les voix de femmes (A) et celle activée plus fortement par les voix d'hommes (B). (Sokhi et al., 2005)	- 63 -
FIGURE 18 - Pourcentage d'identifications du genre réussies par des auditeurs anglophones américains à partir de voyelles normales, filtrées au dessus de 255 Hz, et chuchotées. (D'après Lass et al., 1976)	- 71 -
FIGURE 19 - Pourcentage d'identifications du genre réussies par des auditeurs anglophones américains et francophones à partir de phrases ayant soit leur contour intonatif préservé, soit un F0 plat. (D'après Pépiot, 2010).....	- 75 -
FIGURE 20 - Degré de féminité perçu (échelle de 1 à 5) pour des voyelles [a] produites par sept locutrices du hollandais en mode non-soufflé et soufflé. Les jugements ont été émis par 40 auditeurs. (D'après Van Borsel et al., 2009).....	- 76 -
FIGURE 21 - Frontière perceptive entre hood et hud (valeur du F1 de la voyelle -en hertz-) en fonction du genre du locuteur, chez des auditeurs anglophones américains. (D'après Johnson et al., 1999).....	- 81 -
FIGURE 22 - Pourcentage d'identifications correctes de voyelles isolées /i/ et /u/ en fonction de leur F0, chez des locuteurs anglophones américains. (Diehl et al., 1996)	- 84 -
FIGURE 23 - Temps de réaction moyen (en ms) de sujets anglophones américains ayant pour tâche de répéter un mot, selon que ce dernier est produit par une voix de femme (stéréotypique ou non) ou une voix d'homme (stéréotypique ou non). (Strand, 2000)	- 85 -
FIGURE 24 - Extraction du mot api de la phrase il a dit « api » deux fois prononcée par le locuteur H2FR à l'aide du logiciel <i>Praat</i> . La partie supérieure représente l'onde sonore ; la partie inférieure le spectrogramme et la courbe de F0 (en bleu). La zone sélectionnée figure en rose sur l'onde sonore.	- 104 -
FIGURE 25 - Représentation du mot chipi produit par la locutrice F1FR, segmenté et étiqueté dans le logiciel <i>Praat</i> . La partie supérieure de la figure comprend l'onde sonore, la partie médiane le spectrogramme et la courbe de F0, et la partie inférieure les trois paliers d'étiquetage (phones, C/V, mot).	- 106 -
FIGURE 26 - Courbe de F0 du mot zipi prononcé par le locuteur H2FR, obtenue dans <i>Praat</i> via un fichier Pitch. Dans la partie supérieure figure des données calculées automatiquement par <i>Praat</i> , dont la plage de variation, encadrée ici en rouge, et exprimée en différentes unités de mesure.	- 107 -
FIGURE 27 - Représentation du mot tappy, produit par la locutrice F1AN, dans le logiciel <i>Praat</i> . La partie supérieure de la figure comprend l'onde sonore, la partie inférieure le spectrogramme et la courbe de F0 (en bleu). La sélection, visible en rose sur l'onde sonore, débute au relâchement de la consonne occlusive et se termine au démarrage du voisement. Sa durée (exprimée en s), qui constitue le <i>VOT</i> , est encadrée en rouge dans la partie inférieure gauche de la figure (ici 87 ms).	- 109 -
FIGURE 28 - Représentation de la voyelle [æ] du mot shappy produit par le locuteur H1AN, dans le logiciel <i>Praat</i> . La partie supérieure gauche de la figure comprend l'onde sonore, la partie inférieure gauche le spectrogramme avec en pointillés rouges les formants détectés par <i>Praat</i> . Le spectre correspondant à la partie sélectionnée (en rose sur l'onde sonore) est visible en haut à droite et les valeurs de F1, F2, F3 détectées par <i>Praat</i> pour cette même partie de la voyelle figurent dans la fenêtre se trouvant en bas à droite.	- 110 -
FIGURE 29 - Spectre de la voyelle [æ] du mot appy produit par la locutrice F2AN, obtenu dans le logiciel <i>Praat</i> . Les deux premiers harmoniques (H1 et H2) sont visibles sur la partie gauche du spectre. La différence d'intensité (en dB) entre ces deux items est ici mentionnée en rouge.	- 111 -
FIGURE 30 - Triangles vocaliques moyens des hommes (n=2), en trait rouge, et des femmes francophones (n=2), en trait noir, pour les trois voyelles cardinales [i],[a] et [u]. Les valeurs sont données en hertz dans la partie gauche, en barks dans la partie droite. Le premier formant est représenté sur l'axe vertical et le deuxième formant sur l'axe horizontal.	- 125 -

FIGURE 31 - Triangles vocaliques moyens des hommes (n=2), en trait rouge, et des femmes anglophones (n=2), en trait noir, pour les trois voyelles [i:],[æ] et [u:]. Les valeurs sont données en hertz dans la partie gauche, en barks dans la partie droite. Le premier formant est représenté sur l'axe vertical et le deuxième formant sur l'axe horizontal.....	- 127 -
FIGURE 32 - Répartition temporelle consonnes / voyelles exprimée en pourcentages moyens, pour les 24 mots de type CVCV produits par les hommes (n=2) et les femmes francophones (n=2).....	- 139 -
FIGURE 33 - Répartition temporelle consonnes / voyelles exprimée en pourcentages moyens, pour les 24 mots de type 'CVCV produits par les hommes (n=2) et les femmes anglophones (n=2).....	- 140 -
FIGURE 34 - Graphique des interactions représentant le pic spectral moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs francophones, avec les barres d'erreur-type.	- 141 -
FIGURE 35 - Graphique des interactions représentant le pic spectral moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs anglophones, avec les barres d'erreur-type.....	- 142 -
FIGURE 36 - Graphique des interactions représentant le centre de gravité moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs francophones, avec les barres d'erreur-type.	- 143 -
FIGURE 37 - Graphique des interactions représentant le centre de gravité moyen (en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les quatre locuteurs anglophones, avec les barres d'erreur-type.....	- 144 -
FIGURE 38 - <i>VOT</i> moyen (ms) en fonction du type de consonne initiale et du genre du locuteur (femme ou homme) pour les quatre locuteurs francophones, avec les barres d'erreur-type.....	- 146 -
FIGURE 39 - <i>VOT</i> moyen (ms) en fonction du type de consonne initiale et du genre du locuteur (femme ou homme) pour les quatre locuteurs anglophones, avec les barres d'erreur-type.	- 147 -
FIGURE 40 - Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.....	- 148 -
FIGURE 41 - Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de consonne et du genre du locuteur (femme ou homme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.....	- 149 -
FIGURE 42 - Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.	- 150 -
FIGURE 43 - Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de consonne et du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.....	- 151 -
FIGURE 44 - Graphiques représentant la fréquence moyenne (en Hz) des trois premiers formants vocaliques (F1, F2, F3) pour les trois voyelles cardinales [i], [a], [u] produites par les locutrices (n=2) et locuteurs (n=2) francophones, avec les barres d'erreur-type.	- 153 -
FIGURE 45 - Graphiques représentant la fréquence moyenne (en Hz) des trois premiers formants vocaliques (F1, F2, F3) pour les trois voyelles [i:], [æ], [u:] produites par les locutrices (n=2) et locuteurs (n=2) anglophones, avec les barres d'erreur-type.	- 155 -
FIGURE 46 - Différence d'intensité H1-H2 (dB) sur les voyelles ouvertes en fonction du type de locuteur : femme anglophone (n=2), femme francophone (n=2), homme anglophone (n=2), homme francophone (n=2), avec les barres d'erreur-type.	- 157 -
FIGURE 47 - Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.....	- 158 -

FIGURE 48 - Graphique des interactions représentant le F0 moyen (exprimé en Hz) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.....	- 159 -
FIGURE 49 - Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.	- 160 -
FIGURE 50 - Graphique des interactions représentant la durée moyenne (exprimée en ms) en fonction du type de voyelle et du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.....	- 161 -
FIGURE 51 - F0 moyen sur les mots dissyllabiques (exprimé en Hz) en fonction du genre du locuteur (femme ou homme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.....	- 162 -
FIGURE 52 - F0 moyen sur les mots dissyllabiques (exprimé en Hz) en fonction du genre du locuteur (femme ou homme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.	- 162 -
FIGURE 53 - Plage de variation moyenne sur les mots dissyllabiques (exprimée en dt) en fonction du genre du locuteur (homme ou femme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.	- 164 -
FIGURE 54 - Plage de variation moyenne sur les mots dissyllabiques (exprimée en dt) en fonction du genre du locuteur (homme ou femme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.	- 165 -
FIGURE 55 - Durée moyenne des mots dissyllabiques (exprimée en ms) en fonction du genre du locuteur (femme ou homme) pour les locuteurs francophones (n=4), avec les barres d'erreur-type.....	- 166 -
FIGURE 56 - Durée moyenne des mots dissyllabiques (exprimée en ms) en fonction du genre du locuteur (femme ou homme) pour les locuteurs anglophones (n=4), avec les barres d'erreur-type.	- 166 -
FIGURE 57 - Représentation schématique des découpages opérés sur chaque mot pour obtenir les stimuli de l'expérience. L'exemple utilisé ici est le mot [ʃipi] produit par la locutrice F2FR : son onde sonore et son spectrogramme, obtenus avec <i>Praat</i> , sont visibles respectivement dans la partie centrale et inférieure de la figure. Les barres verticales indiquent les frontières du découpage.	- 183 -
FIGURE 58 - Capture d'écran de l'interface visuelle de l'expérience d'identification du genre par la voix, dans sa version française. Dans la partie supérieure se trouve le dispositif de réponse et la jauge figurant en bas de l'écran indique l'état d'avancement de l'expérience.	- 188 -
FIGURE 59 - Pourcentage d'identifications du genre correctes sur les cinq grandes catégories de stimuli pour les auditeurs francophones, avec les barres d'erreur.	- 206 -
FIGURE 60 - Pourcentage d'identifications du genre correctes sur les cinq grandes catégories de stimuli pour les auditeurs anglophones, avec les barres d'erreur.	- 207 -
FIGURE 61 - Pourcentage d'identifications du genre correctes pour les auditeurs francophones en fonction du locuteur produisant les stimuli, avec les barres d'erreur.	- 209 -
FIGURE 62 - Pourcentage d'identifications du genre correctes pour les auditeurs anglophones en fonction du locuteur produisant les stimuli, avec les barres d'erreur.	- 210 -
FIGURE 63 - Pourcentage d'identifications du genre correctes pour les auditeurs francophones en fonction de la consonne présentée isolément, avec les barres d'erreur.....	- 211 -
FIGURE 64 - Pourcentage d'identifications du genre correctes pour les auditeurs anglophones en fonction de la consonne présentée isolément, avec les barres d'erreur.....	- 214 -
FIGURE 65 - Pourcentage d'identifications du genre correctes pour les auditeurs francophones en fonction de la voyelle présentée isolément, avec les barres d'erreur.	- 216 -
FIGURE 66 - Pourcentage d'identifications du genre correctes pour les auditeurs anglophones en fonction de la voyelle présentée isolément, avec les barres d'erreur.	- 218 -

FIGURE 67 - Pourcentage d'identifications du genre correctes obtenu par les auditeurs francophones en fonction du type de combinaison « C + V », avec les barres d'erreur.	219 -
FIGURE 68 - Pourcentage d'identifications du genre correctes obtenu par les auditeurs anglophones en fonction du type de combinaison « C + V », avec les barres d'erreur.	222 -
FIGURE 69 - Degré de certitude moyen (échelle de 0 à 7) exprimé par les auditeurs francophones en fonction du type de mot dissyllabique, avec les barres d'erreur.	225 -
FIGURE 70 - Degré de certitude moyen (échelle de 0 à 7) exprimé par les auditeurs anglophones en fonction du type de mot dissyllabique, avec les barres d'erreur.	227 -
FIGURE 71 - Extraction du mot <i>cerveau</i> de la phrase <i>il a dit « cerveau »</i> deux fois prononcée par la locutrice F2, à l'aide du logiciel <i>Praat</i> . La partie supérieure représente l'onde sonore ; la partie inférieure le spectrogramme et la courbe de F0 (en bleu). La zone sélectionnée figure en rose sur l'onde sonore.	243 -
FIGURE 72 - Une participante à l'expérience de détection de mots, photographiée au début de l'expérience, lors de l'affichage de la consigne écrite à l'écran. La Button Box est visible dans la partie centrale basse de la photographie.	251 -
FIGURE 73 - Graphique des interactions représentant les temps de réponse moyens (exprimés en millisecondes) en fonction du genre des auditeurs (femme ou homme) et des conditions expérimentales (conditions A, B, C et D), avec les barres d'erreur.	257 -
FIGURE 74 - Graphique représentant les temps de réponse moyens des auditeurs (exprimés en millisecondes) en fonction du locuteur ayant produit le mot-cible, avec les barres d'erreur.	259 -
FIGURE 75 - Nuage de points illustrant l'absence de corrélation entre le temps de réponse des auditeurs et le moment de présentation (ordre chronologique) du mot-cible expérimental.	262 -
FIGURE 76 - Graphique représentant les temps de réponse moyens (exprimés en millisecondes) des 25 participants, toutes conditions et tous mots-cibles confondus, avec les barres d'erreur.	263 -
FIGURE 77 - Graphique représentant les temps de réponse moyens (exprimés en millisecondes) des auditeurs en fonction de la condition expérimentale, avec les barres d'erreur correspondantes.	264 -
FIGURE 78 - Graphique représentant les temps de réponse moyens (exprimés en millisecondes) des auditeurs, en fonction du type de voix (femme ou homme) produisant le mot-cible, avec les barres d'erreur.	266 -
FIGURE 79 - Schéma représentant les interactions entre les différents facteurs jouant un rôle dans la construction des différences acoustiques inter-genres et de leurs représentations mentales chez les auditeurs.	272 -

LISTE DES TABLEAUX

TABLEAU 1 - Transcription orthographique utilisée pour les mots et pseudo-mots présentés aux locuteurs francophones.	101 -
TABLEAU 2 - Transcription orthographique utilisée pour les mots et pseudo-mots présentés aux locuteurs anglophones.	102 -
TABLEAU 3 - Pic spectral moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs francophones.	114 -
TABLEAU 4 - Pic spectral moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs anglophones.	115 -
TABLEAU 5 - Centre de gravité moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs francophones.	116 -
TABLEAU 6 - Centre de gravité moyen (Hz) de chacune des consonnes initiales pour les quatre locuteurs anglophones.	117 -
TABLEAU 7 - Taille des locuteurs francophones (cm) et fréquences de résonance moyennes (pic spectral et centre de gravité, en Hz) des consonnes initiales.	118 -
TABLEAU 8 - Taille des locuteurs anglophones (cm) et fréquences de résonance moyennes (pic spectral et centre de gravité, en Hz) des consonnes initiales.	119 -
TABLEAU 9 - <i>VOT</i> moyen (ms) de chacune des consonnes initiales pour les quatre locuteurs francophones.	119 -
TABLEAU 10 - <i>VOT</i> moyen (ms) de chacune des consonnes initiales pour les quatre locuteurs anglophones.	120 -
TABLEAU 11 - F0 moyen (Hz) de chacune des consonnes initiales voisées pour les quatre locuteurs francophones.	121 -
TABLEAU 12 - F0 moyen (Hz) de chacune des consonnes initiales voisées pour les quatre locuteurs anglophones.	121 -
TABLEAU 13 - Durée moyenne (ms) de chacune des consonnes initiales pour les quatre locuteurs francophones.	122 -
TABLEAU 14 - Durée moyenne (ms) de chacune des consonnes initiales pour les quatre locuteurs anglophones.	123 -
TABLEAU 15 - Fréquence moyenne (Hz), écart-type (Hz) et ratio femmes / hommes pour les F1, F2 et F3 des voyelles en syllabe initiale, chez les locuteurs francophones.	124 -
TABLEAU 16 - Fréquence moyenne (Hz), écart-type (Hz) et ratio femmes / hommes pour les F1, F2 et F3 des voyelles en syllabe initiale et accentuée, chez les locuteurs anglophones.	126 -
TABLEAU 17 - Taille des locuteurs francophones et valeurs moyennes des trois premiers formants vocaliques, toutes voyelles ([i], [a], [u]) confondues.	127 -
TABLEAU 18 - Taille des locuteurs anglophones et valeurs moyennes des trois premiers formants vocaliques, toutes voyelles ([i:], [□], [u:]) confondues.	128 -
TABLEAU 19 - Différence d'intensité entre le premier et le deuxième harmonique pour les neuf occurrences de la voyelle [a] produite par les locuteurs francophones.	129 -
TABLEAU 20 - Différence d'intensité entre le premier et le deuxième harmonique pour les neuf occurrences de la voyelle [a] produites par les locuteurs anglophones.	130 -
TABLEAU 21 - F0 moyen (Hz) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs francophones.	131 -

TABLEAU 22 - F0 moyen (Hz) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs anglophones.	- 131 -
TABLEAU 23 - Durée moyenne (ms) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs francophones.	- 132 -
TABLEAU 24 - Durée moyenne (ms) de chacune des voyelles présentes en syllabe 1 pour les quatre locuteurs anglophones.	- 132 -
TABLEAU 25 - F0 moyen (Hz) pour chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs francophones.	- 134 -
TABLEAU 26 - F0 moyen (Hz) pour chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs anglophones.	- 135 -
TABLEAU 27 - Plage de variation moyenne de F0 sur les 27 mots dissyllabiques prononcés par les locuteurs francophones, exprimée en hertz et en demi-tons.	- 136 -
TABLEAU 28 - Plage de variation moyenne de F0 sur les 27 mots dissyllabiques prononcés par les locuteurs anglophones, exprimée en hertz et en demi-tons.	- 136 -
TABLEAU 29 - Durée moyenne (ms) de chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs francophones.	- 137 -
TABLEAU 30 - Durée moyenne (ms) de chacun des 27 (pseudo-)mots prononcés par les quatre locuteurs anglophones.	- 138 -
TABLEAU 31 - Pourcentage d'identifications réussies et degré de certitude moyen pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre par la voix menée sur les francophones.	- 192 -
TABLEAU 32 - Pourcentage d'identifications réussies et degré de certitude moyen pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre par la voix menée sur les anglophones.	- 193 -
TABLEAU 33 - Répartition des réponses incorrectes pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre menée sur les francophones en fonction des locuteurs ayant produit les stimuli.	- 194 -
TABLEAU 34 - Répartition des réponses incorrectes pour les cinq grandes catégories de stimuli utilisées dans l'expérience d'identification du genre menée sur les anglophones en fonction des locuteurs ayant produit les stimuli.	- 195 -
TABLEAU 35 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de consonnes initiales présentées dans l'expérience d'identification du genre menée sur les francophones.	- 197 -
TABLEAU 36 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de consonnes initiales présentées dans l'expérience d'identification du genre menée sur les anglophones.	- 198 -
TABLEAU 37 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de voyelles initiales présentées dans l'expérience d'identification du genre menée sur les francophones.	- 200 -
TABLEAU 38 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de voyelles initiales présentées dans l'expérience d'identification du genre menée sur les anglophones.	- 201 -
TABLEAU 39 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de combinaisons « consonne + voyelle » présentées dans l'expérience d'identification du genre par la voix menée sur les francophones.	- 202 -
TABLEAU 40 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de combinaisons « consonne + voyelle » présentées dans l'expérience d'identification du genre par la voix menée sur les anglophones.	- 203 -

TABLEAU 41 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de mot dissyllabique présentés dans l'expérience d'identification du genre par la voix menée sur les francophones.	- 204 -
TABLEAU 42 - Pourcentage d'identifications réussies et degré de certitude moyen pour les différents types de mot dissyllabique présentés dans l'expérience d'identification du genre par la voix menée sur les anglophones.	- 205 -
TABLEAU 43 - Résultats des tests de Pearson (corrélation) conduits sur les données relatives aux degrés de certitude moyens des auditeurs francophones pour les stimuli de type C + V. Le symbole « * » indique une corrélation significative entre les deux variables (avec un seuil $p < 0,05$).	- 221 -
TABLEAU 44 - Résultats des tests de Pearson (corrélation) conduits sur les données relatives aux degrés de certitude moyens des auditeurs anglophones pour les stimuli de type C + V. Le symbole « * » indique une corrélation significative entre les deux variables (avec un seuil $p < 0,05$).	- 223 -
TABLEAU 45 - Résultats des tests de Pearson conduits sur les données relatives aux degrés de certitude moyens des auditeurs francophones pour les stimuli de type « mots entiers ». Le symbole « * » indique une corrélation significative entre les deux variables (seuil $p < 0,05$).	- 226 -
TABLEAU 46 - Résultats des tests de Pearson conduits sur les données relatives aux degrés de certitude moyens des auditeurs anglophones pour les stimuli de type « mots entiers ». Le symbole « * » indique une corrélation significative entre les deux variables (seuil $p < 0,05$).	- 228 -
TABLEAU 47 - Degré d'influence estimé des paramètres acoustiques pouvant jouer un rôle dans l'identification du genre par la voix à partir de mots dissyllabique, pour les auditeurs francophones et anglophones américains. Le symbole « \emptyset » indique l'absence d'influence, puis échelle graduée allant de « * » (très faible influence) à « ***** » (très forte influence).	- 234 -
TABLEAU 48 - Répartition des stimuli de type « voix d'hommes » et « voix de femmes » dans l'expérience de détection de mots.	- 247 -
TABLEAU 49 - Ordre de présentation et nature des stimuli de l'expérience de détection de mots. La colonne de gauche indique le type de série. Le début des différents « blocs » est mentionné sur une ligne dédiée. Les mots-cibles de chaque série ainsi que tous les items des séries pré-expérimentales et expérimentales apparaissent en gras. Les mots-cibles expérimentaux sont en rouge.	- 247 -
TABLEAU 50 - Nombre de réponses recueillies, temps de réponse moyens et écarts-types (exprimés en millisecondes) pour les 17 auditrices dans chacune des quatre conditions expérimentales, avec le mot-cible étage.	- 254 -
TABLEAU 51 - Nombre de réponses recueillies, temps de réponse moyens et écarts-types (exprimés en millisecondes) pour les huit auditeurs masculins dans chacune des quatre conditions expérimentales, avec le mot-cible étage.	- 255 -
TABLEAU 52 - Nombre de réponses recueillies, temps de réponse moyens et écarts-types (exprimés en millisecondes) pour l'ensemble des participants dans chacune des quatre conditions expérimentales, avec le mot-cible étage.	- 256 -
TABLEAU 53 - Durée du mot-cible « étage » en fonction du locuteur l'ayant produit, et temps de réponse moyen des auditeurs pour chacun de ces huit items, toutes conditions confondues.	- 260 -
TABLEAU 54 - F0 moyen du mot-cible « étage » en fonction du locuteur l'ayant produit, et temps de réponse moyen des auditeurs pour chacun de ces huit items, toutes conditions confondues.	- 261 -
TABLEAU 55 - Informations détaillées sur les anglophones américains ayant participé à l'expérience d'identification du genre par la voix.	- 309 -
TABLEAU 56 - Informations détaillées sur les francophones ayant participé à l'expérience d'identification du genre par la voix.	- 310 -
TABLEAU 57 - Informations détaillées sur les participants à l'expérience de détection de mots.	- 311 -

ANNEXES

ANNEXE 1 : Téléchargement des fichiers sons analysés et des stimuli utilisés pour les expériences

Les enregistrements **analysés acoustiquement dans le chapitre 2** sont téléchargeables aux adresses suivantes :

- Enregistrements des locuteurs anglophones américains du nord-est des Etats-Unis :
<http://erwan.pepiot.free.fr/doctorat/enregistrementsAN.zip>
- Enregistrements des locuteurs francophones parisiens :
<http://erwan.pepiot.free.fr/doctorat/enregistrementsFR.zip>

Chaque archive contient les 27 mots produits par les différents locuteurs au format *Wave* (échantillonnage à 44 kHz), accompagnés chacun de leur fichier *TextGrid* utilisable dans le logiciel *Praat*.

Les stimuli utilisés dans les **expériences d'identification du genre par la voix du chapitre 3** (format *Wave*, 44 kHz) sont téléchargeables aux adresses suivantes :

- Stimuli de l'expérience conduite sur les auditeurs anglophones américains :
<http://erwan.pepiot.free.fr/doctorat/identdugenrestimuliAN.zip>
- Stimuli de l'expérience conduite sur les auditeurs francophones :
<http://erwan.pepiot.free.fr/doctorat/identdugenrestimuliFR.zip>

Les stimuli utilisés dans l'**expérience de détection de mots du chapitre 4** (format *Wave*, 44 kHz) sont téléchargeables aux adresses suivantes :

- <http://erwan.pepiot.free.fr/doctorat/detectiondemotsstimuli.zip>

L'archive contient 70 fichiers sons correspondant aux 70 séries de stimuli de l'expérience, ainsi qu'un tableau au format *Excel* contenant une description de chaque série (mots, locuteurs).

ANNEXE 2 : Participants anglophones à l'expérience d'identification du genre par la voix

Tableau 55 – Informations détaillées sur les anglophones américains ayant participé à l'expérience d'identification du genre par la voix.

	Initiales	Age	Genre	Profession	Langue(s) maternelle(s)	Etat d'origine²⁵¹
1	AK	18	F	étudiante à AUP ²⁵²	anglais	Californie
2	MM	20	F	étudiante à AUP	anglais	Californie
3	CW	20	F	étudiante à AUP	anglais	Colorado
4	JT	20	F	étudiante à AUP	anglais, espagnol	New Jersey
5	HV	20	F	étudiante à AUP	anglais	Californie
6	JR	21	F	étudiante à AUP	anglais	Maine
7	TA	22	F	étudiante à AUP	anglais	New Jersey
8	AW	22	F	étudiante à AUP	anglais, espagnol	New York
9	ER	23	F	étudiante à NYU ²⁵³ Paris	anglais	New York
10	HS	24	F	étudiante à NYU Paris	anglais	Californie
11	BEL	24	F	étudiante à NYU Paris	anglais	New York
12	REA	25	F	coordinatrice à NYU Paris	anglais	Illinois
13	LT	25	F	étudiante à NYU Paris	anglais	Massachusetts
14	SH	26	F	étudiante à AUP	anglais	New York
15	RW	26	F	étudiante à NYU Paris	anglais	Californie
16	SM	30	F	étudiante en école de FLE	anglais	Texas
17	EM	33	F	étudiante en école de FLE	anglais	Pennsylvanie
18	BW	38	F	étudiante en école de FLE	anglais	Pennsylvanie
19	CH	20	H	étudiant à l'AUP	anglais	Californie
20	TC	21	H	étudiant à l'AUP	anglais	Massachusetts
21	WS	22	H	étudiant à NYU Paris	anglais	Rhode Island
22	TJM	24	H	étudiant à NYU Paris	anglais	Wisconsin
23	JW	30	H	étudiant à NYU Paris	anglais	Pennsylvanie
24	TBJ	31	H	étudiant en école de FLE	anglais	Caroline du Sud
25	TR	35	H	étudiant en école de FLE	anglais	Wisconsin

²⁵¹ Etat dans lequel les participants ont résidé la plus grande partie de leur vie.

²⁵² *American University of Paris.*

²⁵³ *New York University.*

ANNEXE 3 : Participants francophones à l'expérience d'identification du genre par la voix

Tableau 56 – Informations détaillées sur les francophones ayant participé à l'expérience d'identification du genre par la voix.

	Initiales	Age	Genre	Profession	Langue(s) maternelle(s)
1	ST	18	F	étudiante L2 SDL	français
2	LM	19	F	étudiante L2 SDL	français
3	JG	19	F	étudiante L2 SDL	français
4	PW	20	F	étudiante L2 SDL	français
5	AT	20	F	étudiante L3 SDL	français
6	JP	20	F	étudiante L2 SDL	français
7	NC	20	F	étudiante L3 lettres modernes	français
8	AL	20	F	étudiante L2 SDL	français
9	CV	20	F	étudiante L2 SDL	français
10	NL	21	F	étudiante L2 SDL	français
11	CM	21	F	étudiante L2 SDL	français
12	HH	21	F	étudiante L2 SDL	français, hébreu
13	VP	21	F	étudiante L2 SDL	français
14	EM	21	F	étudiante M1 SDL	français
15	IM	22	F	étudiante L2 SDL	français, lingala
16	AG	24	F	étudiante L2 SDL	français, créole
17	GG	46	F	étudiante M1 SDL	français
18	YT	20	H	étudiant L2 LLCE	français
19	CG	22	H	étudiant M2 multimédia	français
20	JV	23	H	étudiant M2 multimédia	français
21	EB	26	H	étudiant M2 SDL	français
22	AG	28	H	étudiant M2 droit	français
23	VP	32	H	technicien informatique	français
24	BP	34	H	adjoint administratif	français
25	PS	48	H	officier de marine	français

ANNEXE 4 : Participants à l'expérience de détection de mots

Tableau 57 – Informations détaillées sur les participants à l'expérience de détection de mots.

	Initiales	Age	Genre	Profession	Langue(s) maternelle(s)	Groupe ²⁵⁴
1	ST	18	F	étudiante L1 LEA	français	1
2	IN	19	F	étudiante L1 LEA	français	2
3	CSA	20	F	étudiante L3 SDL	français	2
4	SI	20	F	étudiante L2 info-com	français	2
5	MB	20	F	étudiante L2 info-com	français	1
6	CF	20	F	étudiante L2 info-com	français	2
7	EM	20	F	étudiante L3 SDL	français	2
8	SS	21	F	étudiante M1 multimédia	français	2
9	PX	22	F	étudiante M1 multimédia	français	1
10	MC	23	F	étudiante M1 SDL	français	2
11	JC	24	F	étudiante L1 LLCE	français	2
12	YZ	24	F	étudiante M1 SDL	français	2
13	ZB	24	F	étudiante M1 SDL	français	1
14	LA	26	F	étudiante M1 SDL	français, berbère	1
15	ER	26	F	étudiante DOC SDL	français	1
16	AT	30	F	bibliothécaire	français	1
17	RT	44	F	étudiante M1 SDL	français, arabe	2
18	ML	21	H	étudiant M1 multimédia	français	1
19	SM	23	H	étudiant M1 multimédia	français	1
20	MP	30	H	instituteur	français	1
21	VP	32	H	technicien informatique	français	1
22	BP	33	H	adjoint administratif	français	1
23	VM	42	H	enseignant en informatique	français	1
24	JLH	43	H	informaticien	français	2
25	JYD	65	H	enseignant-chercheur en linguistique	français	2

²⁵⁴ Le groupe 1 s'est vu présenter les blocs de stimuli dans l'ordre 1, 2, 3, 4, tandis que l'ordre 3, 4, 1, 2 a été utilisé pour le groupe 2. Voir section 1.6 du chapitre 4.

ANNEXE 5 : Scripts Praat programmés et utilisés pour l'analyse acoustique

- **Script utilisé pour les voyelles et les mots entiers** : mesure F0 moyen et durée.

```
for i from x to y
select i
soundname$ = selected$("Sound")
To Pitch... 0 60 350
f0 = Get mean... 0 0 Hertz
select i
duree = Get total duration
resultfile$ = "D:\analyse_F0_durée.txt"
resultline$ = "soundname$'      'f0:0'      'duree:3'      'newline$'"
fileappend "resultfile$" "resultline$"
endfor
```

Ce script est librement utilisable et téléchargeable au format texte à l'adresse suivante :
<http://erwan.pepiot.free.fr/doctorat/scriptpraat1.txt>

- **Script utilisé pour les consonnes** : mesure pic spectral, centre de gravité, F0 moyen et durée.

```
for i from x to y
select i
soundname$ = selected$("Sound")
To Pitch... 0 60 350
f0 = Get mean... 0 0 Hertz
select i
To Ltas... 100
spectral_peak = Get frequency of maximum... 700 15000 None
select i
To Spectrum... yes
cog = Get centre of gravity... 2
select i
duree = Get total duration
resultfile$ = "D:\analyse_COG_spectral_peak_F0_durée.txt"
resultline$ = "soundname$'      'spectral_peak'      'cog:0'      'f0:0'      'duree:3'      'newline$'"
fileappend "resultfile$" "resultline$"
endfor
```

Ce script est librement utilisable et téléchargeable au format texte à l'adresse suivante :
<http://erwan.pepiot.free.fr/doctorat/scriptpraat2.txt>

ANNEXE 6 : Scripts utilisé pour les expériences d'identification du genre par la voix dans le logiciel *Lancelot*

```
[INFORMATION]
TITLE=Expérience d'identification du genre par la voix
DATE=12/10/2011
AUTHOR=Erwan Pépiot
VERSION=2.0
[TRIAL_DATA]255
FIELDS=<wave><genre>
TRIAL1=<chapi_WUFFRC1.wav><Femme>
TRIAL2=<chapi_WUHFR1.wav><Homme>
[...]
TRIAL7=<chapi_F1FRC1.wav><Femme>
TRIAL8=<chapi_F2FRC1.wav><Femme>
[...]
TRIAL295=<toupi F1FR.wav><Femme>
TRIAL296=<toupi F2FR.wav><Femme>
TRIAL297=<toupi H1FR.wav><Homme>
TRIAL298=<toupi H2FR.wav><Homme>
[...]
TRIAL318=<oupi H2FR.wav><Homme>
[TRIAL_TEMPLATE]
T100 = SHOW_TEMPLATE<TemplateBlank.htm>
T105 = PLAY_SOUND<bip.wav>
T110 = WAIT <900>
T300 = PLAY_SOUND<#wave>
T310 = START_CHRONO
T330 = WAIT <1500>
T350 = PLAY_SOUND<#wave>
T355 = WAIT <500>
T360 = SHOW_TEMPLATE<TemplateExperience.htm>
T500 = GET_RESPONSES<RESPONSE_BLOCK 1-2><NON_RESPONSE_AUTHORISED no>
T850 = WAIT <300>
T900 = NEXT_TRIAL_AUTO
[SETTINGS_GROUP1]
INSTRUCTION_FORMAT=<Instructions.htm>
FINAL_MESSAGE_FORMAT=<MessageFinal.htm>
TRAINING_ORDER=<RANDOM 1-2><RANDOM 3-4><RANDOM 5-6>
TRIAL_ORDER=<RANDOM 7-102><RANDOM 103-114><RANDOM 115-210><RANDOM 211-318>
RESPONSE_FORMAT=<${SUBJECT}><${TRIAL}><#wave><#genre><${RESPONSE1}><${RESPONSE2}>
<${RTIME}><${RESP_STATUS1}>
RESPONSE_ITEM1=<ITEM genre><VALUES Homme Femme><CORRECT #genre>
RESPONSE_ITEM2=<ITEM certitude><VALUES 0 1 2 3 4 5 6 7>
TRIAL_COMMANDS=<TRIAL_TEMPLATE>
```

Les scripts (version anglaise et française) au format LSC, directement utilisables dans le logiciel *Lancelot* et accompagnés des fichiers HTML associés, sont téléchargeables à l'adresse suivante :

- <http://erwan.pepiot.free.fr/doctorat/scriptsidentdugenre.zip>

²⁵⁵ Les stimuli mentionnés dans cette version du script sont les stimuli français.

ANNEXE 7 : Script utilisé pour l'expérience de détection de mots dans le logiciel *Perceval*

```
[INFORMATION]
TITLE=Détection de mots
DATE=01/04/2011
AUTHOR=Erwan Pépiot
VERSION=3. 0. 5. 0
[TRIAL_DATA]
FIELDS=<Nom-serie><Fichiers-sons><Duree-pre-MC>
TRIAL1=<Warm up 1><serie1.wav><NC>
TRIAL2=<Warm up 2><serie2.wav><NC>
TRIAL3=<Warm up 3><serie3.wav><NC>
TRIAL4=<Warm up 4><serie4.wav><NC>
TRIAL5=<Warm up 5><serie5.wav><NC>
TRIAL6=<Warm up 6><serie6.wav><NC>
TRIAL7=<Distracteur 1><serie7.wav><NC>
TRIAL8=<Distracteur 2><serie8.wav><NC>
TRIAL9=<Pré XP><serie9.wav><NC>
TRIAL10=<XP COND A><serie10.wav><4357>
TRIAL11=<Distracteur 1><serie11.wav><NC>
TRIAL12=<Distracteur 2><serie12.wav><NC>
TRIAL13=<Pré XP><serie13.wav><NC>
TRIAL14=<XP COND B><serie14.wav><4802>
[...]
TRIAL67=<Distracteur 1><serie67.wav><NC>
TRIAL68=<Distracteur 2><serie68.wav><NC>
TRIAL69=<Pré XP><serie69.wav><NC>
TRIAL70=<XP COND B><serie70.wav><4724>
[TRIAL_EVENTS]
X10=BEGIN
X20=PLAY_SOUND<#Fichiers-sons>
X30=GET_INPUT
X50=END
[SETTINGS_GROUP1]
TRAINING_ORDER=<1-6>
TRIAL_ORDER=<7-70>
INPUT=<Mot VK_SPACE BK_01>
RESPONSE_FORMAT=<#Nom-serie><#Fichiers-sons><#Duree-pre-
MC><$RTIME><$SUBJECT><$TITLE><$GROUP><$WAVEDURATION><$DATE_SUBJECT>
PAUSE=600
[SETTINGS_GROUP2]
TRAINING_ORDER=<1-6>
TRIAL_ORDER=<39-70><7-38>
INPUT=<Mot VK_SPACE BK_01>
RESPONSE_FORMAT=<#Nom-serie><#Fichiers-sons><#Duree-pre-
MC><$RTIME><$SUBJECT><$TITLE><$GROUP><$WAVEDURATION><$DATE_SUBJECT>
PAUSE=600
```

Le script au format SRP, directement utilisable dans le logiciel *Perceval* et accompagné du fichier texte associé, est téléchargeable à l'adresse suivante :

- <http://erwan.pepiot.free.fr/doctorat/scriptdetectiondemots.zip>

ANNEXE 8 : Téléchargement des résultats détaillés de l'analyse acoustique

Les résultats détaillés (item par item) de l'analyse acoustique réalisée dans le chapitre 2 ont été regroupés dans des classeurs au format *Excel* téléchargeables aux adresses suivantes :

- Résultats de l'analyse des productions des **locuteurs anglophones américains du nord-est des Etats-Unis** :

<http://erwan.pepiot.free.fr/doctorat/analyseacoustiqueresultatsAN.xls>

- Résultats de l'analyse des productions des **locuteurs francophones parisiens** :

<http://erwan.pepiot.free.fr/doctorat/analyseacoustiqueresultatsFR.xls>

Ces fichiers contiennent les résultats des mesures suivantes : pic spectral des consonnes initiales, centre de gravité des consonnes initiales, *VOT* des consonnes initiales, fréquence fondamentale moyenne des consonnes initiales, durée des consonnes initiales, valeurs formantiques (F1, F2 et F3) des voyelles en syllabe initiale, différence H1-H2 sur les voyelles ouvertes en syllabe initiale, fréquence fondamentale moyenne des voyelles en syllabe initiale, durée des voyelles en syllabe initiale, fréquence fondamentale moyenne des mots dissyllabiques entiers, plage de variation de F0 sur les mots dissyllabiques entiers (en hertz et demi-tons), durée des mots dissyllabiques entiers. La police de caractères *Doulos SIL*²⁵⁶ a été utilisée pour représenter les symboles phonétiques dans ces fichiers.

²⁵⁶ Téléchargeable à cette adresse :

http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=DoulosSIL_download

ANNEXE 9 : Téléchargement des résultats détaillés des expériences d'identification du genre et de détection de mots

Les résultats détaillés des deux **expériences d'identification du genre par la voix réalisées dans le chapitre 3** ont été regroupés dans des classeurs au format *Excel* téléchargeables aux adresses suivantes :

- Résultats de l'expérience menée sur les auditeurs anglophones américains :
<http://erwan.pepiot.free.fr/doctorat/identdugenresultatsAN.xls>
- Résultats de l'expérience menée sur les auditeurs francophones parisiens :
<http://erwan.pepiot.free.fr/doctorat/identdugenresultatsFR.xls>

Ces deux fichiers sont chacun composés de deux feuilles : l'une présente les résultats détaillés pour chaque stimulus et pour chaque participant, l'autre contient un *tableau croisé dynamique*, permettant une bonne lisibilité dans les résultats. Ce dernier est entièrement paramétrable par l'utilisateur en fonction du type de résultats souhaité (par type de stimuli, par type de voix, etc.).

Les résultats détaillés de l'**expérience de détection de mots, présentée dans le chapitre 4**, sont également téléchargeables au format *Excel* à l'adresse suivante :

- <http://erwan.pepiot.free.fr/doctorat/detectiondemotsresultats.xls>

Ce fichier contient les temps de réponses de chaque participant pour les différents items expérimentaux.