A propos du rôle des formants vocaliques et du f0 moyen dans l'identification du genre par la voix chez les auditeurs francophones parisiens et anglophones américains

Erwan Pépiot

A propos du rôle des formants vocaliques et du f_0 moyen dans l'identification du genre par la voix chez les auditeurs francophones parisiens et anglophones américains

Pépiot, Erwan

Université Paris 8 – EA 1569 « Transferts critiques et dynamique des savoirs » – Groupe LAPS
erwan.pepiot@free.fr

1 Introduction

Un grand nombre d'études ont mis en évidence l'existence de différences acoustiques entre les productions des locuteurs féminins et masculins. La fréquence fondamentale moyenne est communément considérée comme la différence majeure entre ces deux types de voix. Elle serait de l'ordre de 120 Hz chez les locuteurs masculins, et de 200 Hz chez les femmes (Takefuta, Jancosek & Brunt, 1972; Boë, Contini & Rakotofiringa, 1975). Ces chiffres varient sensiblement avec l'âge du locuteur (Pegoraro-Crook, 1988) et sont globalement inférieurs chez les sujets fumeurs (Gilbert & Weismer, 1974). La seconde différence majeure concerne les formants vocaliques. Ceux des voyelles produites par les locutrices tendent à être situés dans fréquences globalement plus élevées que ceux des voyelles prononcées par leurs homologues masculins (Peterson & Barney, 1952; Hillenbrand, Getty, Clark & Wheeler, 1995; Whiteside, 2001; Pépiot, 2013a, 2014). Par ailleurs, des différences femmes-hommes ont également été observées sur d'autres paramètres acoustiques, tels que la plage de variation de f_0, qui semble plus large chez les locutrices (Takefuta, Jancosek & Brunt, 1972; Hann, 2002), le type de phonation, généralement plus soufflé chez les femmes (Klatt & Klatt, 1990; Henton, 1992; Hanson & Chuang, 1999), ou encore le débit de parole, légèrement plus élevé chez les locuteurs masculins (Byrd, 1994; Whiteside, 1996).

Certaines de ces variations inter-genres peuvent s'expliquer en partie par des différences anatomiques et physiologiques qui émergent à la puberté (Fant, 1966). Les plis vocaux s'allongent et s'épaississent de manière plus prononcée chez les sujets masculins (Kahane, 1978), d'où une tendance à des vibrations plus lentes, toutes choses égales par ailleurs. Un second point important concerne le conduit vocal. Sa longueur chez le locuteur adulte atteint environ 14,5 cm chez les femmes, contre 17 à 18 cm chez les hommes (Simpson, 2009) : cela permet d'expliquer, au moins partiellement, les différences inter-genre sur le plan des formants vocaliques.

Certains auteurs suggèrent que les voix de femmes et les voix d'hommes seraient, en raison de ces différences acoustiques, traitées de manière différente par les auditeurs (Sokhi, Hunter, Wilkinson & Woodruff, 2005 ; Pépiot, 2012, 2013b). Ainsi, dès qu'un auditeur est confronté à une voix, il tenterait de manière automatique et inconsciente d'identifier le genre du locuteur (Le Breton, 2011, Simpson, 2009). Cette identification serait effectuée à partir de représentations auditives des voix de femmes et d'hommes, présentes dans le cerveau des auditeurs (Mullennix, Johnson, Topcu-Dorgun & Farnsworth, 1995). Quels sont alors les indices acoustiques sur lesquels les auditeurs fondent leur jugement, et quelle est leur importance respective ?

Pour répondre à ces questions, plusieurs études testant la capacité des auditeurs à identifier le genre d'un locuteur à partir de sa voix ont été réalisées. Différents types de stimuli ont été utilisés dans ces expériences : des fricatives non voisées (Schwartz, 1968; Ingemann, 1968; Whiteside, 1998b), des voyelles (Pausewang Gelfer & Mikos, 2005; Lass, Hughes, Bowyer, Waters & Bourne, 1976; Whiteside; 1998a, 1998c), des syllabes (Bédard et Belin, 2004) et des phrases (Pépiot, 2011; Arnold, 2012). Il

s'agissait dans certains cas de stimuli synthétiques (Whiteside, 1998c; Pausewang Gelfer & Mikos, 2005) ou resynthétisés (Pépiot, 2011; Arnold, 2012). Sur les voyelles isolées, les pourcentages d'identifications réussies varient de 96 % (Lass & al., 1976) à 98.9 % (Whiteside, 1998a). Ils atteignent pratiquement 100 % avec des phrases d'une douzaine de syllabes (Pépiot, 2011). Notons cependant que ces études ont été réalisées avec des méthodes différentes et à plusieurs années voire décennies d'intervalle : il convient donc d'interpréter ces comparaisons avec prudence. L'utilisation du paradigme du *gating*, dans lequel les items sont présentés par segments de durée croissante (Grosjean, 1980), pourrait ainsi apporter des données intéressantes dans le cadre d'une expérience d'identification du genre par la voix.

L'influence relative des paramètres acoustiques dans l'identification du genre par la voix fait quant à lui l'objet de débats. Plusieurs études ont mis en évidence l'importance des formants vocaliques (Schwartz & Rine, 1968; Coleman, 1971; Arnold, 2012), tout comme celle du f_0 moyen (Coleman, 1976, Whiteside, 1998c, Pépiot, 2011). Selon la majorité des auteurs ayant conduit leurs études sur des auditeurs anglophones, la fréquence fondamentale moyenne serait l'indice acoustique le plus important. Pausewang Gelfer & Mikos (2005) ont présenté des voyelles synthétisées à des auditeurs anglophones américains dans deux conditions différentes : timbre vocalique et f_0 cohérents (e.g. valeurs formantiques d'hommes, f_0 à 120 Hz) ou contradictoires (formants vocaliques de femmes avec f_0 à 120 Hz ou valeurs formantiques d'hommes avec f_0 à 240 Hz). Lorsque les auditeurs ont été confrontés à ces stimuli contradictoires, ces derniers se sont principalement basés sur la fréquence fondamentale pour définir le genre du locuteur. Quand des formants vocaliques "féminins" étaient couplés avec un f_0 "masculin", les auditeurs ont identifié une voix d'homme dans près de 80 % des cas, et inversement. Ces donnés confirment des résultats obtenus antérieurement (Coleman, 1976; Lass, Hughes, Bowyer, Waters & Bourne, 1976; Whiteside, 1998c).

Néanmoins, une étude plus récente d'Arnold (2012) sur la contribution relative du f_0 moyen et des fréquences de résonance dans l'identification du genre suggère que ces dernières seraient le paramètre acoustique le plus déterminant. Des stimuli acoustiques (phrases resynthétisées) présentant, toutes choses égales par ailleurs, des fréquences fondamentales moyennes et des fréquences de résonance 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 se sont révélées non significatives. L'auteur en conclut que les fréquences de résonance, et en particulier les formants vocaliques, constituent un indice plus puissant que la fréquence fondamentale moyenne pour identifier le genre par la voix.

Les résultats obtenus dans cette étude d'Arnold semblent donc en contradiction avec ceux de Pausewang Gelfer & Mikos (2005) et de Lass et al. (1976). Remarquons par ailleurs que contrairement à Pausewang Gelfer & Mikos (2005) et Lass et al. (1976), qui utilisaient des stimuli de type « voyelle isolée », Arnold (2012) 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 des participants. Pausewang Gelfer & Mikos (2005) et Lass et al. (1976) ont travaillé avec des auditeurs (et locuteurs) anglophones américains, alors que l'étude d'Arnold (2008) a été menée sur des francophones. Or il est établi que certaines différences acoustiques inter-genres peuvent varier en fonction de la langue parlée (Johnson, 2005). Il est donc tout à fait probable que les stratégies pour identifier le genre par la voix soient également dépendantes de la langue. C'est d'ailleurs ce que semble indiquer une étude de Pépiot (2011), portant sur le rôle joué par le f_0 moyen chez les auditeurs anglophones américains et francophones. Des phrases ont été diffusées aux auditeurs avec différents niveaux de fréquence fondamentale moyenne : f_0 naturel, f_0 ambigu (169 Hz) et f_0 du genre opposé (129 Hz pour les femmes et 209 Hz pour les hommes). Il apparaît que les anglophones américains ont été plus sensibles à ces changements de f_0 moyen que leurs homologues francophones. Cependant, il est important de souligner le fait que les stimuli utilisés pour cette étude sont eux aussi resynthétisés : la

qualité et la naturalité de ces derniers a donc pu affecter les résultats. De plus, le rôle joué par les formants vocaliques n'a pas été testé.

Par conséquent, il semble pertinent de conduire une expérience d'identification du genre par la voix avec des stimuli non-resynthétisés, auprès d'auditeurs francophones parisiens et anglophones américains, dans le but d'établir avec plus de certitude le rôle respectif joué par le f_0 moyen et les formants vocaliques. L'hypothèse de départ est la suivante : le f_0 moyen est le paramètre le plus important pour identifier le genre par la voix chez les anglophones américains, mais les auditeurs francophones privilégient quant à eux la fréquence des formants vocaliques.

2 Méthode

2.1 Matériau linguistique

Du matériau linguistique français et anglais était nécessaire pour entreprendre cette expérience. L'utilisation de mots et pseudo-mots dissyllabiques, a été retenue, car ce paradigme permet de tester un nombre important de combinaisons de phonèmes. De plus, ce type d'unité linguistique est parfaitement adapté pour un dévoilement progressif des items (*gating*), technique de présentation qui sera utilisée dans cette expérience.

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. Les combinaisons suivantes, de type CVCV ou VCV, 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 français (produits par des francophones) seront utilisés dans l'expérience conduite exclusivement sur des francophones.

2.2 Locuteurs pour création des stimuli

Huit locuteurs monolingues ont été enregistrés pour obtenir les stimuli de l'expérience. Quatre d'entre eux sont francophones parisiens (2 femmes et 2 hommes), les quatre autres étant anglophones américains du Nord-Est des Etats-Unis (2 femmes et 2 hommes). En plus de ces huit locuteurs, deux locuteurs supplémentaires dans chaque langue ont été enregistrés (un homme et une femme). Leurs productions sont utilisées dans l'expérience uniquement en tant qu'items d'entraînement. Ces douze locuteurs sont âgés de 23 à 40 ans, non-fumeurs et ne présentent aucun trouble de la parole.

2.3 Locuteurs pour obtention des stimuli

Afin d'homogénéiser les paramètres prosodiques sur les différents items 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 mots français, et « He said 'WORD' three times » pour les items en anglais, les mots ayant tous été écrits

sous forme orthographique. 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 d'une bande passante allant de 20 Hz à 40 kHz. Les enregistrements ont été effectués au format Wave.

2.4 Création des stimuli

Chacun des mots dissyllabiques a été extrait de son contexte « Il a dit 'MOT' deux fois » ou « He said 'WORD' three times ». Cette opération a été réalisée dans le logiciel *Praat*.

Comme annoncé précédemment, ces mots ont ensuite été segmentés en plusieurs éléments, qui permettront de présenter des extraits de longueur croissante à l'auditeur. Le choix du découpage des mots a donc été un élément crucial. La segmentation des mots en moments (consonne 1, consonne 1 plus voyelle 1, etc.) plutôt qu'en extraits de longueur brute équivalente a semblé plus pertinente pour 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, afin de bien contrôler la durée de l'expérience. Tout d'abord, il a paru tout à fait indispensable de présenter isolément la consonne initiale de mot (appelée 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. Cela constituera donc le premier point de segmentation, et implique la création de 96 stimuli pour chaque langue (24 mots x 4 locuteurs). Pour le deuxième palier, une segmentation après la première voyelle (V1) a été choisie. Ainsi, pour les 24 mots de type CVCV, cela permettra d'évaluer l'effet de l'ajout d'une voyelle à la consonne 1. De plus, grâce aux trois mots de type VCV, ce point de segmentation permet également de tester l'identification du genre sur des voyelles isolées. Cette deuxième segmentation implique la création de 108 stimuli supplémentaires pour chaque langue (27 mots x 4 locuteurs). Enfin, 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.

Pour obtenir des 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 (voir section 2.3), 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.

2.5 Analyse acoustique préalable

Les stimuli de l'expérience ont fait l'objet préalable d'une analyse acoustique réalisée avec le logiciel *Praat*, dont les résultats ne seront pas détaillés ici (voir Pépiot, 2013a, 2014). La fréquence des trois premiers formants vocaliques de la V1 ainsi que la fréquence fondamentale moyenne des consonnes initiales voisées, de la V1 et des mots entiers ont notamment été mesurées : ces données seront utilisées pour effectuer des corrélations avec les réponses des auditeurs (catégorisations femme / homme et degrés de certitude).

2.6 Auditeurs

Les auditeurs ayant pris part à l'expérience se divisent 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 dit « parisien »), n'ayant pas l'anglais comme autre langue maternelle. Ils ne présentent

aucun trouble du langage ou de l'audition. La composition des deux groupes de participants est la suivante :

- Groupe des francophones natifs : 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 : 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).

2.7 Procédure expérimentale

La tâche d'identification du genre par la voix (voir Lass et al., 1976 ; Pausewang Gelfer & Mikos, 2005 ; Pépiot, 2011) 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. La technique retenue ici pour la diffusion des stimuli est celle du dévoilement progressif, ou *gating* (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 ([k], [t], [s], etc.).
- 12 items contenant uniquement la V1. 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 consonne), viennent ensuite les extraits contenant une consonne suivie d'une voyelle, et enfin les mots complets. Il a paru logique de procéder de la sorte afin de bien respecter l'idée de dévoilement progressif d'informations : si la présentation successive de chaque mot aux trois points de segmentation ([s], [sa], [sapi] ; [k], [ka], [kapi], etc.) avait été adoptée, l'auditeur aurait pu obtenir rapidement des informations suffisantes pour reconnaître les différentes voix utilisées dans l'expérience, ce qui aurait potentiellement biaisé les résultats.

L'expérience a été réalisée à l'aide d'un ordinateur portable et du logiciel *Perceval* 3.0.5.0. La mise en place de l'expérience dans ce logiciel a nécessité la programmation d'un script dédié. 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.

Les passations se sont déroulées dans une pièce calme, en suivant une procédure rigoureusement identique pour chaque participant. Avant de débuter l'expérience, plusieurs informations sur l'auditeur ont été 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 audio. Les directives suivantes étaient alors affichées, centrées au milieu à l'écran (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 ».

Après s'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. 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 ms, deuxième présentation du stimulus.
- A la fin de cette deuxième présentation, le formulaire permettant de collecter les réponses du participant apparaît.
- Le participant doit alors répondre à la question « Selon vous… La voix que vous avez entendue était-elle une voix de femme ou une voix d'homme ? », 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 (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. » était affiché. A cet instant, si le participant rencontrait une difficulté particulière, il avait la possibilité de demander des précisions à l'expérimentateur assis à proximité. Une fois les éventuels éclaircissements apportés, le participant était invité à continuer l'expérience, en sachant qu'il ne pourrait plus poser de question avant la fin de celle-ci.

Enfin, après la présentation des 312 items expérimentaux, un message de remerciement s'affichait à l'écran. La durée totale de l'expérience était d'environ 35 minutes, ce temps variant sensiblement d'un participant à l'autre en fonction du rythme de réponse adopté par ce dernier.

3 Analyse des données

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.

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.

4 Résultats

4.1 Auditeurs francophones

Les pourcentages d'identifications réussies ainsi que les degrés de certitude moyens (exprimés sur une échelle allant de 0 à 7) obtenus sur les auditeurs francophones pour chacune des cinq grandes catégories de stimuli sont présentés dans le tableau 1, ci-après.

Tableau 1 – 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.

Type de stimulus	**N d'items**	**N de réponses correctes**	**N de réponses incorrectes**	**Pourcentage d'identifications réussies**	**Degré de certitude moyen**
C initiale non voisée	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 non voisée (supérieur à 65 %), en dépit d'un degré de certitude relativement bas (3,56). Lorsque l'indice supplémentaire que constitue le f_0 apparaît (consonnes initiales voisées), le score augmente fortement, pour passer à 94,5 %. En revanche, le degré de certitude moyen s'améliore dans une 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.

Afin de vérifier si ces tendances sont significatives, une ANOVA à un facteur (« type de stimulus ») a été conduite sur les pourcentages d'identifications correctes obtenus par les auditeurs francophones lors de l'expérience d'identification du genre par la voix. 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 réussies ($F(4,7795)=562,081$; $p<0,0001$). 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 non voisées ($p<0,0001$), plus élevé sur les voyelles isolées que sur les consonnes voisées ($p<0,02$), similaire sur les voyelles 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 non voisées ($p<0,0001$), 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 ($p<0,0001$).

Pour estimer le rôle joué par la fréquence fondamentale moyenne sur les jugements des auditeurs, plusieurs tests de Pearson ont été conduits. Les résultats de ces tests sont présentés dans le tableau 2, ci-dessous.

Tableau 2 – Résultats des tests de Pearson pour le f_0 moyen des stimuli et les scores (pourcentage de bonne réponse et degré de certitude moyen) obtenus par les auditeurs francophones, en fonction du type de stimulus.

CORRELATIONS DE PEARSON POUR LE F_0 MOYEN (VARIABLE X) CHEZ LES AUDITEURS FRANCOPHONES		
Variable Y : *pourcentage d'identifications correctes*		
Consonnes initiales	Voix de femmes	$r(24) = 0,141$; $z = 0,652$; $p > 0,5$
	Voix d'hommes	$r(24) = -0,183$; $z = -0,85$; $p > 0,3$
Variable Y : *degré de certitude moyen*		
Consonnes initiales	Voix de femmes	$r(24) = 0,147$; $z = 0,679$; $p > 0,4$
	Voix d'hommes	$r(24) = -0,061$; $z = -0,281$; $p > 0,7$
Combinaisons CV	Voix de femmes	$r(48) = -0,187$; $z = -1,269$; $p > 0,2$
	Voix d'hommes	$r(48) = -0,426$; $z = -3,053$; $p < 0,01$*
Mots entiers	Voix de femmes	$r(54) = 0,273$; $z = 1,997$; $p < 0,05$*
	Voix d'hommes	$r(54) = 0,007$; $z = 0,052$; $p > 0,9$

On constate qu'il existe une corrélation négative significative entre le f_0 moyen des combinaisons CV et les degrés de certitude exprimés pour les voix d'hommes : plus le f_0 était bas plus les auditeurs ont été sûrs de leur choix. Une corrélation positive significative a quant à elle été observée entre le f_0 moyen des mots entiers et les degrés de certitude exprimés pour les voix de femmes. Aucune autre corrélation significative n'a été détectée. Cela suggère que la fréquence fondamentale moyenne joue bien un rôle dans l'identification du genre par la voix chez les auditeurs francophones, mais ce dernier semble relativement limité.

Des tests de Pearson ont également été conduits entre les valeurs des formants vocaliques et les scores des auditeurs. Ces tests ont été réalisés uniquement sur les séquences de type CV, en raison du faible nombre d'items de type « voyelle isolée ». Les pourcentages de bonnes réponses étant ici proches des 100 %, seuls les degrés de certitude des auditeurs ont été testés. Les résultats sont présentés dans le tableau 3, ci-après.

Tableau 3 – Résultats des tests de Pearson pour les valeurs des formants vocaliques et les degrés de certitude moyens exprimés par les auditeurs francophones sur les combinaisons CV.

CORRELATIONS DE PEARSON POUR LES FORMANTS VOCALIQUES SUR LES COMBINAISONS CV CHEZ LES AUDITEURS FRANCOPHONES	
Variable Y : *degré de certitude moyen*	
F1 (Hz) voyelle [i] – Voix de femmes	$r(16) = 0{,}062$; $z = -0{,}222$; $p > 0{,}8$
F2 (Hz) voyelle [i] – Voix de femmes	$r(16) = 0{,}262$; $z = 0{,}966$; $p > 0{,}3$
F3 (Hz) voyelle [i] – Voix de femmes	$r(16) = 0{,}076$; $z = 0{,}274$; $p > 0{,}7$
F1 (Hz) voyelle [i] – Voix d'hommes	$r(16) = 0{,}167$; $z = 0{,}608$; $p > 0{,}5$
F2 (Hz) voyelle [i] – Voix d'hommes	$r(16) = -0{,}569$; $z = -2{,}331$; $p < 0{,}02$*
F3 (Hz) voyelle [i] – Voix d'hommes	$r(16) = -0{,}335$; $z = -1{,}258$; $p > 0{,}2$
F1 (Hz) voyelle [a] – Voix de femmes	$r(16) = -0{,}271$; $z = -1{,}003$; $p > 0{,}3$
F2 (Hz) voyelle [a] – Voix de femmes	$r(16) = -0{,}121$; $z = -0{,}439$; $p > 0{,}6$
F3 (Hz) voyelle [a] – Voix de femmes	$r(16) = -0{,}228$; $z = -0{,}878$; $p > 0{,}3$
F1 (Hz) voyelle [a] – Voix d'hommes	$r(16) = 0{,}195$; $z = 0{,}714$; $p > 0{,}4$
F2 (Hz) voyelle [a] – Voix d'hommes	$r(16) = -0{,}491$; $z = -1{,}899$; $p < 0{,}05$*
F3 (Hz) voyelle [a] – Voix d'hommes	$r(16) = -0{,}371$; $z = -1{,}407$; $p > 0{,}1$
F1 (Hz) voyelle [u] – Voix de femmes	$r(16) = 0{,}634$; $z = 2{,}695$; $p < 0{,}01$*
F2 (Hz) voyelle [u] – Voix de femmes	$r(16) = 0{,}728$; $z = 3{,}331$; $p < 0{,}001$*
F3 (Hz) voyelle [u] – Voix de femmes	$r(16) = 0{,}759$; $z = 3{,}582$; $p < 0{,}001$*
F1 (Hz) voyelle [u] – Voix d'hommes	$r(16) = -0{,}426$; $z = -1{,}642$; $p > 0{,}1$
F2 (Hz) voyelle [u] – Voix d'hommes	$r(16) = -0{,}497$; $z = -1{,}927$; $p < 0{,}05$*
F3 (Hz) voyelle [u] – Voix d'hommes	$r(16) = -0{,}403$; $z = -1{,}541$; $p > 0{,}1$

De nombreuses corrélations significatives ont donc été détectées entre la fréquence des formants vocaliques et les degrés de certitude exprimés par les auditeurs francophones. Sur les voix d'hommes, ces corrélations sont, sans surprise, négatives (plus la fréquence du formant est bas, plus les auditeurs ont été sûrs de leur catégorisation) : elles apparaissent sur le F2 du [i], du [a] et du [u]. Pour les stimuli de type « voix de femmes », des corrélations positives ont été détectées sur les F1, F2 et F3 du [u]. Une tendance se dessine également sur le F2 du [i], mais cette corrélation n'atteint pas le seuil de significativité. Ces résultats suggèrent que la fréquence des formants vocaliques, et plus particulièrement du deuxième formant, jouent un rôle important dans le processus d'identification du genre par la voix chez les auditeurs francophones.

4.2 Auditeurs anglophones

Les résultats pour l'expérience menée sur les auditeurs anglophones américains figurent dans le tableau 4. Notons que chez les anglophones, les consonnes initiales /d/ et /g/ sont phonétiquement dévoisées : elles ont donc été classées dans la catégorie des consonnes initiales *non voisées*.

Tableau 4 – 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 non voisée	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 non voisées (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 %).

Des tests statistiques 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 du genre réussies.

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 non voisées ($p<0,0001$). 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 ($p<0,0001$ pour chaque type).

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 non voisées ($p<0,0001$), 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 ($p<0,001$).

A l'instar des auditeurs francophones, plusieurs tests de Pearson ont été conduits afin de tester le rôle joué par la fréquence fondamentale moyenne sur les jugements des auditeurs anglophones. Les résultats de ces tests sont présentés dans le tableau 5, ci-dessous.

Tableau 5 – Résultats des tests de Pearson pour le f_0 moyen des stimuli et les scores (pourcentage de bonne réponse et degré de certitude moyen) obtenus par les auditeurs anglophones américains, en fonction du type de stimulus.

CORRELATIONS DE PEARSON POUR LE F_0 MOYEN (VARIABLE X) CHEZ LES AUDITEURS ANGLOPHONES AMERICAINS		
Variable Y : *pourcentage d'identifications correctes*		
Consonnes initiales	Voix de femmes	$r(12) = 0,537$; $z = 1,801$; $p > 0,05$
	Voix d'hommes	Tous les scores à 100 %
Variable Y : *degré de certitude moyen*		
Consonnes initiales	Voix de femmes	$r(12) = 0,793$; $z = 3,237$; $p < 0,01$*
	Voix d'hommes	$r(12) = -0,625$; $z = -2,197$; $p < 0,05$*
Combinaisons CV	Voix de femmes	$r(48) = 0,005$; $z = 0,036$; $p > 0,9$
	Voix d'hommes	$r(48) = -0,341$; $z = -2,38$; $p < 0,02$*
Mots entiers	Voix de femmes	$r(54) = 0,587$; $z = 4,812$; $p < 0,0001$*
	Voix d'hommes	$r(54) = -0,137$; $z = -0,985$; $p > 0,3$

De fortes corrélations significatives ont été détectées. Sur les voix de femmes, ces corrélations sont positives et apparaissent sur les degrés de certitude exprimés pour les consonnes initiales et les mots entiers. Une tendance non significative se dessine également sur le pourcentage d'identifications correctes pour les consonnes initiales voisées. Pour les voix d'hommes, les corrélations sont négatives et ont été observées sur les degrés de certitude exprimés pour les consonnes initiales et les combinaisons « consonne + voyelle ». Ces nombreuses corrélations significatives suggèrent que le f_0 moyen constitue un indice particulièrement décisif chez les auditeurs anglophones américains pour identifier le genre du locuteur par la voix.

De la même manière, plusieurs tests de Pearson ont également été conduits entre les valeurs des formants vocaliques et les scores des auditeurs. Tout comme pour les francophones, ces tests ont été réalisés uniquement sur les séquences de type CV, avec les degrés de certitude des auditeurs. Les résultats sont visibles dans le tableau 6, ci-après.

Tableau 6 – Résultats des tests de Pearson pour les valeurs des formants vocaliques et les degrés de certitude moyens exprimés par les auditeurs anglophones américains sur les combinaisons CV.

CORRELATIONS DE PEARSON POUR LES FORMANTS VOCALIQUES SUR LES COMBINAISONS CV CHEZ LES AUDITEURS ANGLOPHONES AMERICAINS	
Variable Y : *degré de certitude moyen*	
F1 (Hz) voyelle [i:] – Voix de femmes	$r(16) = 0,372$; $z = 1,41$; $p > 0,1$
F2 (Hz) voyelle [i:] – Voix de femmes	$r(16) = -0,164$; $z = -0,596$; $p > 0,5$
F3 (Hz) voyelle [i:] – Voix de femmes	$r(16) = -0,328$; $z = -1,228$; $p > 0,2$
F1 (Hz) voyelle [i:] – Voix d'hommes	$r(16) = -0,82$; $z = -0,296$; $p > 0,7$
F2 (Hz) voyelle [i:] – Voix d'hommes	$r(16) = -0,138$; $z = -0,501$; $p > 0,6$
F3 (Hz) voyelle [i:] – Voix d'hommes	$r(16) = -0,138$; $z = -0,501$; $p > 0,6$
F1 (Hz) voyelle [æ] – Voix de femmes	$r(16) = 0,266$; $z = 0,984$; $p > 0,3$
F2 (Hz) voyelle [æ] – Voix de femmes	$r(16) = -0,221$; $z = -0,811$; $p > 0,4$
F3 (Hz) voyelle [æ] – Voix de femmes	$r(16) = 0,261$; $z = 0,962$; $p > 0,3$
F1 (Hz) voyelle [æ] – Voix d'hommes	$r(16) = -0,369$; $z = -1,398$; $p > 0,1$
F2 (Hz) voyelle [æ] – Voix d'hommes	$r(16) = -0,058$; $z = -0,208$; $p > 0,8$
F3 (Hz) voyelle [æ] – Voix d'hommes	$r(16) = -0,094$; $z = -0,339$; $p > 0,7$
F1 (Hz) voyelle [u:] – Voix de femmes	$r(16) = 0,220$; $z = 0,807$; $p > 0,4$
F2 (Hz) voyelle [u:] – Voix de femmes	$r(16) = 0,649$; $z = 2,788$; $p < 0,01$*
F3 (Hz) voyelle [u:] – Voix de femmes	$r(16) = 0,454$; $z = 1,767$; $p > 0,05$
F1 (Hz) voyelle [u:] – Voix d'hommes	$r(16) = -0,397$; $z = -1,514$; $p > 0,1$
F2 (Hz) voyelle [u:] – Voix d'hommes	$r(16) = -0,019$; $z = -0,068$; $p > 0,9$
F3 (Hz) voyelle [u:] – Voix d'hommes	$r(16) = -0,111$; $z = -0,402$; $p > 0,6$

Contrairement aux observations faites sur les données des auditeurs francophones, ces tests s'avèrent assez peu concluants. Une seule corrélation significative a été détectée : elle concerne le F2 de la voyelle [u:] pour les stimuli de type « voix de femmes » (corrélation positive). Une autre tendance non significative apparaît également sur F1 du [æ] pour les voix d'hommes (corrélation négative). Ces résultats suggèrent que la fréquence des formants vocaliques joue un rôle limité dans l'identification du genre par les voix chez les auditeurs anglophones américains.

5 Conclusion - Discussion

L'expérience d'identification du genre par la voix présentée ici se distingue 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. Ces particularités méthodologiques ont permis d'obtenir des résultats intéressants.

Tout d'abord, les auditeurs anglophones comme francophones sont parvenus à identifier le genre des locuteurs à partir de stimuli très courts, y compris des consonnes non voisées (plus de 65 % d'identifications réussies pour les deux groupes d'auditeurs). Cela confirme les résultats obtenus antérieurement par Schwartz (1968), Ingemann, 1968 et Whiteside (1998b).

D'autre part, l'hypothèse de départ selon laquelle le f_0 moyen serait le paramètre le plus important chez les anglophones américains, mais que les auditeurs francophones privilégieraient quant à eux la fréquence des formants vocaliques, semble largement vérifiée. En effet, de nombreuses et fortes corrélations significatives ont été observées entre le f_0 moyen des stimuli et les scores des auditeurs anglophones

américains, mais très peu de corrélations sont apparues avec la fréquence des formants vocaliques. Une tendance inverse est apparue pour les auditeurs francophones, pour qui la fréquence des formants vocaliques et en particulier celle de F2 semblent avoir joué un rôle particulièrement décisif.

Ces données vont clairement dans le sens des résultats obtenus par Pépiot (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 f_0 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, 2012).

Le rôle moindre 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, Pisoni & De Jong, 2005). D'autre part, les formants des voyelles de l'anglais américain sont nettement moins stables que ceux des voyelles du français. Cela est vrai non seulement pour les diphtongues, mais également pour certaines « monophtongues » longues, telles que le /i:/ et le /u:/, qui sont phonétiquement légèrement diphtongués (Pike, 1946). Cette instabilité formantique pourrait donc être une raison supplémentaire pour laquelle les auditeurs anglophones américains s'appuient moins sur cet indice acoustique que leurs homologues francophones lorsqu'ils tentent d'identifier le genre d'un locuteur par sa voix.

Références bibliographiques

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.

Bédard, C., & Belin, P. (2004). A "voice inversion effect?". *Brain and Cognition*, *55*, 247-249.

Boë, L-J., Contini, M., & Rakotofiringa, H. (1975). Étude statistique de la fréquence laryngienne. *Phonetica*, *32*, 1-23.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39-54.

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.

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.

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.

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.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, *28*, 267-283.

Haan, J. (2002). *Speaking of questions. An exploration of Dutch question intonation*. Utrecht, Pays-Bas : LOT.

Hanson, H., & Chuang, E. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustic Society of America*, *106*, 1064-77.

Henton, C. G. (1992). Sex and speech synthesis: techniques, successes, and challenges. *Proceedings of the Fourth Australian International Conference on Speech Science and Technology – Brisbane*, 738-743.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustic Society of America*, *97*, 3099-3111.

Ingemann, F. (1968). Identification of the speaker's sex from voiceless fricatives. *Journal of the Acoustic Society of America*, *44*, 1142-1144.

Johnson, K. (2005). Speaker normalization in speech perception. In Pisoni, D. & Remez, R. (Eds.). *The Handbook of Speech Perception* (pp. 363-389). Oxford, Royaume-Uni : Blackwell Publishers.

Kahane, J. (1978). A morphological study of the human prepubertal and pubertal larynx. *American Journal of Anatomy*, *151*, 11-20.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of the Acoustic Society of America*, *87*, 820-857.

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 Acoustic Society of America*, *59*, 675-678.

Le Breton, D. (2011). *Eclats de voix : une anthropologie des voix*. Paris : Editions Métailié.

Mullennix, J., Johnson, K., Topcu-Dorgun, M. & Farnsworth, L. (1995). The perceptual representation of voice gender. *Journal of the Acoustical Society of America*, *98*, 3080-3095.

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.

Pegoraro-Krook, M. I. (1988). Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia Phoniatrica*, *40*, 82-90.

Pépiot, 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*, 418-430.

Pépiot, E. (2012). Les temps de traitement des voix de femmes et d'hommes sont-ils équivalents ? *Actes des JEP-TALN-RECITAL 2012*, 153-160.

Pépiot, E. (2013a). *Voix de femmes, voix d'hommes : différences acoustiques, identification du genre par la voix et implications psycholinguistiques, chez les locuteurs anglophones et francophones* (316 p.). Thèse de doctorat. Université Paris 8.

Pépiot, E. (2013b). Processing male and female voices : a word spotting experiment. *Perceptual and Motor Skills*, *117*, 903-912.

Pépiot, E. (2014). 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.* (en cours de publication).

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the identification of vowels. *Journal of the Acoustic Society of America*, *24*, 175-184.

Pike, K. L. (1947). On the phonemic status of English diphthongs. *Language*, *23*, 151-159.

Schwartz, M. F. (1968). Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustic Society of America*, *43*, 1178-1179.

Schwartz, M. F., & Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustic Society of America*, *44*, 1736-1737.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*, 621-640.

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.

Takefuta, Y., Jancosek, E. G., & Brunt, M. (1972). A statistical analysis of melody curves in the intonation of American English. *Proceedings of the 7th International Congress of Phonetic Sciences - Montreal (1971)*, 1035-1039.

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 Acoustic Society of America*, *110*, 464-478.