

**Voix de femmes, voix d'hommes : à propos de  
l'identification du genre par la voix chez des auditeurs  
anglophones et francophones**

Erwan Pépiot

► **To cite this version:**

Erwan Pépiot. Voix de femmes, voix d'hommes : à propos de l'identification du genre par la voix chez des auditeurs anglophones et francophones. Scientific Works, Plovdiv : Visš selskostopanski institut, 2011, 49, pp.418-430. halshs-00763232

**HAL Id: halshs-00763232**

**<https://halshs.archives-ouvertes.fr/halshs-00763232>**

Submitted on 10 Dec 2012

**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.

# VOIX DE FEMMES, VOIX D'HOMMES : A PROPOS DE L'IDENTIFICATION DU GENRE PAR LA VOIX CHEZ DES AUDITEURS ANGLOPHONES ET FRANCOPHONES

Erwan Pépiot, Université Paris 8

## Abstract

This experimental study investigates the role played by different acoustic parameters on voice-based gender identification. A perceptual experiment was jointly conducted on French and American English hearers. The results show that intonation is an important clue for the French hearers but not for the American English participants, who are more sensitive to mean F0. This suggests that one's mother tongue plays a role in the strategies adopted to identify the speakers' gender from voice.

Keywords: phonetics, gender identification, male voices, female voices, fundamental frequency, intonation, cross-language differences.

## Introduction

La fréquence fondamentale moyenne est communément considérée comme étant la principale différence entre les voix d'hommes et les voix de femmes : elle se situerait respectivement aux alentours des 120 Hz et des 240 Hz (Vaissière 2006). Cependant, de nombreuses études ont mis en évidence d'autres différences : le timbre, c'est-à-dire la position des formants (Coleman 1976 ; Pépiot 2009), l'intonation<sup>1</sup> (Brend 1975), ou encore la position de la zone de bruit des fricatives (Schwartz 1968). Ces différences s'expliqueraient, tout du moins partiellement, par des éléments physiologiques : plis vocaux plus épais et conduit vocal plus long chez les locuteurs masculins (Fant 1966).

Le rôle joué par chacun de ces paramètres dans l'identification du genre<sup>2</sup> par la voix fait quant à lui l'objet de débats. Si la majorité des auteurs considère que la fréquence fondamentale moyenne est le principal paramètre (Pausewang Gelfer & Mikos 2005 ; Coleman 1976), quelques auteurs suggèrent qu'il s'agirait du timbre (Arnold 2008). Cette dernière étude a été menée sur des francophones, contrairement aux autres expériences de ce type, menées sur des anglophones. En outre, le rôle de l'intonation n'a à ma connaissance pas fait l'objet de recherches.

La plupart des études menées dans ce domaine semblent avoir largement sous-exploité un autre élément important : les différences inter-langues. Sur le plan acoustique, les différences de timbre entre les hommes et les femmes, par exemple, varient considérablement d'une langue à l'autre (Johnson 2005). Ces différences inter-langues pourraient permettre d'expliquer les résultats en apparence contradictoires mentionnés plus haut, sur le rôle respectif du timbre et de la fréquence fondamentale dans l'identification du genre par la voix.

---

<sup>1</sup> Le terme *intonation* désigne ici les variations de la courbe de F0 dans le temps.

<sup>2</sup> J'ai choisi d'utiliser ici le mot *genre*, traduction de l'anglais *gender*, plutôt que le mot *sexe* afin de souligner la dimension culturelle des différences entre hommes et femmes.

Dans le cadre de cette étude, portant sur l'identification du genre par la voix, j'ai souhaité travailler sur l'influence de ce facteur « langue », sur le rôle joué par la fréquence fondamentale moyenne (à différents niveaux), et sur celui joué par l'intonation. Pour cela, il m'a fallu mettre au point une expérience d'identification du genre par la voix, qui permette de tester tous ces paramètres. Mes hypothèses de départ sont les suivantes :

**H1** : *La fréquence fondamentale moyenne joue un rôle dans l'identification du genre par la voix.*

**H2** : *Cette dernière joue un rôle plus important chez les anglophones que chez les francophones.*

**H3** : *L'intonation joue un rôle dans l'identification du genre par la voix.*

L'intérêt d'une telle étude est multiple. D'une part, cela permettrait une meilleure compréhension des processus cognitifs entrant en jeu lors de l'identification du genre par la voix. D'autre part, ces données pourraient être largement exploitées dans d'autres disciplines, tels que la sociologie, et plus spécifiquement les études de genre. Enfin, on peut parfaitement envisager des applications pratiques dans le domaine du traitement automatique des langues.

## 1. Méthodologie

Dans cette section seront décrits le choix des matériaux linguistiques et des locuteurs, la procédure d'enregistrement et enfin les modifications acoustiques effectuées a posteriori pour obtenir les stimuli de l'expérience.

### 1.1 Matériau linguistique

Afin de mener à bien cette étude, du matériau linguistique en langue française et en langue anglaise était nécessaire. Pour plusieurs raisons, j'ai choisi d'opter pour des phrases (une dans chaque langue) plutôt que pour des mots ou des voyelles isolées. L'utilisation de ces unités n'aurait pas permis de tester l'influence de l'intonation sur l'identification du genre.

Le choix des phrases a été réalisé sur la base de plusieurs critères. Le premier de ces critères était la recherche d'une certaine neutralité émotionnelle. Le deuxième critère était d'ordre syntaxique : il fallait une phrase dotée d'une structure suffisamment riche pour permettre l'émergence potentielle de schémas prosodiques distincts selon les locuteurs. Pour ces raisons, j'ai retenu la structure [Sujet – Verbe – Objet – Complément] appliquée à une phrase affirmative. Un troisième critère, d'ordre phonétique, a été pris en considération : les différences formantiques hommes-femmes étant plus importantes sur les voyelles antérieures, la présence d'au moins trois de ces voyelles m'a paru indispensable. Enfin, le dernier critère a été de choisir deux phrases de longueur équivalente et sémantiquement proches. Sur la base de ces critères, les deux phrases suivantes ont été retenues :

- **Phrase française** : « Caroline a vendu son vélo cet après-midi. »

- **Phrase anglaise** : « Steven sold his bicycle yesterday afternoon. »

## 1.2 Choix des locuteurs pour les enregistrements

Pour ces enregistrements, j'ai fait appel à 10 participants, pour moitié francophones (locuteurs natifs du français parisien), et pour moitié anglophones (locuteurs natifs de l'anglais américain). Chacun des groupes est constitué de deux hommes et deux femmes, dont les enregistrements seront utilisés et pris en compte dans l'expérience, ainsi que d'un cinquième locuteur, homme ou femme, dont les enregistrements seront utilisés en tant qu'items d'échauffement. Voici la composition détaillée des deux groupes :

- **Locuteurs francophones** : « femme francophone 1 » (FF1), 25 ans ; « femme francophone 2 » (FF2), 27 ans ; « homme francophone 1 » (HF1), 23 ans ; « homme francophone 2 » (HF2): 33 ans ; « locutrice francophone pour *warm-up* » (WUF): 26 ans.
- **Locuteurs anglophones** : « femme anglophone 1 » (FA1), 49 ans ; « femme anglophone 2 » (FA2), 51 ans ; « homme anglophone 1 » (HA1), 20 ans ; « homme anglophone 2 (HA2), 55 ans ; « locuteur anglophone pour *warm-up* » (WUA): 19 ans.

## 1.3 Procédure d'enregistrement

Le matériel utilisé est un enregistreur numérique portable *Edirol R-09HR* de marque Roland, doté de deux microphones omnidirectionnels intégrés. Les enregistrements ont été effectués en chambre sourde, et sauvegardés en format Wave.

Voyons maintenant en détails les opérations de resynthèse effectuées sur les enregistrements afin d'obtenir les stimuli de l'expérience.

## 1.4 Modification acoustique des enregistrements pour l'obtention des stimuli

J'ai créé quatre séries de stimuli pour l'expérience. La première série est constituée des enregistrements bruts, sans aucune modification acoustique. Les trois autres séries ont été obtenues via un travail de resynthèse effectué sur les voix naturelles.

### Resynthèse du F0 moyen au même niveau pour les hommes et les femmes

Les stimuli de la deuxième session ont été créés en resynthétisant les F0 moyens de chaque locuteur (hommes et femmes) à un même niveau. Ce niveau intermédiaire se situe à 169 Hz pour mon échantillon composé de 5 hommes (moyenne : 129 Hz) et 5 femmes (moyenne : 209 Hz).

La perception de l'oreille humaine n'étant pas linéaire mais logarithmique (voir Martin, 2008), j'ai choisi d'utiliser des coefficients multiplicateurs plutôt qu'une élévation ou un abaissement absolu de toute la courbe F0. Après avoir calculé ce coefficient ( $F0 \text{ cible} / F0 \text{ d'origine}$ ) pour chaque locuteur, l'opération de resynthèse a été effectuée à l'aide du logiciel *Praat* et de sa fonction « *multiply pitch frequencies* ».

### Aplatissement de la courbe de F0

Pour obtenir les stimuli de la troisième session de l'expérience, j'ai dû aplatir la courbe de F0 de tous les locuteurs, au même F0 moyen que pour la session 2 (169 Hz). Cette opération de resynthèse a elle aussi été effectuée avec le logiciel *Praat*, via la fonction « *Set pitch point at...* ».

## **Permutation des F0 moyens des hommes et des femmes**

Les stimuli de la session 4 ont été créés en resynthétisant le F0 moyen des locuteurs masculins au niveau moyen des locuteurs féminins (209 Hz), et celui des locuteurs féminins au niveau moyen des locuteurs masculins (129 Hz). La procédure a été similaire à celle utilisée pour l'obtention des stimuli de la deuxième session.

### **1.5 Participants**

Les individus ayant pris part à l'expérience constituent un groupe d'anglophones natifs (n'ayant pas le français comme autre langue maternelle) et un groupe de francophones natifs (n'ayant pas l'anglais comme autre langue maternelle). Voici la composition des deux groupes :

- **Auditeurs francophones natifs** : 4 hommes, 28 femmes. Tous étudiants de première année de licence, et âgés de 18 à 25 ans. Moyenne d'âge : 19,3 ans (18,8 ans pour les hommes, 19,4 ans pour les femmes)
- **Auditeurs anglophones natifs** : 5 hommes, 11 femmes. Tous étudiants de première année, et âgés de 18 à 25 ans. Moyenne d'âge : 20,2 ans (20,8 ans pour les hommes, 19,8 ans pour les femmes)

### **1.6 Procédure expérimentale**

L'expérience s'est donc déroulée en quatre sessions. Les voix des dix mêmes locuteurs ont été utilisées, mais des modifications acoustiques de nature différente pour chaque session ont été effectuées sur elles. Voici la nature des stimuli, session par session :

- **Session 1** : Voix naturelles.
- **Session 2** : Voix resynthétisées au F0 moyen de 169 Hz.
- **Session 3** : Voix resynthétisées au F0 moyen de 169 Hz avec courbe de F0 aplatie.
- **Session 4** : Voix resynthétisées au F0 moyen de 209 Hz pour les locuteurs masculins, et au F0 moyen de 129 Hz pour les locutrices.

Chaque session a été espacée de la suivante d'au moins une semaine. Cette précaution a pour but d'éviter que les auditeurs ne se souviennent des voix diffusées d'une session à l'autre.

Une attention particulière a été apportée à l'ordre de présentation des stimuli. Ce dernier a été changé pour chaque session, de manière à ce que les auditeurs ne puissent le mémoriser. D'autres contraintes ont été appliquées pour éviter tout effet d'accommodation (voir Schweinberger & al. 2008) : pas plus de deux stimuli consécutifs dans la même langue et pas plus de deux stimuli de locuteurs du même genre (homme ou femme) d'affilée.

J'ai procédé à deux passations groupées : l'une pour les auditeurs francophones, l'autre pour les auditeurs anglophones. Les stimuli et leur ordre de diffusion étant strictement les mêmes pour les deux groupes, tout comme le déroulement de l'expérience.

Le déroulement a été la même pour les quatre sessions. Tout d'abord, faire lire la consigne écrite figurant sur le questionnaire distribué aux participants : « *Selon vous, la personne qui parle est-elle un homme ou une femme ? Cochez la case correspondant à*

votre réponse, et précisez aussi le degré de certitude de votre réponse, sur une échelle de 0 (pas sûr du tout) à 7 (tout à fait sûr) »<sup>3</sup>. J’ajoutais ensuite quelques précisions orales sur la consigne. Après m’être assuré que cette dernière avait été parfaitement comprise par tous les participants, j’ai commencé la diffusion des stimuli. Comme indiqué plus haut, chaque stimulus a été diffusé deux fois. Je laissais passer quelques secondes entre chaque stimulus pour que les auditeurs aient le temps de répondre. Les enceintes utilisées pour la diffusion des stimuli sont des *JLB Duet Aluminium 2.0* d’une puissance de 2x6 Watts RMS et disposant d’une bande passante allant de 40 Hz à 20 KHz.

La méthodologie a maintenant été exposée. Dans la section suivante, je décris les résultats de l’expérience.

## 2. Résultats

Dans un premier temps, les résultats sont présentés de manière brute et synthétique, session par session. Une deuxième partie est consacrée à leur analyse statistique.

### 2.1 Résultats bruts

Une synthèse des résultats obtenus pour les auditeurs anglophones et les auditeurs francophones est présentée dans le tableau ci-dessous.

**Tableau 1** – Synthèse des résultats de la session 1.

|                  | Auditeurs anglophones                            |                                       | Auditeurs francophones              |                          |
|------------------|--|---------------------------------------|-------------------------------------|--------------------------|
|                  | Pourcentage de réponses congruentes <sup>4</sup> | Degré de certitude moyen <sup>5</sup> | Pourcentage de réponses congruentes | Degré de certitude moyen |
| <b>Session 1</b> | 100,00   | 6,87                                  | 100,00                              | 6,87                     |
| <b>Session 2</b> | 94,17  | 6,64                                  | 97,40                               | 6,61                     |
| <b>Session 3</b> | 94,17  | 6,46                                  | 91,41                               | 5,78                     |
| <b>Session 4</b> | 82,50  | 6,16                                  | 88,24                               | 5,69                     |

Les résultats de la session 1 (voix naturelles) sont donc parfaitement clairs : les 328 jugements émis par les deux groupes réunis sont tous congruents. Tous les auditeurs sans exception, ont donc été à même d’identifier le genre de chaque locuteur à partir des voix naturelles, avec un degré de certitude extrêmement élevé. Quelques jugements non-congruents ont été émis dans la session 2 (F0 moyens égalisés au niveau intermédiaire de 169 Hz) mais le taux de réponses congruentes et les degrés de certitude demeurent élevés pour les deux groupes. Pour la troisième session (F0 moyens égalisés et courbes mélodiques aplaties), on remarque que le taux de réponses congruentes a baissé par rapport à la session 2 chez les francophones, mais pas chez les anglophones. Enfin, concernant la session 4 (F0 moyens des hommes et des femmes permutés -209 Hz et 129 Hz-), le taux de réponses congruentes et le degré de certitude moyen ont fortement baissé par rapport aux sessions précédentes pour les deux groupes d’auditeurs. Cette tendance est encore plus marquée chez les anglophones.

<sup>3</sup> La consigne était traduite en anglais pour le groupe des anglophones.

<sup>4</sup> Qui correspond au genre du locuteur ayant été enregistré pour la création du stimulus.

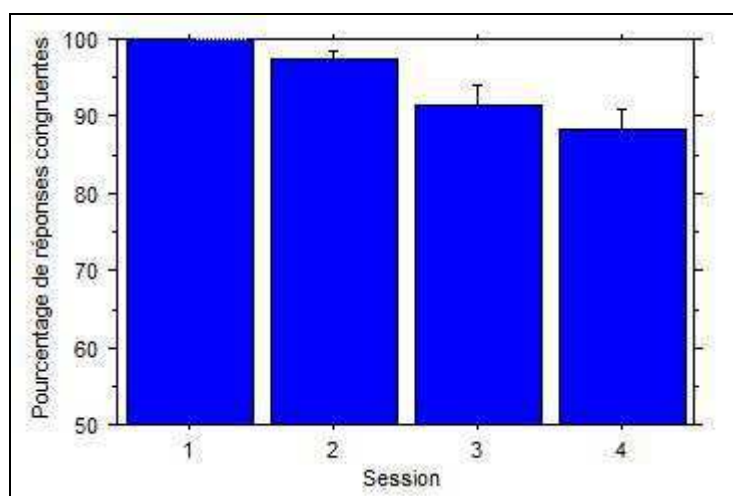
<sup>5</sup> Pour rappel, les degrés de certitude ont été exprimés sur une échelle allant de 0 à 7.

## 2.2 Analyse statistique

Afin d'établir si les différences observées (pour les pourcentages de réponses congruentes et les degrés de certitude) étaient significatives, j'ai procédé à diverses analyses statistiques à l'aide du logiciel *StatView*.

### Différences entre chaque session pour les auditeurs francophones

J'ai réalisé une ANOVA à un facteur sur les réponses données par les auditeurs francophones en fonction du facteur « session ». Le graphique correspondant à cette analyse est visible ci-dessous (figure 1). Les pourcentages de réponses congruentes sont présentés dans tous les graphiques avec un seuil minimum fixé à 50 %, car ce dernier correspond au seuil de chance (choix forcé entre deux items).



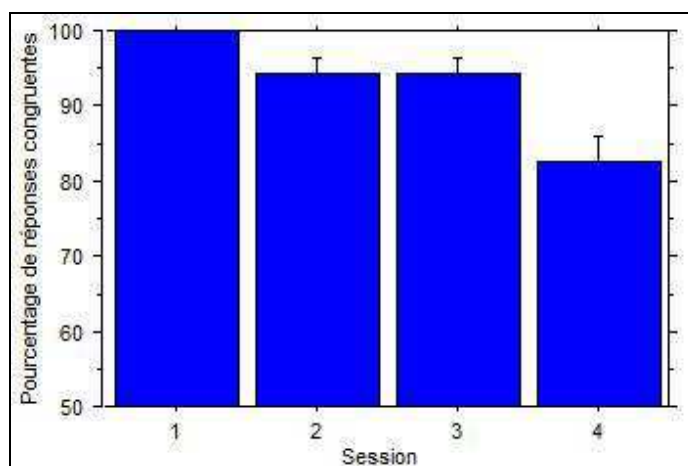
**Figure 1** – Graphique représentant le pourcentage de réponses congruentes en fonction de la session (1, 2, 3 et 4) pour les auditeurs francophones, avec les barres d'erreur-types.

Le résultat de l'ANOVA globale est le suivant : il existe un effet très significatif du facteur « session » sur le pourcentage de réponses congruentes données par les auditeurs, avec  $F(3,660)=10,72$  et  $p<0,001$ . Le test PLSD de Fisher pour les différentes sessions prises deux à deux met en évidence des différences significatives entre les sessions 1 et 3 ( $p<0,001$ ), 1 et 4 ( $p<0,001$ ), 2 et 3 ( $p<0,02$ ) et 2 et 4 ( $p<0,001$ ). Il y a donc bien eu un rôle important joué par l'intonation (seul paramètre modifié entre les sessions 2 et 3) sur la catégorisation « homme » / « femme » effectuée par les auditeurs francophones. De même, le F0 moyen a joué un rôle significatif lorsqu'il a été « permuté » (session 4), mais pas lorsqu'il a été resynthétisé au niveau intermédiaire de 169 Hz (session 2).

J'ai réalisé une analyse identique pour les degrés de certitude émis par les auditeurs francophones en fonction du facteur « session ». Le résultat de l'ANOVA est le suivant : il existe un effet globalement très significatif du facteur « session », avec  $F(3,660)=58,41$  et  $p<0,001$ . Le test PLSD de Fisher fait état de différences significatives entre les sessions 1 et 2 ( $p<0,01$ ), 1 et 3 ( $p<0,001$ ), 1 et 4 ( $p<0,001$ ), 2 et 3 ( $p<0,001$ ) et 2 et 4 ( $p<0,001$ ). Les auditeurs francophones ont donc été, de manière significative, moins sûrs de leurs réponses, d'une part lorsqu'ils étaient dépourvus de l'information véhiculée par l'intonation (différence entre les sessions 2 et 3), d'autre part lorsque le F0 moyen était ambigu (différence entre les sessions 1 et 2) et enfin lorsque le F0 moyen était permuté (différences entre les sessions 1 et 4, et entre les sessions 2 et 4).

## Différences entre chaque session pour les auditeurs anglophones

De la même façon que pour les auditeurs francophones, j'ai réalisé une ANOVA à un facteur sur les réponses données par les auditeurs anglophones (variable dépendante) en fonction du facteur « session ». Le graphique correspondant à cette analyse est visible ci-dessous (figure 2).



**Figure 2** – Graphique représentant le pourcentage de réponses congruentes en fonction de la session (1, 2, 3 et 4) pour les auditeurs anglophones, avec les barres d'erreur-types.

L'analyse a donné le résultat suivant : comme chez les francophones, il existe un effet globalement très significatif du facteur « session » sur le pourcentage de réponses congruentes données par les auditeurs, avec  $F(3,474)=9,99$  et  $p<0,0001$ . Le test PLSD de Fisher met en évidence des différences très significatives entre les sessions 1 et 4 ( $p<0,0001$ ), 2 et 4 ( $p<0,001$ ) et 3 et 4 ( $p<0,001$ ). Contrairement à ce qui a été observé chez les auditeurs francophones, l'intonation (seul paramètre modifié entre les sessions 2 et 3) n'a donc pas eu d'influence sur la catégorisation « homme » / « femme » opérée par les anglophones. En revanche, le F0 moyen a joué un rôle extrêmement important lorsqu'il a été « permuté » (session 4).

Une analyse similaire menée pour les degrés de certitude montre qu'il existe bien un effet globalement très significatif du facteur « session », avec  $F(3,474)=12,79$  et  $p<0,0001$ . Le test PLSD de Fisher révèle des différences significatives entre les sessions 1 et 2 ( $p<0,05$ ), 1 et 3 ( $p<0,001$ ), 1 et 4 ( $p<0,0001$ ), et 2 et 4 ( $p<0,0001$ ). Les auditeurs anglophones ont donc été, de manière significative, moins sûrs de leurs réponses lorsque le F0 moyen était ambigu (différence entre les sessions 1 et 2) et lorsque le F0 moyen était permuté (différences entre les sessions 1 et 4, ainsi que 2 et 4). En revanche, la privation de l'information véhiculée par l'intonation n'a pas eu d'influence significative sur les degrés de certitude exprimés (différence entre les sessions 2 et 3 non significative).

## Différences entre les pourcentages de réponses congruentes obtenus par les auditeurs francophones et ceux obtenus par les anglophones en fonction du stimulus

Afin de poursuivre la mise en évidence de stratégies différentes dans l'identification du genre entre le groupe d'auditeurs anglophones et celui des francophones, j'ai décidé de tester le rôle du facteur « groupe d'auditeurs » pour le pourcentage de réponses congruentes données en fonction du stimulus (c'est-à-dire des

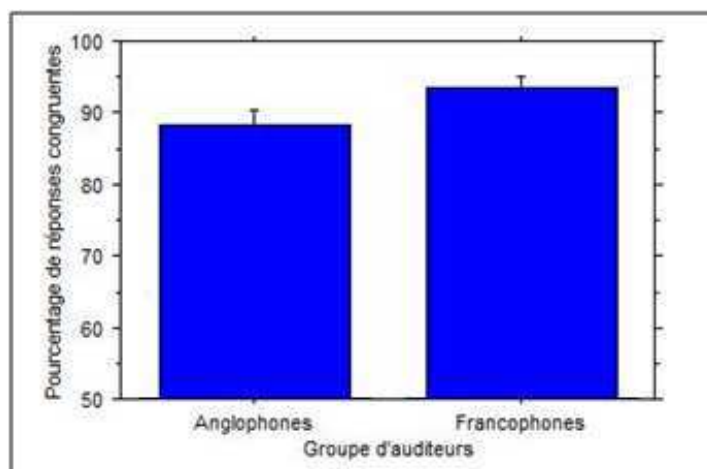


différentes voix), toutes sessions confondues. Pour cela j'ai effectué une analyse ANOVA à deux facteurs (« stimulus »\* « groupe d'auditeurs »).

Les résultats du test montrent qu'il existe bien une très forte interaction entre les facteurs « stimulus » et « groupe d'auditeurs » :  $F(7,1128)=4,23$  avec  $p<0,001$ . On remarque par exemple que les voix FA2, HA1 et HF2 ont posé bien plus de problèmes aux auditeurs anglophones (fort taux de réponses non congruentes toutes sessions confondues) qu'aux auditeurs francophones, à l'inverse de la voix FF1. Les deux groupes d'auditeurs n'ont donc pas réagi de la même manière aux différentes voix.

### **Comparaison des pourcentages de bonnes réponses obtenus par les francophones et les anglophones pour les sessions dans lesquelles seul le F0 moyen a été modifié (sessions 2 et 4)**

Sur les quatre sessions qui composent l'expérience, deux avaient pour but de tester l'influence du F0 moyen à des niveaux différents : il s'agit des sessions 2 et 4. Pour chacune d'entre elles, les anglophones ont obtenu des scores inférieurs aux francophones en termes de réponses congruentes. Cela pourrait suggérer que dans une tâche d'identification du genre, les anglophones sont plus sensibles à la fréquence fondamentale moyenne que les francophones. Afin de confirmer cette hypothèse, j'ai regroupé les résultats (pourcentage de réponses congruentes) des sessions 2 et 4, et testé le rôle du facteur « groupe d'auditeurs » (c'est-à-dire « langue maternelle des auditeurs ») *via* un test ANOVA. Le graphique correspondant est présenté ci-dessous (Figure 3).



**Figure 3** – Graphique représentant le pourcentage de réponses congruentes pour les sessions où le F0 moyen a été modifié (sessions 2 et 4) en fonction du groupe d'auditeurs, avec les barres d'erreur-types.

Les résultats de l'ANOVA sont les suivants : il existe bien un effet significatif du facteur « groupe d'auditeurs » sur le pourcentage de réponses congruentes pour les sessions 2 et 4 réunies, avec  $F(1,566)=4,897$  et  $p<0,05$ . Cela signifie que le facteur « fréquence fondamentale moyenne » a bien joué un rôle plus important chez les anglophones.

### **3. Conclusion - Discussion**

Cette expérience d'identification du genre par la voix a permis d'obtenir un certain nombre de résultats intéressants. Tout d'abord, le rôle de la fréquence fondamentale

moyenne a été largement mis en évidence : l'hypothèse 1 est donc vérifiée. Notons également que les auditeurs des deux groupes ont vu leur pourcentage de réponses congruentes diminuer très significativement lorsque les F0 moyens des hommes et des femmes ont été permutés. En revanche, lorsque le F0 moyen était resynthétisé à un niveau ambigu, seuls les degrés de certitude ont significativement baissé. Cela signifie qu'à un niveau ambigu, il semble que le rôle du F0 moyen diminue et soit compensé par les autres indices acoustiques, permettant ainsi à l'auditeur de reconnaître de manière très fiable le genre du locuteur, bien qu'étant moins sûr de lui.

L'hypothèse 2, selon laquelle la fréquence fondamentale moyenne joue un rôle plus important chez les anglophones que chez les francophones, semble elle aussi être confirmée. Le pourcentage de réponses congruentes dans les deux sessions qui testaient l'influence du F0 moyen, est significativement plus bas chez les anglophones que chez les francophones. Le groupe des anglophones s'est donc basé plus fortement sur ce paramètre pour identifier le genre.

L'aplatissement des courbes de F0, a entraîné une baisse significative du pourcentage de réponses congruentes et du degré de certitude moyen chez les francophones. En revanche, aucune modification significative n'a pu être observée chez les anglophones. L'hypothèse 3, qui prédisait une influence de l'intonation dans l'identification du genre par la voix, est donc partiellement vérifiée : ce paramètre semble bien jouer un rôle pour les auditeurs francophones, mais pas pour les anglophones.

Une analyse statistique complémentaire a permis de confirmer qu'il existe bien, de manière générale, une grande influence de la langue maternelle des auditeurs dans l'identification du genre par la voix : ce ne sont pas les mêmes voix qui ont posé problème aux anglophones et aux francophones. Cela renforce donc l'idée qu'ils n'utilisent pas forcément les mêmes paramètres acoustiques, ou tout du moins pas de la même manière.

Les résultats de cette étude sont à mettre en relation directe avec des recherches antérieures. Tout d'abord, la mise en évidence de l'importance accrue du F0 moyen chez les anglophones par rapport aux francophones apporte un éclairage nouveau sur les expériences menées d'un côté par Pausewang Gelfer et Mikos (2005) et Coleman (1976), et de l'autre par Arnold (2008). Pour rappel, les deux premières, réalisées sur des auditeurs anglophones américains, attribuent au F0 moyen un rôle plus important qu'au timbre lors d'une tâche d'identification du genre par la voix, contrairement à la troisième, menée sur des auditeurs francophones, qui présente le timbre comme plus déterminant. En considérant les résultats obtenus ici, ces études n'apparaissent plus contradictoires : il est fort probable que pour les anglophones américains la fréquence fondamentale moyenne constitue le paramètre acoustique le plus important, et qu'à l'inverse, pour les auditeurs francophones, il s'agisse bien du timbre.

La principale conclusion à tirer de ces observations est qu'il conviendrait de ne pas chercher à lister et hiérarchiser les indices acoustiques jouant un rôle dans l'identification du genre par la voix de manière universelle, mais de considérer l'influence de la langue maternelle des auditeurs sur leur(s) stratégie(s) d'identification.

## Références bibliographiques

**Arnold 2008:** Arnold, A. *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, 2008.

**Brend 1975:** Brend, R. M. Male-female intonation patterns in American English. In : *Proceedings of the Seventh International Congress of Phonetic Sciences*, La Haye (1971), 866–870. Reprinted in Thorne, B. & Henley, N. (1975). *Language and Sex: Difference and Dominance*, Newbury House: Rowley, 1975, 84-87.

**Coleman 1976:** Coleman, R. O. 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*, 1976, N° 19, 168-180.

**Fant 1966:** Fant, G. A note on vocal tract size factors and non-uniform F-pattern scaling. // *Speech Transmission Laboratory, Quarterly Progress and Status Report*, 1976, N° 81, 21-38.

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

**Martin 2008:** Martin, P. *Phonétique acoustique, introduction à l'analyse acoustique de la parole*. Paris : Armand Colin, 2008.

**Pausewang Gelfer & Mikos 2005:** Pausewang Gelfer, E., Mikos, V. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. // *Journal of Voice*, 2005, N° 19, 544-554.

**Pépiot 2009:** Pépiot, E. *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 8, 2009.

**Schwartz 1968:** Schwartz, M. F. Identification of speaker sex from isolated voiceless fricatives. // *Journal of the Acoustical Society of America*, 1968, N° 43, 1178-1179.

**Schweinberger & al. 2008:** Schweinberger, S. & al. Auditory adaptation in voice perception. // *Current Biology*, 2008, N° 18, 684-688.

**Vaissière 2006:** Vaissière, J. *La phonétique*. Paris : Presses Universitaires de France, 2006.