



HAL
open science

Voice, speech and gender: male-female acoustic differences and cross-language variation in English and French speakers.

Erwan Pépiot

► **To cite this version:**

Erwan Pépiot. Voice, speech and gender: male-female acoustic differences and cross-language variation in English and French speakers.. CORELA - COgnition, REprésentation, LAngage, 2015, HS-16, 10.4000/corela.3783 . hal-03998746

HAL Id: hal-03998746

<https://hal.science/hal-03998746>

Submitted on 21 Feb 2023

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.



Corela

Cognition, représentation, langage

HS-16 | 2015

Diversité des pratiques de recherche en science du langage

Voice, speech and gender:

Male-female acoustic differences and cross-language variation in English and French speakers

Erwan Pépiot



Édition électronique

URL : <https://journals.openedition.org/corela/3783>

DOI : 10.4000/corela.3783

ISSN : 1638-573X

Éditeur

Cercle linguistique du Centre et de l'Ouest - CerLICO

Référence électronique

Erwan Pépiot, « Voice, speech and gender: », *Corela* [En ligne], HS-16 | 2015, mis en ligne le 15 juin 2015, consulté le 21 septembre 2021. URL : <http://journals.openedition.org/corela/3783> ; DOI : <https://doi.org/10.4000/corela.3783>

Ce document a été généré automatiquement le 21 septembre 2021.



Corela – cognition, représentation, langage est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Voice, speech and gender:

Male-female acoustic differences and cross-language variation in English and French speakers

Erwan Pépiot

1. Introduction

- 1 Differences between female and male voices are linked to complex and multidisciplinary issues. They not only refer to acoustic (fundamental frequency, resonant frequencies, etc.) and perceptual measurements, but also to anatomy and physiology (differences in the vocal organs), sociology and even philosophy (construction of gender identity, innate versus learned behavior). The present study focuses on acoustic differences: I am thus adopting a phonetician's point of view.
- 2 Mean fundamental frequency, which is associated with the perceptual notion of pitch, is commonly considered as the major difference between adult male and female voices. Mean F0 would be around 120 Hz for men and 200 Hz for women (Takefuta *et al.*, 1972), but these values slightly vary through age (Pegoraro-Crook, 1988) and are broadly lower for smokers (Gilbert & Weismer, 1974). This acoustic parameter is indeed a decisive clue in the perception of gender from voice (Pépiot, 2010; 2011). A number of studies have brought to light other cross-gender acoustic differences. First of all, vowel formants of female speakers tend to be located at higher frequencies (Hillenbrand *et al.*, 1995; Pépiot, 2009), as well as consonant noise (Schwartz, 1968). Some studies (Takefuta *et al.*, 1972; Olsen, 1981) suggest that F0 range would be larger for female than for male speakers, even though there is no consensus on this point (see Simpson, 2009). Phonation type also seems to depend on the speaker's gender: female voices are often considered more breathy than male voices (Klatt & Klatt, 1990).
- 3 According to a majority of authors, cross-gender acoustic variations can mainly be accounted for by anatomical and physiological differences that arise during puberty (Fant, 1966). Vocal folds become longer and thicker in male speakers (Kahane, 1978): that would explain why they tend to vibrate more slowly than those of women. A second

important anatomical issue is vocal tract length, that is, the distance from the vocal folds to the lips: all things being equal, the longer the vocal tract, the lower resonant frequencies (Fant, 1970). The average length of the adult female vocal tract is about 14.5 cm, while the average male vocal tract is 17 to 18 cm long (Simpson, 2009). These would account, at least in part, for cross-gender differences observed in vowel formants and consonant noise.

- 4 How can one account for cross-language differences? For example, in a Chinese dialect, mean F0 is almost equivalent for male and female speakers (Rose, 1991). Furthermore, if one compares various acoustic studies about vowel formant frequencies conducted on different languages (Johnson, 2005), one can notice that cross-gender differences vary from one language to another: for instance, male-female differences are relatively small in Danish but appear to be much greater in Russian. Nonetheless, we need to take into account that comparisons made by Johnson were based on several studies led by different authors, at different times and with different methods. Therefore, we must be very careful when interpreting such results.
- 5 Given such facts, it seems quite interesting to conduct a cross-language study on acoustic differences between female and male voices. Additionally, we can notice that most studies in this field focus on a single acoustic parameter, but a multiparametric analysis would probably be much more productive. In the present study, I chose to work on cross-gender acoustic differences in Parisian French and Northeastern American English speakers, with the following hypothesis: *cross-gender acoustic differences are language dependent*.

2. Method

2.1 Linguistic material

- 6 To conduct this study, an English and a French corpus were necessary. I used “CVCV” dissyllabic words or pseudo-words, so that many phoneme combinations could be tested. Their selection was based on two main criteria: make the two corpora as similar as possible (e.g. English inter-dental fricatives were dismissed as there is no equivalent in French), and limit the number of combinations by choosing only the most pertinent phonemes (e.g. cardinal vowels) while holding constant the last “CV” sequence (/pi/ was chosen as it can appear on word final position in both languages). Twenty-seven words or pseudo-words for each language were finally chosen:

-/C (**plosive**) - V - p - i / **combinations**: /tipi/, /tapi/, /tupi/, /dipi/, /dapi/, /dupi/, /kipi/, /kapi/, /kupi/, /gipi/, /gapi/, /gupi/ for the French corpus, /'ti:pi/, /'tʌpi/, /'tu:pi/, /'di:pi/, /'dʌpi/, /'du:pi/, /'ki:pi/, /'kʌpi/, /'ku:pi/, /'gi:pi/, /'gʌpi/, /'gu:pi/ for the English corpus.

-/C (**fricative**) - V - p - i / **combinations**: /sipi/, /sapi/, /supi/, /zipi/, /zapi/, /zupi/, /ʃipi/, /ʃapi/, /ʃupi/, /ʒipi/, /ʒapi/, /ʒupi/ for the French corpus, /'si:pi/, /'sʌpi/, /'su:pi/, /'zi:pi/, /'zʌpi/, /'zu:pi/, /'ʃi:pi/, /'ʃʌpi/, /'ʃu:pi/, /'ʒi:pi/, /'ʒʌpi/, /'ʒu:pi/ for the English corpus.

-/V - p - i / **combinations**: /ipi/, /api/, /upi/ for the French corpus, /'i:pi/, /'ʌpi/, /'u:pi/ for the English corpus.

2.2 Speakers

7 Eight monolingual speakers participated in the experiment. Four of them are native Parisian French speakers (2 women, 2 men) and four others are native Northeastern American English speakers (2 women and 2 men). They are aged from 23 to 40, are non-smokers and have no reported speech or voice disorder. Here is a brief description of each speaker:

- **French female speaker 1 (F1FR):** 27, student, Paris area.
- **French female speaker 2 (F2FR):** 23, student, Paris area.
- **French male speaker 1 (M1FR):** 23, student, Paris area.
- **French male speaker 2 (M2FR):** 24, student, Paris area.
- **American English female speaker 1 (F1EN):** 40, teacher, Northampton (MA).
- **American English female speaker 2 (F2EN):** 23, student, Brattleboro (VT).
- **American English male speaker 1 (M1EN):** 39, student, Philadelphia (PN).
- **American English male speaker 2 (M2EN):** 26, teacher, Binghamton (NY).

2.3 Recording procedure

8 Recordings took place in a quiet room, using a digital recorder *Edirol R09-HR* by Roland. English speakers read the English corpus aloud and French speakers the French one. Words were presented to the participants with an orthographical transcription. Moreover, in order to make prosodic parameters consistent, words were placed into a frame sentence: “He said WORD twice” for the English corpus and “Il a dit MOT deux fois” for the French one. Speakers were asked to say each sentence twice, at a normal speech rate.

3. Data analysis

9 Data analysis was conducted with *Praat* software². The different steps of the analysis are described below.

3.1 Segmentation and labelling

10 Words were first extracted from the frame sentence. Since all the items were recorded twice, only the most acoustically satisfactory occurrence was selected, making up a total of 108 words for each language (27 items * 4 speakers). I then segmented and labeled words into phones. These tasks were performed manually with *Praat*. Segmentation was based jointly on waveform and spectrogram and each segment boundary was located at a zero crossing. To make further acoustic analysis more convenient, each phone was then extracted into a separate sound file.

3.2 Acoustic analysis of consonants

11 Duration³ and center of gravity of each initial consonant were computed. Voice onset time of plosives was measured as well as mean F0 of voiced consonants. As expected in an initial position followed by a vowel and under lexical stress, English plosives /t/ and /k/

are phonetically performed as [t^h] and [k^h] and their counterparts /d/ and /g/ as voiceless non-aspirated plosives (i.e. [d̥] and [k̥]). Therefore, mean F0 could not be measured on these segments.

- 12 To obtain duration and mean F0, *Get total duration* and *Get mean* commands were once again used. Center of gravity was measured by using the *Get centre of gravity* command on the *spectrum* object created for each sound file. All these procedures were automated with a script, but I performed an *a posteriori* verification on the data: when results seemed incoherent, a manual measurement was made. Finally, VOT was measured manually for each initial plosive consonant. To do so, I had to localize the consonant release and the beginning of voicing on spectrograms, voice onset time being the temporal spacing between the first point and the second. A reminder: if voicing begins after the release, VOT is positive, if it begins before the release, VOT is negative.

3.3 Acoustic analysis of vowels

- 13 Multiple measures were made on first syllable vowels. Duration (also measured on second syllable vowels) and mean F0 were obtained using the same procedure as in 3.2 and 3.3. Frequencies of the first three formants (F1, F2 and F3) were manually measured using spectrograms, automatic formant track detection and spectra. Values were taken in a central and stable portion of vowels, in order to limit the influence of coarticulation.
- 14 I also took into account speakers' phonation type. The most reliable acoustic measurement (Gordon & Ladefoged, 2001) seems to be the relative intensity of H1 (first harmonic) compared with H2 (second harmonic). According to Klatt & Klatt (1990) and Gordon & Ladefoged (2001), the relative strength of H1 is correlated with glottal open quotient (GOQ): the stronger it is, the higher the GOQ. A voice with a high GOQ will tend to be perceived as breathy, while a low GOQ is associated with a creaky voice (Gordon & Ladefoged, 2001). Nevertheless, certain precautions have to be taken. This measurement should not be performed on isolated vowels and on vowels followed or preceded by a nasal consonant (Simpson, 2012). Furthermore, H1-H2 can only be measured on open vowels: F1 would otherwise distort the results (Klatt & Klatt, 1990). Thus, only vowel [a] for French speakers and vowel [ɑ] for English speakers were taken into account. A 5 period selection was made on a central part of the vowel. The corresponding spectrum was displayed and the difference between H1 and H2 intensity (in dB) was then calculated.

3.4 F0 and duration measurements of entire words

- 15 Duration and mean F0 of entire words were obtained by creating a *Pitch file* for each word, and performing *Get total duration* and *Get mean* commands. This operation was automated by a *Praat* script. The third measurement performed on entire words was F0 range: these data were collected in semitones, through the *Pitch info* window.

3.5 Statistical analysis

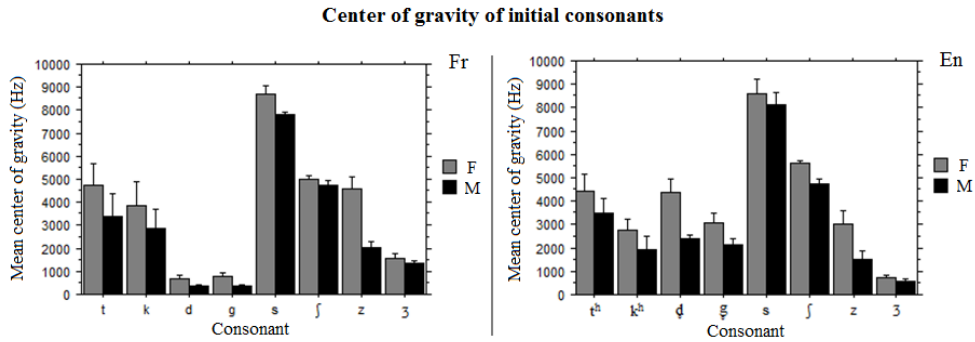
- 16 In order to test if cross-gender differences are statistically significant, ANOVAs were conducted on the data, for each acoustic parameter.

4. Results

4.1 Center of gravity

17 Results for the center of gravity of initial consonants are presented in the figure below⁴.

FIGURE 1: CENTER OF GRAVITY OF INITIAL CONSONANTS FOR MALE (M) AND FEMALE (F) FRENCH SPEAKERS (LEFT PART) AND AMERICAN ENGLISH SPEAKERS (RIGHT PART).

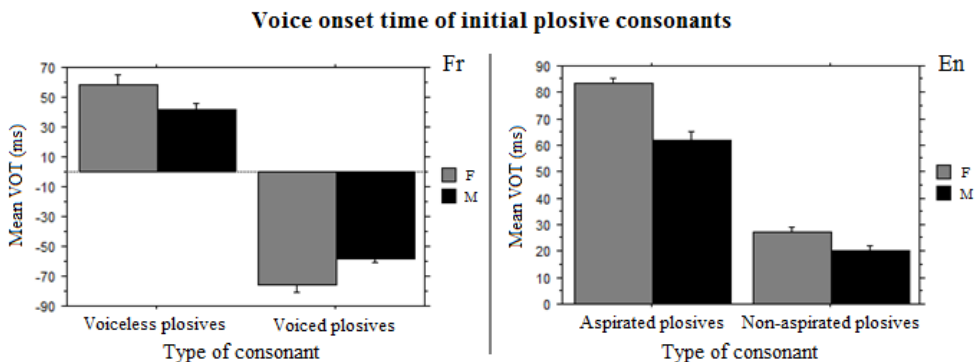


18 For French speakers, the center of gravity is higher for women than for men on every consonant. I performed a two-factor ANOVA (“speaker’s gender” and “consonant”) on these data. Results show that there is a significant overall effect of the speaker’s gender on the center of gravity: it is much higher for female speakers ($F(1,80)=11.501$, $p<0.01$). Furthermore, there is no interaction between the two factors ($F(7,80)=1.143$, $p>0.3$), which means that cross-gender difference remains relatively constant across consonants. Similar tendencies are found in American English speakers. Women’s center of gravity is significantly higher than men’s ($F(1,80)=18.863$, $p<0.0001$) and there is no interaction between factors “speaker’s gender” and “consonant” ($F(7,80)=0.811$, $p>0.5$).

4.2 Voice onset time

19 Results for the voice onset time of initial plosive consonants are shown in figure 2.

FIGURE 2: VOICE ONSET TIME (MS) OF INITIAL PLOSIVE CONSONANTS FOR MALE (M) AND FEMALE (F) FRENCH SPEAKERS (LEFT PART) AND AMERICAN ENGLISH SPEAKERS (RIGHT PART).



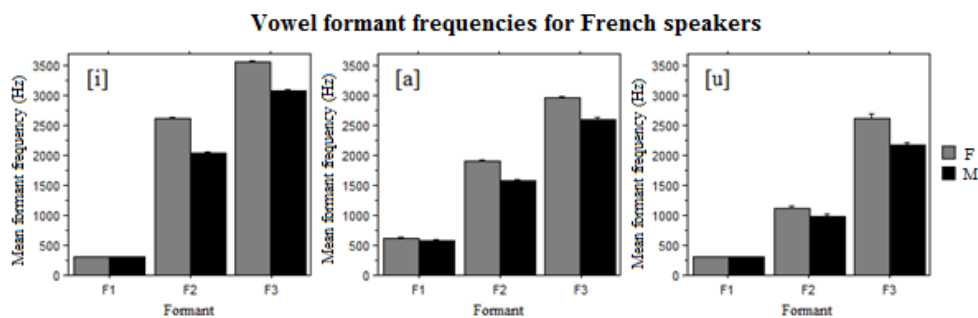
20 Regarding Parisian French speakers, a one-factor ANOVA (“speaker’s gender”) shows that women’s mean VOT is significantly longer than men’s in voiceless plosives ($F(1,22)=4.332$, $p<0.05$), while it is significantly shorter for voiced plosives ($F(1,22)=9.87$, $p<0.01$).

Unsurprisingly, if we consider mean VOT contrast between the two types of plosives, a one-factor ANOVA (“speaker’s gender”) indicates that it is significantly greater for female than for male speakers ($F(1,22)=18.195$, $p<0.001$). Concerning the Northeastern American English speakers, similar statistical tests show that mean VOT is significantly longer for female speakers in aspirated plosives ($F(1,22)=29.584$, $p<0.0001$). Unlike French speakers, it is also slightly but significantly longer for women in non-aspirated plosives ($F(1,22)=10.42$, $p<0.01$). However, the mean VOT contrast between the two types of plosives (here aspirated versus non-aspirated) remains significantly greater for female speakers ($F(1,22)=10.816$, $p<0.01$).

4.3 Vowel formants

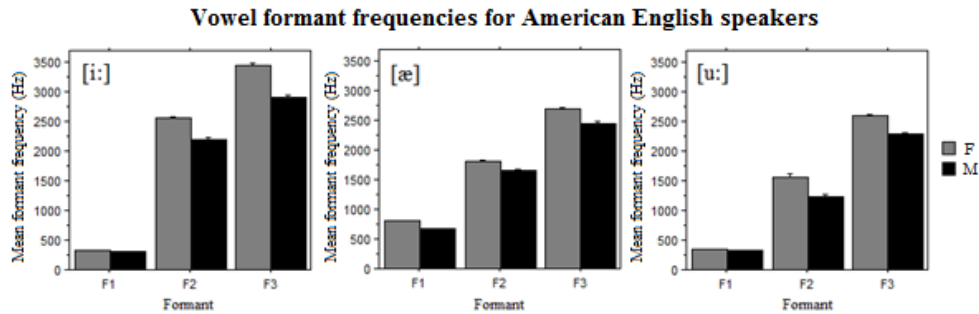
- 21 Vowel formant frequencies for the Parisian French speakers are presented in figure 4.

FIGURE 3: VOWEL FORMANT FREQUENCIES (HZ) FOR MALE (M) AND FEMALE (F) FRENCH SPEAKERS.



- 22 As expected, overall formant frequencies of female speakers are higher than those of male French speakers. I performed a two-factor ANOVA (“speaker’s gender” and “vowel”) for each formant to check if differences are significant. Results for the first formant (F1) show that there is no overall significant cross-gender difference ($F(1,102)=0.914$, $p>0.3$). No interaction was found between the two factors ($F(2,102)=2.494$, $p>0,05$). For F2, the ANOVA shows a very significant overall gender effect ($F(1,102)=247.477$, $p<0.0001$): frequencies are significantly higher for female speakers. Unlike what was found for F1, there is now a strong interaction between the two factors “speaker’s gender” and “vowel” ($F(2,102)=34,684$; $p<0,0001$). Three one-factor ANOVAs (“speaker’s gender”) were then conducted for each vowel individually. A widely significant gender effect was found for the F2 of [i] ($F(1,34)=525.914$, $p<0.0001$) and [a] ($F(1,34)=98.642$, $p<0.0001$), but it was barely significant for back vowel [u] ($F(1,34)=6.521$, $p<0.02$). Concerning the third formant (F3), the two-factor ANOVA reveals a widely significant overall gender effect ($F(1,102)=240.17$, $p<0.0001$) and no interaction with the “vowel” factor ($F(2,102)=1.433$, $p>0.2$).

FIGURE 4: VOWEL FORMANT FREQUENCIES (HZ) FOR MALE (M) AND FEMALE (F) AMERICAN ENGLISH SPEAKERS.

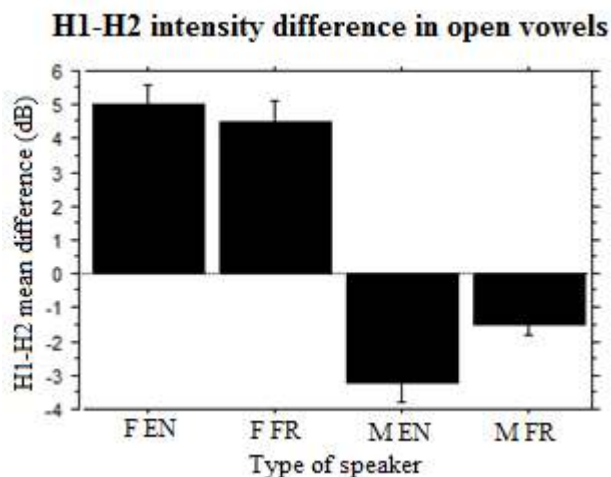


- 23 For American English speakers, overall formant frequencies also appear to be globally higher for women. Similar statistical tests were performed again. Contrary to French speakers, a significant gender effect was found for F1 ($F(1,102)=364.857$, $p<0.0001$). There is a large interaction between factors “speaker’s gender” and “vowel” for this formant. Individual one-factor ANOVAs show a very large and significant cross-gender difference for open vowel [æ] ($F(1,34)=236.665$, $p<0.0001$) and smaller but significant differences for [i:] ($F(1,34)=92.298$, $p<0.0001$) and [u:] ($F(1,34)=62.373$, $p<0.001$). Regarding the second formant (F2), there is a highly significant gender effect ($F(1,102)=98.541$, $p<0.0001$) and a low, albeit significant, interaction between “speaker’s gender” and “vowel” ($F(2,102)=5.002$, $p<0.01$). Nonetheless, separated ANOVAs show that male-female differences remain constantly strong among vowels [i:] ($F(1,34)=54.372$; $p<0.0001$), [æ] ($F(1,34)=132.237$; $p<0.0001$) and [u:] ($F(1,34)=23.207$; $p<0.0001$). Finally, the ANOVA performed on F3 data shows a very significant overall gender effect ($F(1,102)=290.178$, $p<0.0001$), with an important interaction between factors “speaker’s gender” and “vowel” ($F(2,102)=18.578$, $p<0.0001$). Individual one-factor ANOVAs reveal that cross-gender difference for F3 is greater for close vowels [i:] ($F(1,34)=132.54$, $p<0.0001$) and [u:] ($F(1,34)=135.443$, $p<0.0001$) than for [æ] ($F(1,34)=50.129$, $p<0.0001$).

4.4 H1-H2

- 24 Results for H1-H2 intensity difference in open vowels are shown in the figure below.

FIGURE 5: H1-H2 INTENSITY DIFFERENCE (DB) IN OPEN VOWELS FOR ENGLISH (EN) AND FRENCH (FR) MALE (M) AND FEMALE (F) SPEAKERS.

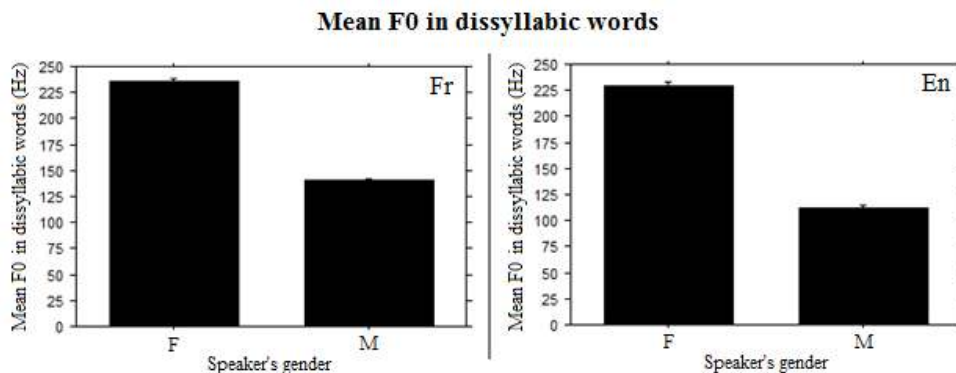


- 25 French speakers' H1-H2 difference in open vowel [a] appears to be much greater for women than for men. A one-factor ANOVA ("Speaker's gender") indicates that this difference is widely significant ($F(1,34)=69.516$, $p<0.0001$). This suggests that French female speakers have a higher GOQ, thus a more breathy voice than male speakers. We can notice an analogous tendency for American English speakers. Indeed, a similar statistical test shows that H1-H2 difference in open vowel [ɑ] is significantly greater for female speakers ($F(1,34)=101.079$, $p<0.0001$), hence a more breathy voice quality. Finally, I conducted another one-factor ANOVA ("type of speaker") for speakers of both languages at the same time. The overall effect of this factor is obviously significant ($F(3,68)=58.62$, $p<0.0001$). More interestingly, Fisher's PLSD test shows that there is no significant difference between French and English female speakers ($p>0.4$) but a significant difference between French and English male speakers ($p<0.05$): H1-H2 is lower for English speakers, which suggests they have a smaller GOQ, thus a more creaky voice.

4.5 Mean F0

- 26 Mean F0 in dissyllabic words for both English and French speakers is displayed in figure 6.

FIGURE 6: MEAN F0 (HZ) IN DISSYLLABIC WORDS FOR FRENCH (LEFT PART) AND AMERICAN ENGLISH (RIGHT PART) MALE (M) AND FEMALE (F) SPEAKERS.

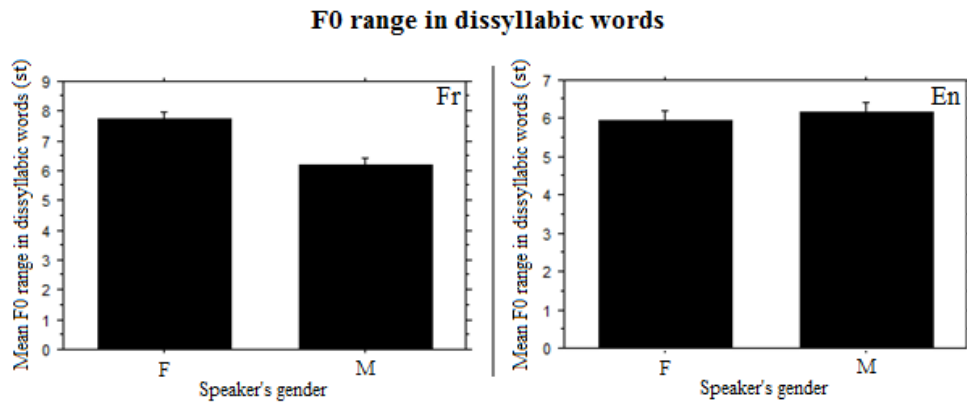


- 27 Unsurprisingly, mean F0 is much higher for female speakers in both languages. A one-factor ANOVA ("speaker's gender") shows that this difference is highly significant for French speakers ($F(1,106)=951.013$, $p<0.0001$) as well as for American English speakers ($F(1,106)=1159.938$, $p<0.0001$).

4.6 F0 range

- 28 Results for F0 range in dissyllabic words are presented in the following figure.

FIGURE 7: MEAN F0 RANGE (ST) IN DISSYLLABIC WORDS FOR FRENCH (LEFT PART) AND AMERICAN ENGLISH (RIGHT PART) MALE (M) AND FEMALE (F) SPEAKERS.

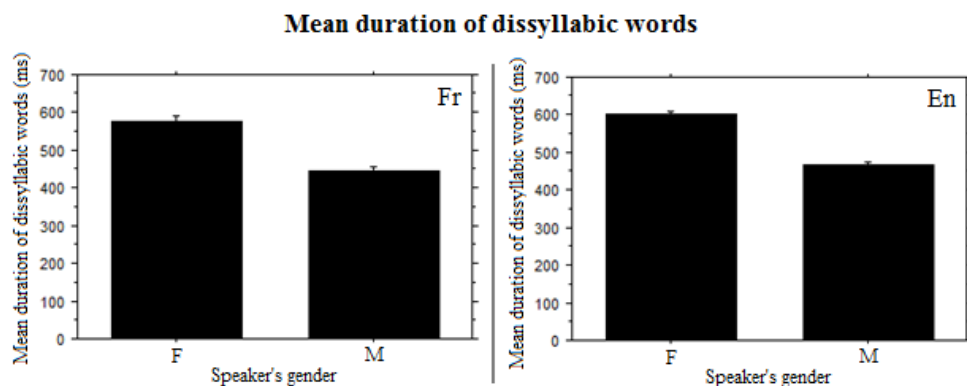


- 29 For French speakers, F0 range in semitones is greater for women than for men. A one-factor ANOVA (“speaker’s gender”) indicates that the difference is widely significant ($F(1,106)=22.489$, $p<0.0001$). On the contrary, there is no significant cross-gender difference for American English speakers ($F(1,106)=0.383$, $p>0.5$).

4.7 Duration

- 30 Mean duration of dissyllabic words for French and English speakers is shown in figure 8.

FIGURE 8: MEAN DURATION (MS) OF DISSYLLABIC WORDS FOR FRENCH (LEFT PART) AND AMERICAN ENGLISH (RIGHT PART) MALE (M) AND FEMALE (F) SPEAKERS.



- 31 Word duration appears to be greater for female than for male speakers. I conducted one-factor ANOVAs (“speaker’s gender”) to check if these differences were significant. Results show that it is highly significant, for both French ($F(1,106)=67.524$, $p<0.0001$) and American English speakers ($F(1,106)=123.6$, $p<0.0001$). If we take a closer look at duration measurements, it appears that “consonant / vowel” temporal distribution in words are slightly different for male and female speakers in both languages. For French speakers, consonants represent 52 % of total word duration for women but only 46 % for men. A similar tendency is found for American English speakers (48 % for female versus 42 % for male speakers). One-factor ANOVAs (“speaker’s gender”) show that this difference is significant for French ($F(1,94)=17.409$, $p<0.0001$) as well as for American English speakers ($F(1,94)=17.975$, $p<0.0001$).

5. General discussion and conclusions

- 32 This acoustic analysis has given interesting results. Concerning fundamental frequency, we saw that mean F0 of dissyllabic words was significantly higher for women in both languages, which broadly confirms results obtained in previous studies (e.g. Takefuta *et al.*, 1972). F0 range measurements have highlighted an interesting cross-language variation. While F0 range (in semitones) was significantly larger for female French speakers comparing to their male counterparts, there is no such difference for American English speakers. These results support a former perceptual study (Pépiot, 2010; 2011) that showed a tendency for French listeners to associate flat F0 sentences with male voices, whereas no such effect was observed for American English listeners.
- 33 Resonant frequency analysis also pointed out cross-gender and cross-language differences. Consonants' center of gravity was significantly higher in female voices for French as well as for American English speakers. These results are close to those obtained by Schwartz (1968). Vowel formant frequencies appeared to be generally higher for women, but results varied strongly depending on vowel, language and formant number. No significant cross-gender F1 difference was found for French speakers, but F1 values were significantly greater for female English speakers, especially for open vowel [æ]. F2 values were significantly higher for female speakers in both languages, particularly for close front vowels. However, while the difference was quite small in [u] for French speakers, it was fairly large in American English [u:]. These observations seem to support former results (Hillenbrand *et al.*, 1995; Pépiot, 2009) and suggest that American English female speakers tend to slightly centralize their close back vowel [u:] compared to French female speakers. Third formant analysis showed significantly higher values for female voices in both languages and no important variations among vowels.
- 34 H1-H2 measurements for open vowels gave precious indications about speakers' phonation type. For French as for English speakers, significant cross-gender differences were found: female voices had much higher H1-H2 values than males, which suggests they tend to speak with a more breathy voice quality. Furthermore, American English male speakers had a significantly lower H1-H2 than French male speakers. This indicates a very low GOQ, hence a more creaky voice. These results support the claim that female speakers' breathy voice quality could have a physiological origin (Simpson, 2009) whereas male speakers' use of creaky voice would rather be socio-phonetic and language dependant (Henton, 1989).
- 35 VOT measures led to conclusive results. For French speakers, women's mean VOT was significantly longer in voiceless plosives and significantly shorter in voiced plosives. Results for American English speakers were slightly different: VOT was significantly longer for female speakers on aspirated plosives, but also in a less extent on non-aspirated plosives. Nevertheless, mean VOT contrast between the two types of plosives (voiced/voiceless in French, aspirated/non-aspirated in English) was significantly larger for female speakers in both languages. Thus, women would tend to mark a greater distinction between these phonetic pairs. This phenomenon could be explained by socio-phonetic and cultural factors and favors the idea that female speakers would try to achieve a more intelligible speech than male speakers (Simpson, 2009).
- 36 Besides VOT, other duration measurements were conducted. Dissyllabic words' overall duration was significantly greater for female speakers in both languages. This supports

former results obtained by Byrd (1994) and holds with the socio-phonetic explanation mentioned above. Another interesting result was found for consonant / vowel temporal distribution. Consonants were proportionally longer in words produced by female speakers than by men. It is known that consonants are likely to be more important than vowels in oral word recognition (Owren & Cardillo, 2006). These results could therefore be linked, once again, to female speakers' tendency to produce "clearer" speech.

- 37 This multiparametric acoustic analysis has brought to light several cross-gender differences, but also some cross-language variation between Parisian French speakers and Northeastern American English speakers. This tends to validate the general hypothesis which claimed that cross-gender acoustic differences are language dependent, even though many cross-gender differences were found in both languages. Moreover, most of the differences found in this study are unlikely to be explained by physiological and anatomical factors. A large part of cross-gender variation can probably be accounted for by gender social construction. Therefore, these data may be of interest for improving vocal rehabilitation for transgender people (Wiltshire, 1995).
- 38 Nonetheless, results obtained in the present study have to be interpreted with caution. First of all, only two men and two women were recorded for each language. Despite the restrictive selection criteria and the very small intra-gender variation, it seems difficult to generalize the results to the whole Parisian French and Northeastern American English speaker populations. Besides, corpora were made of read disyllabic words: it is uncertain whether similar results would be obtained with spontaneous speech.

BIBLIOGRAPHIE

- BYRD, Dani (1994), « Relations of sex and dialect to reduction », *Speech Communication*, 15, 39-54.
- FANT, Gunnar (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, Gunnar (1970), *Acoustic Theory of Speech Production*, The Hague : Mouton.
- GILBERT, Harvey & WEISMER, Gary (1974), « The effects of smoking on the speaking fundamental frequency of adult women », *Journal of Psycholinguistic Research*, 3, 225-231.
- GORDON, Matthew & LADEFOGED, Peter (2001), « Phonation types: A crosslinguistic overview », *Journal of Phonetics*, 29, 383-406.
- HENTON, Caroline (1989), « Sociophonetic aspects of creaky voice », *Journal of the Acoustical Society of America*, 86, S26.
- HILLENBRAND, James *et al.* (1995), « Acoustic characteristics of American English vowels », *Journal of the Acoustical Society of America*, 97, 3099-3111.
- JOHNSON, Keith (2005), « Speaker normalization in speech perception », in David Pisoni & Robert Remez (dir.), *The Handbook of Speech Perception*, Oxford : Blackwell Publishers, 363-389.

- KAHANE, Joel (1978), « A morphological study of the human prepubertal and pubertal larynx », *American Journal of Anatomy*, 151, 11-20.
- KLATT, Dennis & KLATT, Laura (1990), « Analysis, synthesis and perception of voice quality variations among female and male talkers », *Journal of the Acoustical Society of America*, 87, 820-857.
- OLSEN, Carrol (1981), « Sex differences in English intonation observed in female impersonation », *Toronto Papers of the Speech and Voice Society*, 2, 30-49.
- OWREN, Mickael & CARDILLO, Gina (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.
- PEGORARO-CROOK, Maria (1988), « Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis », *Folia Phoniatica*, 40, 82-90.
- PÉPIOT, Erwan (2009), *The making of French vocalic triangles: the case of a woman's voice versus a man's voice*, Master 1 thesis, University Paris 8.
- PÉPIOT, Erwan (2010), *Sur l'identification du genre par la voix chez des auditeurs anglophones et francophones*, Master 2 thesis, University Paris 8.
- PÉPIOT, Erwan (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*, 49, (to be published).
- ROSE, Phil (1991), « How effective are long term mean and standard deviation as normalization parameters for tonal fundamental frequency? », *Speech Communication*, 10, 229-247.
- SCHWARTZ, Martin (1968), « Identification of speaker sex from isolated voiceless fricatives », *Journal of the Acoustical Society of America*, 43, 1178-1179.
- SIMPSON, Adrian (2009), « Phonetic differences between male and female speech », *Language and Linguistics Compass*, 3, 621-640.
- SIMPSON, Adrian (2012), « The first and second harmonics should not be used to measure breathiness in male and female voices », *Journal of Phonetics*, 40, 477-490.
- TAKEFUTA, Yukio *et al.* (1972), « A statistical analysis of melody curves in the intonation of American English », in *Proceedings of the 7th International Congress of Phonetic Sciences*, Montreal, 1035-1039.
- TITZE, Ingo (1989), « Physiologic and acoustic differences between male and female voices », *Journal of the Acoustical Society of America*, 85, 1699-1707.
- WILTSHIRE, Anne (1995). Not by pitch alone: a view of transsexual vocal rehabilitation. *National Student Speech Language Hearing Association Journal*, 22, 53-57.

NOTES

1. In the English corpus, lexical stress is always on the first syllable.
2. Praat version 5.1.43
3. Duration was also measured on the words' second consonant [p], in order to establish C/V temporal distribution on entire words.
4. All the figures displayed in this section contain error-bars.

RÉSUMÉS

Un grand nombre d'études ont été menées sur les différences acoustiques entre les voix de femmes et d'hommes. 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 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, 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.

Many studies were conducted on acoustic differences between female and male voices. However, they were generally led on speakers of only one language, and focused on a single acoustic parameter. The present study is an acoustic analysis of dissyllabic words or 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 socially constructed and language dependent.

INDEX

Mots-clés : phonétique, voix et genre, différences acoustiques inter-genres, variations inter-langues, français parisien, anglais américain

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

AUTEUR

ERWAN PÉPIOT

Université Paris 8
EA1569 - Groupe LAPS
erwan.pepiot@free.fr