ПЛОВДИВСКИ УНИВЕРСИТЕТ „ПАИСИЙ ХИЛЕНДАРСКИ“ – БЪЛГАРИЯ
НАУЧНИ ТРУДОВЕ, ТОМ 53, КН. 1, СБ. А, 2015 – ФИЛОЛОГИЯ,
PAISII HILENDARSKI UNIVERSITY OF PLOVDIV – BULGARIA
RESEARCH PAPERS, VOL. 53, BOOK 1, PART A, 2015 – LANGUAGES AND LITERATURE

FEMALE AND MALE SPEECH: A STUDY OF VOWEL FORMANTS AND CONSONANT NOISE IN PARISIAN FRENCH AND AMERICAN ENGLISH SPEAKERS

Erwan Pépiot
Université Paris 8

The present study is an acoustic analysis of vowels and consonants in disyllabic words produced by 10 Northeastern American English speakers (5 females, 5 males) and 10 Parisian French speakers (5 females, 5 males). Vowel formant frequencies were measured, as well as initial voiceless consonants' spectral centre of gravity. Significant cross-gender differences were obtained for each parameter, with higher frequencies for female speakers. Moreover, cross-language variations were found: female/male differences on F1 appeared to be greater in American English than in Parisian French speakers, and the opposite was true with consonant noise. Such results support the idea that cross-gender acoustic differences are partly language-dependent and therefore, socially constructed.

Key words: phonetics, speech and gender, vowel formants, consonant noise, American English, Parisian French

Introduction

Mean fundamental frequency is generally considered the main difference between adult male and female speech. It would be around 120 Hz for men and 200 Hz for women (Takefuta et al. 1975; Boë et al. 1975). Several studies have brought to light other cross-gender acoustic differences, especially in resonant frequencies.

First, vowel formants of female speakers tend to be located at higher frequencies (Hillenbrand et al. 1995; Whiteside 2001; Ericsdotter 2005). The scope of cross-gender differences varies from one study to another, from one formant to another, and seems to depend on vowel type. Such tendency was also found in consonants. Indeed, consonant noise in voiceless fricatives has been shown to be located at higher frequencies when produced by female speakers (Shadle et al. 1991; Nittrouer 1995; Jongman et al. 2000), except for labio-dental and inter-dental fricatives (Schwartz 1968; Fox & Nissen 2005).

Some of these cross-gender acoustic variations could be accounted for by anatomical and physiological differences that arise during puberty

(Fant 1966). Vocal folds then become longer and thicker in male speakers (Kahane 1978): that would explain the differences on fundamental frequency. A second important anatomical issue is vocal tract length. In adult female speakers it averages 14.5 cm versus 17 to 18 cm long in adult male speakers (Simpson 2009). These characteristics would account, at least in part, for cross-gender differences observed in vowel formants and consonant noise (Fant 1975; Nordstrom 1977).

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. 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. Moreover, no similar investigation has been conducted on consonant noise.

Given such facts, it seemed relevant to conduct a cross-language study on acoustic differences between female and male speech. In the present study, cross-gender acoustic differences in Northeastern American English and Parisian French speakers on vowel formats and consonant noise were investigated. The general hypothesis was the following: *cross-gender acoustic differences are language dependent.*

1. Material and method

1.1. Linguistic material

French and English linguistic material was needed for this study. Disyllabic words and pseudo-words were used, so that many phoneme combinations could be tested. Their selection was based on two main criteria: make the two corpora as similar as possible, and limit the number of combinations by choosing only the most relevant phonemes while holding the last CV sequence constant: /pi/ was chosen as it can appear in word final position in both languages. Twenty-seven (C)VCV words were finally chosen for each language:

- /C (plosive) – V – p – i/ combinations: /tipi/, /tapi/, /tupi/, /dipi/, /dapi/, /dupi/, /kipi/, /kapi/, /kupi/, /gipi/, /gapi/, /gupi/ for the French corpus, /'ti:pi/ , /'tæpi/, /'tu:pi/, /'di:pi/, /'dæpi/, /'du:pi/, /'ki:pi/, /'kæpi/, /'ku:pi/, /'gi:pi/, /'gæpi/, /'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/, /'sæpi/, /'su:pi/, /'zi:pi/, /'zæpi/, /'zu:pi/, /'ʃi:pi/, /'ʃæpi/, /'ʃu:pi/, /'ʒi:pi/, /'ʒæpi/, /'ʒu:pi/ for the English corpus.

- /V – p – i / combinations: /ipi/, /api/, /upi/ for the French corpus, /ˈi:pi/, /ˈæpi/, /ˈu:pi/ for the English corpus.

There is no phonological lexical stress in French (Di Cristo 1999), but within the frame sentence used for the recordings (see 1.3) French speakers naturally produced an emphatic stress on the first syllable of each experimental word.

1.2 Speakers

Twenty monolingual speakers were recorded. Ten of them were French native speakers (5 women, 5 men) and ten others were American English native speakers (5 women and 5 men). The 10 American speakers all came from the same northeastern area of the United States (Pennsylvania, Massachusetts, New York State, or southern Vermont). The 10 French speakers all came from Paris area (Ile-de-France). Speakers were aged from 20 to 40 (SD = 6.5 years). Mean age was 28.2 for US speakers (29.4 for females, 27 for males) and 26.6 for French speakers (27.2 for females, 26 for males).

All speakers were non-smokers and had reported no speech disorder. Each of them received a USB memory stick for their participation in the study and was informed that the data from the recordings would be treated with confidentiality.

1.3 Recording procedure

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 in an orthographical transcription. 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.

Words were first extracted from the frame sentence. All the items having been recorded twice, only the most acoustically satisfactory occurrence was selected, thus making up a total of 270 words for each language (27 items * 10 speakers). Words were segmented and labeled into phones. Each phone was then extracted into a separate sound file. These tasks were performed manually in *Praat*.

Frequencies of the first three formants (F1, F2 and F3) were manually measured on the first syllable vowels, using spectrograms and spectra. Values were taken in a central and stable portion of the vowel.

Spectral centre of gravity of each initial voiceless consonant was computed in *Praat*. This parameter is a measure for how high the frequencies in a spectrum are on average. It was measured by using the *Get centre of gravity* command on the *spectrum* object created for each sound file. This procedure was automated with a script.

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

2. Results

2.1 Vowel formants (F1, F2 and F3)

American English speakers

Mean vowel formant frequencies (Hz) for American English speakers as a function of speaker's gender are presented in Table 1, below.

Table 1. Mean vowel formant frequencies (Hz) for F1, F2 and F3 for vowels [i:] [Θ] and [u:] produced by female (n = 5) and male (n = 5) American English speakers. SD among the 45 measurements (9 occurrences * 5 speakers) per formant for each vowel is also mentioned

		Mean frequency (Hz)		
Vowel	**Formant**	**Female speakers**	**Male speakers**	**Ratio F/M**
[i:]	**F1**	342	302	1,13
	σ	*23*	*8*	
	F2	2649	2210	1,20
	σ	*121*	*188*	
	F3	3459	2948	1,17
	σ	*113*	*164*	
[æ]	**F1**	804	684	1,18
	σ	*34*	*36*	
	F2	1826	1653	1,10
	σ	*41*	*48*	
	F3	2672	2453	1,09
	σ	*64*	*112*	
[u:]	**F1**	355	316	1,13
	σ	*19*	*14*	
	F2	1629	1299	1,25
	σ	*240*	*212*	
	F3	2628	2281	1,15
	σ	*107*	*130*	

A two factor ANOVA ("vowel" and "speaker's gender") was conducted for F1. It showed a very significant global effect of the speaker's gender: $F(1,264) = 491.421$; $p < 0.0001$. This cross-gender difference appeared to be significant for each vowel taken separately, with $p < 0.0001$ in all cases.

The same analysis was performed for F2 values. There is a significant overall effect of the speaker's gender: $F(1,264) = 254,159$; $p < 0.0001$. Female/male difference is significant for each of the three vowels, with $p < 0.0001$.

For the third formant, similar tendencies were found, with a strong and significant global effect of the speaker's gender: $F(1,264) = 617,626$; $p < 0.0001$. This cross-gender difference is once again significant for each vowel taken individually, with $p<0.001$ in all cases.

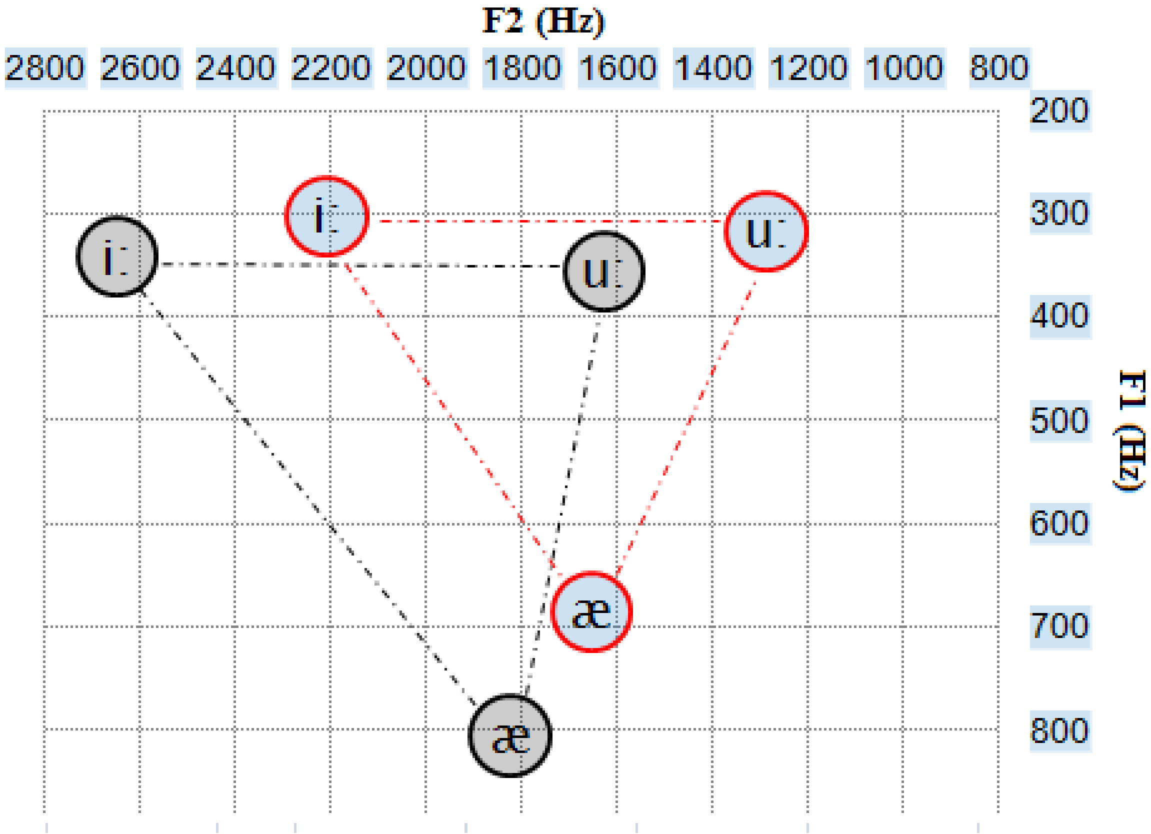

Figure 1. Vowel chart representing mean vowel formant frequencies (Hz) for vowels [i:] [Θ] and [u:] produced by female and male American English speakers.

In order to make such cross-gender differences clearly visible, results for the two first formants are presented in a vowel chart (Figure 1), using a template from *SaRP* software (Nikolov et al. 2011).

Parisian French speakers

Vowel formant frequencies (Hz) in Parisian French speakers as a function of speaker's gender are presented in Table 2.

Table 2. Mean vowel formant frequencies (Hz) for F1, F2 and F3 for vowels [i] [a] and [u] produced by female (n = 5) and male (n = 5) Parisian French speakers.
SD among the 45 measurements (9 occurrences * 5 speakers) per formant for each vowel is also mentioned

		Mean frequency (Hz)		
Vowel	**Formant**	**Female speakers**	**Male speakers**	**Ratio F/M**
[i]	**F1**	302	294	1,03
	σ	*13*	*12*	
	F2	2530	1969	1,29
	σ	*120*	*99*	
	F3	3523	2999	1,17
	σ	*129*	*105*	
[a]	**F1**	635	570	1,11
	σ	*67*	*59*	
	F2	1927	1599	1,20
	σ	*89*	*94*	
	F3	2884	2504	1,15
	σ	*133*	*134*	
[u]	**F1**	312	305	1,02
	σ	*11*	*11*	
	F2	1023	980	1,04
	σ	*157*	*119*	
	F3	2585	2135	1,21
	σ	*210*	*100*	

A two factor ANOVA ("vowel" and "speaker's gender") was performed for F1 values in French speakers. There is moderate but significant overall effect of the speaker's gender: $F(1,264) = 33.464$; $p < 0.0001$. This cross-gender difference is significant for each vowel taken separately, with $p < 0.0001$ for [a] and $p < 0.01$ for [i] and [u].

For the second formant, similar tendencies were found, with a strong and significant overall effect of the speaker's gender: $F(1,264) = 488.791$;

$p < 0.0001$. Female/male difference appeared to be significant for [i] and [a], with $p < 0.0001$, but not for [u]: $p = 0.1454$.

The same analysis was performed for F3 values. It showed a very significant global effect of the speaker's gender: $F(1,264) = 701.907$; $p < 0.0001$. This cross-gender difference is significant for each vowel taken individually, with $p < 0.001$ in all cases.

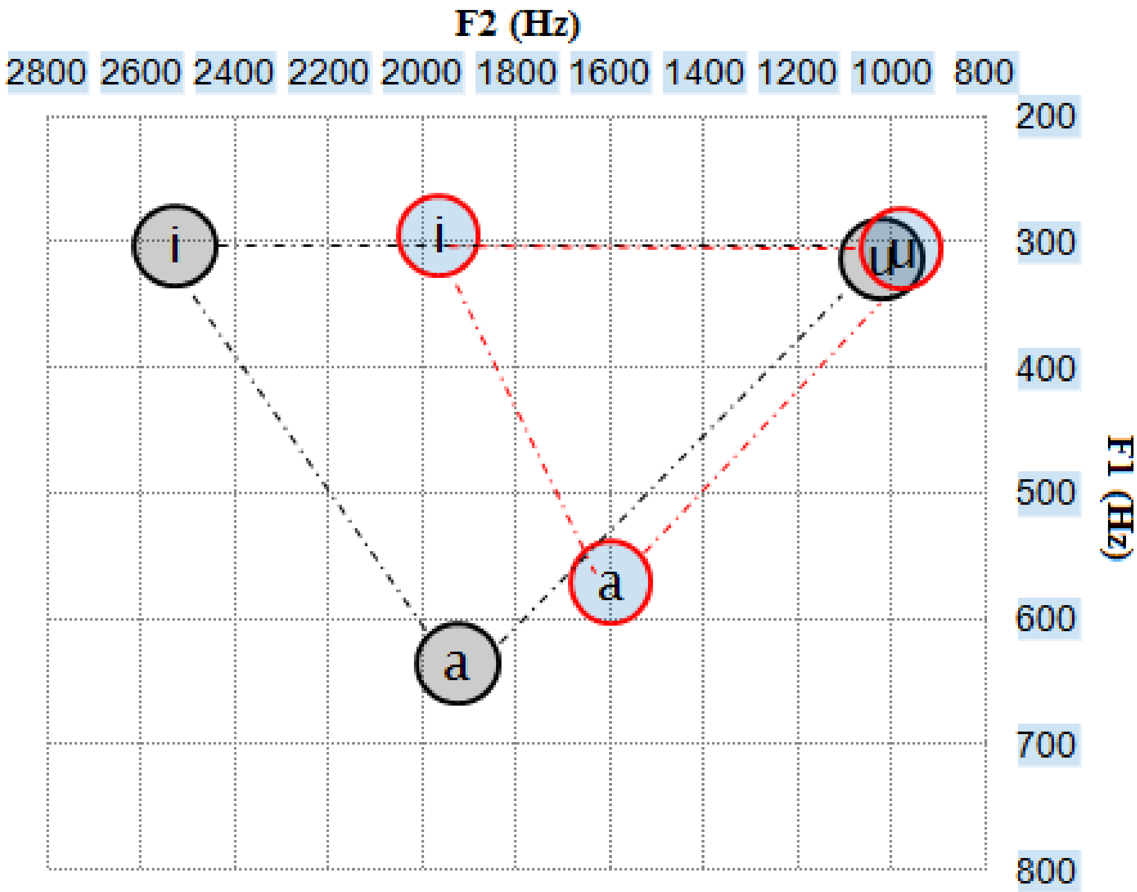

Figure 2. Vowel chart representing mean vowel formant frequencies (Hz) for vowels [i] [a] and [u] produced by female and male Parisian French speakers.

Results for the two first formants are presented in a vowel chart (Figure 2), using a template from *SaRP* software.

2.2 Voiceless consonants – Spectral centre of gravity

American English speakers

Spectral centre of gravity (Hz) of initial voiceless consonants produced by American English speakers is presented in Table 3, as a function of speaker's gender.

Table 3. Mean spectral centre of gravity (Hz) for consonants [s] [ʃ] and [tʰ] and [kʰ] produced by female (n = 5) and male (n = 5) American English speakers.
SD among the 5 measurements (1 occurrence * 5 speakers) per context for each consonant is also mentioned

		Mean centre of gravity (Hz)		
Cons.	**Context**	**Female speakers**	**Male speakers**	**Ratio F/M**
[s]	C + [æ]	9599	8638	1,11
	σ	*1317*	*815*	
	C + [i:]	9812	8328	1,18
	σ	*1496*	*1380*	
	C + [u:]	9665	8464	1,14
	σ	*1300*	*970*	
	All	**9692**	**8477**	**1,14**
[ʃ]	C + [æ]	5360	4460	1,20
	σ	*496*	*284*	
	C + [i:]	5609	4471	1,25
	σ	*497*	*236*	
	C + [u:]	5645	4650	1,21
	σ	*579*	*940*	
	All	**5538**	**4527**	**1,22**
[tʰ]	C + [æ]	4528	3826	1,18
	σ	*2455*	*2041*	
	C + [i:]	5886	5234	1,12
	σ	*1113*	*1683*	
	C + [u:]	5710	4612	1,24
	σ	*2182*	*1950*	
	All	**5374**	**4557**	**1,18**
[kʰ]	C + [æ]	2595	2376	1,09
	σ	*1104*	*1578*	
	C + [i:]	4535	4047	1,12
	σ	*2394*	*1377*	
	C + [u:]	1787	1689	1,06
	σ	*392*	*844*	
	All	**2972**	**2704**	**1,10**

A two factor ANOVA ("consonant" and "speaker's gender") was performed for centre of gravity values. All consonants taken together, it showed a significant global effect of the speaker's gender: $F(1,112) = 10{,}011$; $p < 0.01$. This cross-gender difference is significant for [s] ($p < 0.01$) and [ʃ] ($p < 0.0001$) but not for [tʰ] ($p = 0.2502$) and [kʰ] ($p = 0.6112$).

Parisian French speakers

Mean centre of gravity (Hz) of initial voiceless consonants for Parisian French speakers as a function of speaker's gender is presented in Table 4, below.

Table 4. Mean spectral centre of gravity (Hz) for consonants [s] [ʃ] and [t] and [k] produced by female (n = 5) and male (n = 5) Parisian French speakers.
SD among the 5 measurements (1 occurrence * 5 speakers) per context for each consonant is also mentioned

		Mean centre of gravity (Hz)		
Cons.	**Context**	**Female speakers**	**Male speakers**	**Ratio F/M**
[s]	C + [a]	9360	8013	1,17
	σ	*685*	*241*	
	C + [i]	9463	7914	1,20
	σ	*475*	*418*	
	C + [u]	9247	7755	1,19
	σ	*887*	*338*	
	All	**9414**	**7858**	**1,20**
[ʃ]	C + [a]	5333	4347	1,23
	σ	*951*	*428*	
	C + [i]	5731	4877	1,18
	σ	*969*	*317*	
	C + [u]	5068	3888	1,30
	σ	*1376*	*795*	
	All	**5417**	**4331**	**1,25**
[t]	C + [a]	3343	2199	1,52
	σ	*1799*	*313*	
	C + [i]	7397	6569	1,13
	σ	*859*	*399*	
	C + [u]	5034	3364	1,50
	σ	*866*	*731*	
	All	**5258**	**4044**	**1,30**
[k]	C + [a]	4636	3068	1,51
	σ	*1272*	*626*	
	C + [i]	6085	5224	1,16
	σ	*1847*	*939*	
	C + [u]	1154	1019	1,13
	σ	*101*	*552*	
	All	**3959**	**3104**	**1,28**

A two factor ANOVA ("consonant" and "speaker's gender") was conducted for centre of gravity date in French speakers. There is a very significant global effect of the speaker's gender: $F(1,112) = 16.130$; $p < 0.0001$. Contrary to English speakers, Female-male difference is here significant for each consonant: [s] ($p < 0.0001$), [ʃ] ($p < 0.01$) [t^h] ($p < 0.01$) and [k^h] ($p < 0.05$).

3. Conclusion – Discussion

This acoustic analysis has revealed several cross-gender acoustic differences, for both English and French speakers. As expected, higher resonant frequencies were found for female speakers. More interestingly, some cross-language variations were observed.

Regarding vowel formants, female/male differences on F1 appeared to be much greater in American English than in Parisian French speakers. Furthermore, there was a very large and significant cross-gender difference on F2 for vowel [u:] in American English speakers, whereas no significant difference was found between female and male French speaker on F2, for vowel [u]. This tendency could be accounted for by compensatory gestures in French female speakers when producing this vowel, in order to reach a low F2 frequency (Fant 1966; Fant 1975). On the other hand, cross-gender differences on F2 appeared to be larger in French speakers, for front vowels ([i]/[i:] and [a]/[æ]). Regarding consonant noise, which was measured through spectral centre of gravity, female-male differences were greater in French than in English speakers. This was true for each tested consonant. Such tendency is consistent with previous findings by Arnold (2012) and Pépiot (2011; 2013; 2014a), indicating that French listeners rely more on resonant frequencies than American English listeners when trying to identify a speaker's gender.

Overall, these results clearly support the idea that cross-gender acoustic differences are language dependent, as suggested by Johnson (2005) and Pépiot (2014b). Therefore, even if anatomical and physiological elements play a role in female/male acoustic differences, it is very likely that such differences are partly *socially constructed* (Johnson 2005; Johnson 2006; Simpson 2009). These data could be of interest for improving vocal rehabilitation of transgender people (Wiltshire 1995) and may also be useful in automatic speech processing and forensic phonetics.

Nonetheless, such data should be interpreted with caution. Five men and 5 women were recorded for each language. Despite the restrictive selection criteria and the small intra-gender variation, it seems quite difficult to generalize the results to the whole Parisian French and American English speakers' population. The present study could hence be replicated with a greater number of participants.

REFERENCES

Arnold 2012: Arnold, A. 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*, 2012, № 28, 1–18.

Boë et al. 1975: Boë, L-J., Contini, M., & Rakotofiringa, H. Étude statistique de la fréquence laryngienne. // *Phonetica*, 1975, № 32, 1–23.

Di Cristo 1999: Di Cristo, A. Vers une modélisation de l'accentuation du français : première partie. // *Journal of French Language Studies*, 1999, № 9, 143–179.

Ericsdotter 2005: Ericsdotter, C. *Articulatory-acoustic relationships in Swedish vowel sounds*, PhD Thesis, 2005, Stockholm University.

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, № 81, 21–38.

Fant 1975: Fant, G. Non-uniform vowel normalization. // *Speech Transmission Laboratory, Quarterly Progress and Status Report*, № 16, 1–19.

Fox & Nissen 2005: Fox, R. A. & Nissen, S. L. Sex-related acoustic changes in voiceless English fricatives. // *Journal of Speech, Language, and Hearing Research*, 2005, № 48, 753–765.

Hillenbrand et al. 1995: Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. Acoustic characteristics of American English vowels. // *Journal of the Acoustic Society of America*, 1995, № 97, 3099–3111.

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

Johnson 2006: Johnson, K. Resonance in an exemplar-based lexicon: the emergence of social identity and phonology. // *Journal of Phonetics*, 2006, № 34, 485–499.

Jongman et al. 2000: Jongman, A., Wayland, R., & Wong, S. Acoustic characteristics of English fricatives. // *Journal of the Acoustic Society of America*, 2000, № 108, 1252–1263.

Kahane 1978: Kahane, J. A. morphological study of the human prepubertal and pubertal larynx. // *American Journal of Anatomy*, 1978, № 151, 11–20.

Nikolov et al. 2011: Nikolov, R., Ditcheva, M., Dommergues, J.-Y., Herry-Bénit**,** N. & Ryst, E. (2011). *SaRP: Speech Analyzer Rapid Plot (2006–2011) Plotting vowels in F2-F1 scatter charts with multiple data sets*. http://web.uni-plovdiv.bg/rousni/sarp

Nittrouer 1995: Nittrouer, S. Children learn separate aspects of speech production at different rates: Evidence from spectral moments. // *Journal of the Acoustic Society of America*, 1995, № 97, 520–530.

Nordstrom 1977: Nordstrom, P-E. Female and infant vocal tracts simulated from male area functions. // *Journal of Phonetics*, 1977, № 5, 81–92.

Pépiot 2011: Pépiot, E. 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*, 2011, № 49, 418–430.

Pépiot 2013: Pépiot, E. *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.). PhD Thesis, 2013, University Paris 8.

Pépiot 2014a: Pepiot, E. 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. // *Actes du Congrès Mondial de Linguistique Française 2014 – Berlin*, 1365–1379.

Pépiot 2014b: Pepiot, E. Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. // *Proceedings of the 7th International Conference on Speech Prosody – Dublin*, 305–309.

Schwartz 1968: Schwartz, M. F. Identification of speaker sex from isolated voiceless fricatives. // *Journal of the Acoustical Society of America*, 1968, № 43, 1178–1179.

Shadle et al. 1991: Shadle, C. H., Badin, P. & Moulinier, A. Towards the spectral characteristics of fricative consonants. // *Proceedings of the 12th International Congress of Phonetic Sciences – Université de Provence*, 1991, 42–45.

Simpson 2009: Simpson, A. P. Phonetic differences between male and female speech. // *Language and Linguistics Compass*, 2009, № 3, 621–640.

Takefuta et al. 1975: Takefuta, Y., Jancosek, E. G., & Brunt, M. 1975. 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 2001: Whiteside, S. P. Sex-specific fundamental and formant frequency patterns in a cross-sectional study. // *Journal of the Acoustic Society of America*, 2001, № 110, 464–478.

Wiltshire 1995: Wiltshire, A. Not by pitch alone: a view of transsexual vocal rehabilitation. // *National Student Speech Language Hearing Association Journal*, 1995, № 22, 53–57.