Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Do speakers show different F0 when they speak in different languages? The case of English, French and German

Sandra Schwab[1, 2], *Jean-Philippe Goldman*[3]

[1] Institut für Vergleichende Sprachwissenschaft, Universität Zürich, Switzerland
[2] Ecole de langue et de civilisation françaises, Université de Genève, Switzerland
[3] Département de linguistique, Université de Genève, Switzerland
sandra.schwab@uzh.ch, jean-philippe.goldman@unige.ch

Abstract

Based on the facts that the voice quality that allows the recognition of a speaker is characterized, among other features, by his/her fundamental frequency (F0) and that F0 may differ across languages, we investigated, in the present research, whether speakers show different F0 when they speak in two different languages. To do this, we carried out a study with a within-speaker design, in which long-term distributional (LTD) F0 level and span measures were examined in early or late bilingual speakers of English and French, of English and German, and of French and German.

The results are the following: English-French speakers presented a lower F0 in English than in French. Along the same line, English-German speakers showed a lower F0 in English than in German. Moreover, they showed more variability in English than in German, especially when English was the speakers' mother tongue. Finally, French-German showed no differences in F0 level or span between both languages. These findings, which are partially in agreement with previous studies, not only highlight the advantage of using a within-speaker design in order to neutralize individual differences, but they also support the idea that the language spoken by the speaker is important for his/her identification.

Keywords: Long-term distribution F0, within-speaker design, English, German, French.

1. Introduction

Do speakers speak with the same voice in different languages? This is the question we aim at answering in this research. This question is particularly relevant given the fact that more than half of the world's population uses two or more languages (or dialects) in everyday life ([12]). Within the framework of forensic phonetics, this issue also has important consequences, since the voice differences that are observed across languages within a same speaker might constitute an obstacle to the recognition of that speaker.

Based on the facts that the voice quality of a speaker (i.e. the characteristic auditory "coloring" of the speaker's voice) is characterized, among other features, by his/her fundamental frequency (F0), on the one hand, and, on the other hand, that F0 may differ across languages (e.g., [21]), one may wonder whether the same speaker presents different F0 when he/she speaks in two different languages.

Two aspects are generally taken into consideration in the examination of the speaker's F0: its level (i.e., the height of F0) and its span (i.e., the range of F0) ([19]). In cross-linguistic studies, these two aspects are examined by means of long-term distributional (LTD) measures based on the analysis of the F0 distribution within the speaker's speech. For example, among the LTD measures of F0 level are the mean F0 and median F0, and among LTD measures of F0 span are the standard deviation of the values of F0, the F0 range, etc. (see for example, [2], [6], [14], [15], [21]). The examination of LTD level and span measures is complementary, since a speaker can be characterized not only by his/her F0 level (i.e. how high or low his/her F0 is), but also by the variability in his/her F0 (i.e. how much he/she varies his/her F0).

To our knowledge, very few cross-linguistic studies have described the F0 differences between English, German and French, which are the three languages under study in the present research. We selected these three languages because of the similarities and differences between them. English and German, both Germanic languages, are closely-related, sharing (besides morphological and syntactic properties) phonetic, rhythmic and accentual characteristics (i.e., vowel reduction, syllabic complexity, stress-timed rhythm, free lexical stress), while French, a Romance language, is distantly-related from English and German, since it presents, for example, no vowel reduction, a simpler syllabic structure, a syllable-timed rhythm and a fixed stress.

In the few studies dealing with English, German and French, the speakers of the various languages were different; in other words, a between-speaker design was used. For example, [6] investigated the mean F0 (among other measures) in speakers of English, French, German, Italian, Spanish. They found differences between all languages, but especially between German and French (i.e., from lowest to highest F0: German, Spanish, Italian, English, French). [21] examined various LTD F0 level and span measures (besides linguistic measures; [22]) in speakers of English and speakers of German and found no differences in F0 level measures, but differences in span measures, with larger span in English than in German. [2] investigated LTD F0 level and span measures in Slavic (Polish and Bulgarian) and Germanic languages (English and German). They came to the conclusion that, while Germanic languages differ from Slavic languages in F0 level and F0 span, English and German do not significantly differ. Nevertheless, due to the between-subject design used in these experiments (and despite the highly controlled selection of the speakers), individual (i.e., physiological) differences cannot be ruled out to account for the results.

To solve this issue, a small number of researchers have used a within-speaker design in the examination of cross-language F0 differences (e.g. [1], [11], [20]), but, to our knowledge, none of them have dealt with English, German and French.

The objective of the present preliminary study is to investigate whether speakers show different F0 when they speak in different languages. To do this, we carried out a study with a within-speaker design, in which the same speaker (early or late bilingual) was recorded in two languages. Various LTD F0 level and span measures were examined in the productions of three groups of speakers: 1) speakers that produced sentences in English and French; 2) speakers that produced sentences in English and German; 3) speakers that produced sentences in French and German. It is important to keep in mind that, within each group, the same speakers produced the sentences in both languages, hence the within-speakers design.

 doi: 10.21437/SpeechProsody.2016-2

2. Material

The material we used in this research consisted of speech samples extracted from the SIWIS database ([8]). The goal of the SIWIS project is to set up a framework for speech-to-speech translation for English, French, German and Italian with the use of statistical speech synthesis and recognition systems and with cross-language speaker adaptation techniques for HMM-based speech synthesis as in [23]. The detection and the generation of accents were also developed in the SIWIS project, in order to convey more prosodic information throughout the speech-to-speech translation chain.

Within this framework, speakers of various languages, including not only early but also late bilingual or trilingual speakers, were recruited on the basis of an evaluation of the degree of their accentedness in the different languages. The speakers were evaluated for each language by three native speakers who judged the degree of accent on a 0-3 scale (0 = strong foreign accent, 1 = noticeable accent, 2 = very slight accent, 3 = no foreign accent). Only candidates with a minimum averaged score of 2.67 were selected for the SIWIS project, as they were considered as speakers with no foreign accent. The speakers also indicated their level in each language (A = native, B = active, C = passive).

Speaker	Gender	Age	EN	FR	DE
1	H	25	A (UK)	A (CH)	
20	H	22	A (UK)	A (CH)	
2	F	22	A	A (CH)	
9	F	22	A	A	
10	H	32	A (UK)	A (FR)	
18	F	23	B	A	
22	F	21	B	A	
6	F	22	A (CA)	A	A
13	F	25	A (UK)	A (CH)	A
12	F	47	A (UK)	A (FR)	C (DE)
3	H	56	A	B	B
11	F	23	B	A	B
5	F	24	B	B (FR)	A (DE)
4	F	25	C	B	A
7	F	28	B (USA)		A (CH)
8	F	54		B	A
15	F	29		B (CH)	A (AT)

Table 1. *Speakers' gender, age, language level (A = native, B = active, C = passive) and regional variety.*

Speakers were instructed to read carefully at a normal rate each of the sentences that appeared on the computer screen. In the case of errors, they were asked to repeat the sentence. They began with the recordings in their weakest language(s) (B or C) and finished it in their native language(s) (A).

Since the present study focuses on English, German and French, we selected, among the SIWIS speakers, 14 English-French speakers, 8 English-German speakers and 9 French-German speakers. Among these speakers, 7 spoke English, German and French. Speakers' gender, age and language level are presented in Table 1, as well as their regional variety, when this information was available.

Among all the recordings performed within the framework of the SIWIS project [8], we selected only declarative sentences taken from Europarl statements [16] and declarative sentences taken from the journalistic texts (which represented between 99 and 125 sentences per speaker and per language). The duration of the each sentence was between 0.8 sec to 12 sec, and the total duration of the corpus used in this study was 290 min (7 min per speaker and per language, in average).

3. Data analysis

F0 values were extracted in semi-tones with Praat ([5], "semitones re 1 Hz") using the Hirst algorithm ([13]) in order to avoid some Praat F0 detection errors. For each sentence of each speaker, F0 value was extracted every 10ms. On the basis of all these values, various long-term distribution (LTD) F0 measures were computed. Regarding the level, we calculated the mean F0 and median F0. As for the span, we computed the standard deviation (SD), the maximum-minimum range, the 90% range and the Fischer coefficient of skewness and kurtosis[1] of F0. Then, for each speaker, we inspected the distribution of each measure (the two languages grouped together) in order to remove extreme outliers from the data (i.e., datapoints inferior to Quartile1-(2*1.5*Interquartile Range) or superior to Quartile3+(2*1.5* Interquartile Range); 0.05%-1-05% of the datapoints for the different parameters).

We ran separate analyses for the English-French speakers, for the English-German speakers, and for French-German speakers. We analyzed the data by means of mixed-effects regression models ([4]) in R software (version 3.1.3). The random effects of the model included random intercepts for participants and sentence as well as random slopes allowing the effect of the language to differ across participants. The fixed effects of the models included the language, and, given that the speakers were males and females and had different language levels in both languages (see Table 1), we also included the interaction between Language and Gender and the interaction between Language and Level[2] into the initial models, in order to

[1] A positive skewness coefficient reflects a right-skewed distribution (i.e. most values are concentrated below the mean, with extreme values to the right), while a negative skewness coefficient reflects a left-skewed distribution (i.e. most values are concentrated above the mean, with extreme values to the left). A positive kurtosis coefficient indicates a leptokurtic distribution, (i.e. a sharper than a normal distribution), whereas a negative kurtosis coefficient indicates a platykurtic distribution (i.e. flatter than a normal distribution with a wider peak).

[2] Level was coded as followed: for English-French language pair: AA = 2 mother tongues; AB = dominance in English, BA = dominance in French; BB = same dominance in both languages (but not mother tongues); CB = dominance in French (but not mother tongue). For English-German language pair: AA = 2 mother tongues; AB = dominance in English; BA = dominance in German and BB = same dominance in both languages (but not mother tongues). For French-German language pair: AA = 2 mother tongues; AB = dominance in

ensure that the effect of language was not modulated by the effect of gender or language level. If not significant, these interactions (and the simple terms[3]) were removed from the final models (and not commented in the Results sections). Significance was assessed using a p-value (from the Satterthwaite approximation for degrees of freedom implemented in the *lmerTest* package; [18]) below 0.05 for the main effects and a t-value above 1.96 for the estimates. Following [3], in order to ensure that the results in our final models were not driven by a few atypical data points, residuals larger than 2.5 times the standard deviation were considered outliers and removed.

4. Results and discussion

4.1. English-French

As can be seen in Table 2, for the 14 English-French speakers (6 males and 8 females), the LTD level measures and the coefficient of skewness differ significantly between the two languages (mean F0: β = 0.66; SE = 0.18; t = 3.65; F(1, 14) = 13.34 p < .001; median F0: β = 0.80; SE = 0.18; t = 4.37; F(1, 13) = 19.12, p < .001; skewness: β = -0.15; SE = 0.04; t = -3.52; F(1, 14) = 12.39, p < .01). The speakers' F0 is lower in English than in French[4], and their skewness coefficient is higher in English than in French. In other words, English F0 is lower than French F0, but with more very high values in the right tail of the distribution.

Measure	English	French
LTD Level		
Mean F0	**87.90 (4.93)**	**88.55 (4.86)**
Median F0	**87.64 (5)**	**88.40 (4.94)**
LTD Span		
SD	2.85 (0.84)	2.72 (0.7)
Max-Min	13.33 (3.06)	12.95 (2.7)
90% range	9.37 (2.72)	8.99 (2.28)
Skewness	**0.36 (0.42)**	**0.21 (0.39)**
Kurtosis	-0.15 (0.85)	-0.22 (0.63)

Table 2. *F0 measures (in semi-tones; with standard deviations in brackets) for the English-French speakers. The parameters that present a significant difference between both languages are in bold*.

4.2. English-German

Table 3 presents the results for the 8 English-German speakers (2 males and 6 females). As to the LTD level measures, mean F0 is marginally lower in English than in German (β = -0.35; SE = 0.16; t = -2.15; F(1, 7) = 4.63, p = .07), whereas median F0 is significantly lower in English than in German (β = -0.51; SE = 0.18; t = -2.90; F(1, 7) = 8.37, p < .05).

Measure	English[5]	German
LTD Level		
Mean F0	89.59 (4.16)	89.94 (4.25)
Median F0	**89.27 (4.29)**	**89.78 (4.39)**
LTD Span		
SD	**2.84 (0.87)**	**2.77 (0.83)**
Max-Min	13.12 (3.20)	12.92 (3.06)
90% range	**9.32 (2.82)**	**9.08 (2.65)**
Skewness	**0.37 (0.39)**	**0.20 (0.41)**
Kurtosis	-0.21 (0.69)	-0.26 (0.65)

Table 3. *F0 measures (in semi-tones; with standard deviations in brackets) for the English-German speakers. The parameters that present a significant difference between both languages are in bold.*

Regarding the LTD span measures, although English globally presents a higher SD (i.e., more variability) than German (β = 0.25; SE = 0.05; t = 5.00; F(1, 11) = 13.00, p < .01), there is more variability in the speakers' mother tongue (i.e., English or German, see Table 1) (interaction Language x Level: F(3, 7) = 6.64, p < .01)[6]. Along the same line, speakers present a wider 90% range in English than in German (β = 0.84; SE = 0.14; t = 5.70; F(1, 72) = 19.48, p < .001), especially when English is their mother tongue (interaction Language x Level: F(3, 43) = 24.52, p < .001)[7]. Finally, the coefficient of skewness is higher in English than in German (β = 0.18; SE = 0.05; t = 3.23; F(1, 7) = 10.46, p < .05), indicating that their F0 distribution is more right-skewed in English than in German.

4.3. French-German

Table 4 presents the results for the 9 French-German speakers (2 males and 7 females). None of the seven F0 measures present significant differences between French and German.

Measure	French	German
LTD Level		
Mean F0	89.95 (3.92)	89.9 (3.98)
Median F0	89.81 (4.08)	89.76 (4.13)
LTD Span		
SD	2.63 (0.73)	2.76 (0.79)
Max-Min	12.49 (2.81)	12.88 (2.94)
90% range	8.65 (2.37)	9.02 (2.53)
Skewness	0.18 (0.4)	0.19 (0.4)
Kurtosis	-0.27 (0.56)	-0.29 (0.61)

Table 4. *F0 measures (in semi-tones; with standard deviations in brackets) for the French-German speakers.*

French; BA = dominance in German and BB = same dominance in both languages (but not mother tongues).

[3] Since we used a within-speaker design in which we compared the productions of the same speaker in two languages, the main effects of Gender and Level were not of interest in the present study.

[4] For example, a F0 difference of 0.76 semi-tones is equivalent to a 4.5 Hz difference at 100 Hz and to a 9 Hz difference at 200 Hz.

[5] The differences between the English LTD measures in the English-French and in the English German analyses (for example, mean F0 of 87.90 semi-tones Hz and 89.59 semi-tones, respectively) are due to the fact that males are more numerous in the English-French analysis than in the English-German analysis.

[6] The F0 SD (in semi-tones) for English and German respectively is as follows: AA: 2.78 and 2.53; AB: 3.77 and 3.60; BA: 2.49 and 2.6; BB: 2.15 and 2.06.

[7] The 90% range (in semi-tones) for English and German respectively is as follows: AA: 9.25 and 8.36; AB: 12.30 and 11.69; BA: 8.15 and 8.49; BB: 7.12 and 6.81.

5. General discussion

The objective of this research was to determine whether speakers show different F0 when they speak in different languages. For this, we examined various LTD F0 level and span measures in early or late bilingual speakers of English and French, of English and German and of French and German.

As far as the English-French speakers are concerned, the results showed, on the one hand, that their F0 was lower in English than in French, which is in agreement with [6]. As to the English-German speakers, the results showed lower F0 in English than in German, a finding that does not agree with [2] and [21]. On the other hand, they showed more variability in English than in German, what is in agreement with [21], especially when English was the speakers' mother tongue. Finally, as far as the French-German speakers are concerned, results showed no differences in F0 level or span, and therefore do not support the results reported by [6]. Methodological issues might account for these discrepancies. As already mentioned, the experimental design used in the previous studies was between-speaker, whereas we used in the present preliminary experiment a within-speaker design, which allowed the neutralization of individual physiological differences.

In this respect, the fact that English not only differs from French (i.e., a distantly-related language with a very different prosodic system), but also from German (i.e., a closely-related language with a similar prosodic system) and the fact that French and German (i.e., two distantly-related languages with very different prosodic systems) do not differ are difficult to explain. One could think that the presence of glottalizations in English (e.g. [7]), which present a lower F0, might be responsible for the lower English F0, but glottalizations are also common in German ([17]). Another explanation involves creaky voice. It is known that "creaky voice is associated with lowered fundamental frequency values" ([10], p. 400), and also that (especially female) speakers of English employ creaky voice to a large extent (e.g., [24]). Having this in mind, we can hypothesize that the finding of a lower F0 in English than in German or French may be due to more presence of creaky voice in English than in the other two languages.

Regarding the language level, its effect on the difference between languages (i.e., interaction Language x Level) concerned only some LTD span measures of English-German speakers, but not LTD F0 level measures: the variability was larger in the mother tongue. It seems thus that the language level does not have an impact on how high or low the speakers' F0 is, but rather on how much the speakers are able to vary their F0. Nevertheless, these results should be considered with precaution, since the language level was not entirely controlled in the two languages.

In conclusion, this research shows that a speaker may not present similar F0 in the different languages he/she speaks, especially when one of the two languages is English. This finding has direct implications for speaker identification in forensic phonetics, and supports the idea that the language spoken by the individual should be taken into account in speaker recognition.

Further research is needed to explore more deeply these preliminary findings, especially the possible reasons for the lower English F0. In particular, we will include in our analysis linguistic measures, as proposed in [22], and we will increase the number of speakers in our within-speaker design study with similar language level in both languages.

6. AKNOWLEDGMENTS

This study is part of the SIWIS project "Spoken Interaction with Interpretation in Switzerland" financed by the Swiss National Science Foundation (n°141903). Thanks to Lorraine Baqué and Hanna Ruch for their helpful comments on an earlier version of this manuscript. This research was partially carried out during an *Ambizione* grant period financed by the Swiss National Science Foundation (PZ00P1_148036).

7. References

[1] Altenberg, E., Ferrand, C. 2006. Fundamental frequency in monolingual English, bilingual Russian/English, bilingual Chinese/English, young adult women. *Journal of Voice* 20, 89-96.

[2] Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Jastrzebska, M., Trouvain, J. 2014. Comparison of Pitch Range and Pitch Variation in Slavic and Germanic Languages. *Proc. Speech Prosody 2014*, 20-23 May, Dublin, Ireland, 776-780.

[3] Baayen, H. (2008). *Analysing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press

[4] Baayen, R. H., Davidson, D. J., Bates, D. M. 2008. Mixed effects modeling with crossed random effects for subjects and items, *Journal of Memory and Language* 59: 390-412

[5] Boersma, P., Weenink, D. 2011. Praat: doing phonetics by computer (Version 5.2). www.praat.org.

[6] Campione, E., Véronis, J. 1998. A statistical study of pitch target points in five languages, *5th International Conference on Spoken Language Processing (ICSLP'98)* (pp. 1391-1394). Sidney.

[7] Dilley, L., Shattuck-Hufnagel, S. and Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24: 423-444.

[8] Garner, P. N., Clark, R., Goldman, J.-P., Honnet, P.-E., Ivanova, M., Lazaridis, A., Liang, H., Pfister, B., Ribeiro, M. S., Wehrli, E., Yamagishi, J. 2014. Translation and Prosody in Swiss Languages in A. Auchlin, T. Prsir (eds.) *SWIP 3 - Swiss Workshop In Prosody*, Cahiers de Linguistique Française, 31, 211-221.

[9] Goldman, J.-P. 2011. EasyAlign: an Automatic Phonetic Alignment Tool under Praat. *Proceedings of Interspeech*, 3233–3236.

[10] Gordon, M. & Ladefoged, P. (2001). Phonation Types: A Crosslinguistic Overview. *Journal of Phonetics* 29: 383-406.

[11] Graham, C. 2013. Revisiting f0 Range Production in Japanese-English Simultaneous Bilinguals. *Annual Report of UC Berkeley Phonology Lab* 110–125.

[12] Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, Mass: Harvard University Press.

[13] Hirst, D. 2011. The analysis by synthesis of speech melody: From data to models, *Journal of Speech Sciences* 1(1): 55-83.

[14] Keating, P., Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *Journal of the Acoustical Society of America* 132, 1050-1060.

[15] Kinoshita, Y., Ishihara, S., Rose, P. 2009. Exploring the discriminatory potential of F0 distribution parameters in traditional forensic speaker recognition. *The International Journal of Speech, Language and the Law* 16(1), 91-111.

[16] Koehn, P. 2005. Europarl: A parallel corpus for statistical machine translation, *MT Summit* 2005.

[17] Kohler, K. J. (1994). Glottal stops and glottalization in German, *Phonetica*, 51, 38-51.

[18] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-20. http://CRAN.R-project.org/package=lmerTest.

[19] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

[20] Marquina, M. 2011. *Estudio acústico de la variación inter e intralocutor en la frecuencia fundamental de hablantes bilingües de catalán y de castellano*. Trabajo de investigación de doctorado. Universitat Autònoma de Barcelona.

[21] Mennen, I., Schaeffler, F., Docherty, G. 2012. Cross-language differences in fundamental frequency range: A comparison of English & German. *Journal of the Acoustical Society of America* 131 (3), 2249-2260.

[22] Patterson, D. 2000. *A linguistic approach to pitch range modelling*, Ph.D. thesis, University of Edinburgh, Edinburgh.

[23] Yamagishi, J., Nose, T., Zen, H., Zhen-Hua, L., Toda, T., Tokuda, K., King, S., Renals, S. 2009. A robust speaker-adaptive HMM-based text-to-speech synthesis. *IEEE TSALP*, 17(6), 1208–1230.

[24] Yuasa, I. P. (2010). Creaky voice: a new feminine voice quality for young urban upwardly mobile American women? *American Speech Fall 2010 85(3)*: 315-337.