

Published in: Campbell, Nick/Gibbon, Dafydd/Hirst, Daniel (eds.): Proceedings to the 7th Speech Prosody Conference. Trinity College Dublin, Ireland. May 20-23, 2014. - Baixas: International Speech Communication Association, 2014. Pp. 1037-1041.

## Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers

Frank Zimmerer, Jeanin Jügler, Bistra Andreeva, Bernd Möbius, Jürgen Trouvain

Department of Computational Linguistics and Phonetics, Saarland University, Germany

{zimmerer|juegler|andreeva|moebius|trouvain}@coli.uni-saarland.de

### Abstract

This article presents preliminary results indicating that speakers have a different pitch range when they speak a foreign language compared to the pitch variation that occurs when they speak their native language. To this end, a learner corpus with French and German speakers was analyzed. Results suggest that speakers indeed produce a smaller pitch range in the respective L2. This is true for both groups of native speakers. A possible explanation for this finding is that speakers are less confident in their productions, therefore, they concentrate more on segments and words and subsequently refrain from realizing pitch range more native-like. For language teaching, the results suggest that learners should be trained extensively on the more pronounced use of pitch in the foreign language.

**Index Terms:** pitch variation, L1, L2, language learning

### 1. Introduction

When learning a foreign language, especially as adults, it is extremely hard to reach native-like skills in phonetics and phonology of this language. One of the reasons for this hardship is that the phonetic and phonological knowledge of the native language (L1) can interfere with the phonetic and phonological system of the foreign language (L2) (e.g., among many more, [6], [14], [18]). For instance, in German, voiceless plosives are produced with a long Voice Onset Time (VOT), whereas in French, VOT for voiceless plosives is rather short. When producing stops in L2, German speakers usually do not adapt their production, and their voiceless French stops do not sound like a native French production would.

However, apart from segmental differences that are hard to be learned perfectly, a foreign accent might also occur due to prosodic interference from the native language in L2 (e.g. [2], [26], [27]). Languages have been shown to differ with respect to the pitch range they use, their exact pitch contours and the exact placement of pitch changes (e.g. [3], [4], [5], [7], [10], [11], [13], [17], [19], [21], [22]). A study by Mennen and colleagues [22] suggests, for example, that there are differences both in level as well as range for English and German speakers in the respective L1. This finding is supported by data presented by Andreeva and colleagues [3] who found differences in level and range for Bulgarian, English, German, and Polish speakers. Keating and Kuo [17] found several differences between English and Mandarin speakers in pitch level and range, also depending, for instance, on the task the speakers were engaged in. Some of the differences that occur between languages can also be partly explained by (socio-) cultural factors (e.g. [17], [30]).

The difficulty to reach native-like performance in the prosodic realization of an L2 is arguably aggravated by the fact that when speaking a foreign language pitch variation is apparently compressed compared to the pitch range that is

standard for native speakers (e.g. [9], [10], [16], [20], [29]), which can result in a foreign accent. Furthermore, the lack of correct pitch variation can lead to be perceived as speaking in a monotonous way (e.g. [15], [16]). One possible explanation for this compression is that L2 learners are less confident about speaking the foreign language, or that they focus on getting the segmental pronunciation and the placement of stress correctly before expanding the pitch range as native speakers do. For instance, Mennen [20] showed that Dutch (L1) speakers of Modern Greek failed to produce the same pitch range as native Greek speakers. Also, in a study investigating pitch range of Finnish (L1) speakers of Russian (L2), Ullakonoja [29] found smaller pitch ranges in the L2 production for these speakers compared to native Russian speakers. Furthermore, the results show that the pitch range was also different compared to the Finnish L1 productions, and that extended stays in Russia led to a larger pitch range, implying the learnability of this intonation feature. Similarly, Busà and colleagues [10] found (non significant) long-term distributional (LTD) differences in the production of Italian speakers of English compared to native English productions. In a study of Arabic native speakers a comparison of their native productions with productions in English (L2) did not show a language effect [1]. In this study, however, the speaker group was quite advanced and living in the L2 environment. Thus, the results maybe due to the higher L2 proficiency level of the speakers.

Other studies have assumed a compressed pitch range *a priori*, and focused on the improvement of pitch range suggesting that the compression of pitch range can be overcome with enhanced training methods (e.g. [7], [15], [16]).

This short overview suggests a clear trend for a compressed pitch range in L2 speech production. Moreover, it seems that training helps to decrease the degree of pitch range compression. However, it is not clear whether the finding of reduced pitch range in the L2 is an universal tendency or whether pitch range compression is dependent on the language pair under investigation. This paper contributes to the research on L2 pitch range by investigating whether learners of French and German compress their pitch range when speaking their non-native language (German and French), compared to their L1 productions, to find out whether there is a general trend for pitch compression in L2 production, irrespective of L1. The construction of the corpus allows for a direct comparison for each speaker in each of the languages, because the same speakers were recorded in both languages (see also section 2.2).

A second question that will be touched upon is the extent to which language learners are able to learn to suppress pitch compression, that is, we investigate whether advanced learners compress pitch range less than beginners. The results reported in this paper are preliminary: the number of speakers (7 per native language) is rather small, the number of male and

female speakers is not evenly distributed across the two native languages and the number of advanced learners and beginners is also not equal. Even so, stable effects found in the present study will provide a strong motivation for an investigation based on a larger number of speakers.

## 2. Methods and materials

Different LTD measures have been used to quantify pitch range differences in the past (e.g. [15], [17], [22]). The quantifications that have been analyzed range from linguistically defined tonal structures to different measures of  $F_0$  (e.g. in Hz or semitones) or a combination thereof. Furthermore, there is the overall level of pitch (usually calculated as mean value over time) and the range of variation within a given speech sample. In this paper we concentrate on the latter. It is outside the scope of this paper to discuss the (dis)advantages of one measure over another. We focus on the so-called Pitch Dynamism Quotient (PDQ), which allows for a normalized evaluation of pitch variation [15] (see also [25]) where it is called frequency modulation factor – and below, section 2.3.).

### 2.1. General corpus description

A bilingual learner corpus served as the basis for the analysis reported here. The corpus was created with French (L1) learners of German (L2) and German (L1) learners of French (L2) at the LORIA institute in Nancy, France and the institute of phonetics at Saarland University in Saarbrücken, Germany [28].

For the corpus, 7 speakers of each language were recorded. They were recorded both in the respective L2 and in their native language. This design allows for a within-subject comparison of the productions in the two languages. Furthermore, recording settings and the text data are comparable across the languages (as L1 or L2) and identical within a language (i.e. French texts read by both French and German participants and *vice versa*). Therefore, comparisons between individual speakers with different L1 can be drawn as well as differences between a group of speakers that differ with respect to what language they speak as native language. In each group (French and German native speakers) there were 5 beginners (A1-A2 level according to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)). Additionally, 2 advanced learners (B2-C1 Level) of the respective language were recorded. In the group of French native speakers, 6 male speakers and 1 female speaker participated (15-22 years, M: 20.1 years, SD: 2.4 years), whereas for the German native speaker group, 5 female speakers and 2 male speakers were recorded (15-26 years old, M: 20.6 years, SD: 4.2 years). The 3 teenage speakers (all 15 years of age, all had completed the change of the voice, 1 French and 2 German highschool students) being part of the corpus were all male. The data of all 14 speakers were used for the analysis in this study.

Recordings were made in quiet office rooms with head-mounted microphones, which were amplified and digitized (16kHz, 16 bit) in a M-AUDIO Fast Track USB device. Recordings were saved on Windows Laptop computers with a custom-made software that was developed at LORIA (“Corpusrecorder”, [12]). Each sentence (and each story) was saved as a separate audiofile.

### 2.2. Text materials for the analysis

The corpus data analyzed for this paper consisted of read sentences and read short stories [28]. Both groups of speakers read a set of 25 sentences (*sentence-condition*) in their respective L2, before reading two short stories (an advertising text about ecological economic development and the story of the three little pigs), also in the respective L2 (*story-condition*). Then, the same tasks (reading a set of 25 sentences and a translated version of the two stories, slightly longer in the German than in French version) were recorded in the respective L1 of the participants. The sentences were different in the two languages, but similar in content and length. Due to a technical defect, only 24 sentences were recorded for 1 native speaker of French in the L2. Otherwise, all audiofiles could be used for further analysis.

### 2.3. Pitch analysis

Pitch measurements were done in three steps. First,  $F_0$  was extracted automatically by means of the ESPTS algorithm (“get\_f0” [24]) from all files. For male speakers time steps of 0.01 seconds were used, whereas the computation of  $F_0$  for female speakers was done in time steps of 0.005 seconds.

Then, manual correction was performed with PRAAT [8] to exclude cases of octave jumps, wrong measurements (e.g. when the algorithm found voicing in silent intervals or mistakes due to creaky voice, or other artifacts produced by the algorithm that were not based on changes in  $F_0$ , but possibly distorting the results). Other microprosodic variation (e.g. due to stop realization) was retained. In a final step, the first two  $F_0$  values after an unvoiced segment and the last  $F_0$  value before an unvoiced segment were excluded. This was done because the vocal folds need some time to achieve their intended vibration rate (i.e. start or stop of vibration) and this kind of irregular vibration could also exaggerate the  $F_0$  variation. The resulting data was analyzed by means of the JMP software [23].

The number of female and male speakers differed strongly between the two groups of native speakers. This was one of the main reasons to investigate pitch variation with the PDQ as defined as the standard deviation divided by the  $F_0$  mean [15]. This measure normalizes  $F_0$  variation data and allowed us to concentrate on language differences while minimizing group differences at the same time. The lower the PDQ, the smaller is the variation.

Note, however, that due to the uneven distribution of male and female speakers across languages, comparisons between the native speaker groups (e.g. a claim like “French is generally more variable than German”) cannot be made despite the use of PDQ (and thus, the normalization). First of all, the preliminary analysis of seven speakers per native language is not enough for general, far reaching claims concerning two languages. But more importantly, pitch *range* has been found to differ between male and female speakers (e.g. [25]). Therefore, possible differences in the results between the speakers of the two languages cannot be attributed with certainty to language differences, the source could as well be a gender difference. However, on a general level, differences between German and French pitch range may be expected on the basis of prior research (e.g. [5], [19]). Koreman and colleagues [19] presented evidence that both French and German speakers use  $F_0$  to indicate accentuation, but they do so differently.

Therefore, what is possible, and this is done in this paper, is to investigate the pitch variation of the German group in the native and non-native language, as well as the same comparison for the French native speakers. The results will be reported in the next sections.

### 3. Results

Mean  $F_0$  and the standard deviations were calculated for every audiofile (henceforth “item”). Subsequently, the PDQ was calculated for these items. These PDQ values were the basis for all subsequent analyses.

Overall mean PDQ for French speakers in their L1 is 0.134 when reading sentences, and 0.142 when reading the stories. In their L2, French speakers produce a mean PDQ of 0.119 for the sentences and 0.13 for the stories. German speakers have a mean PDQ in their L1 of 0.142 for the sentences and 0.146 for the stories. In their French productions, the mean PDQ drops to 0.118 for the sentences and 0.133 for the stories. Figure 1 depicts the overall mean PDQ results for the two speaker groups in the respective task languages.

Subsequently, the PDQ values were entered into a linear mixed model with *PDQ* as dependent factor, *speaker* and *item* as random factors and *native language* (French/German), *task language* (French/German), and *task* (sentence/story) as independent factors, as well as all their possible interactions.

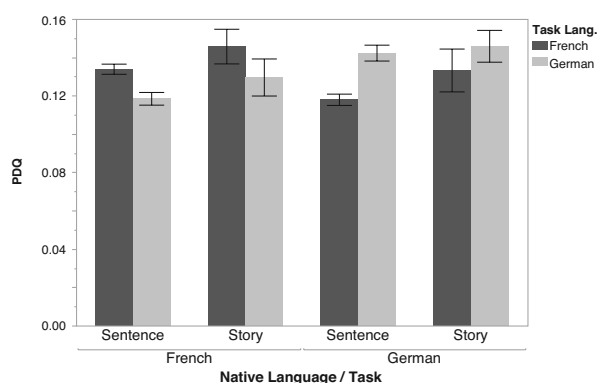


Figure 1: Mean PDQ for French and German speakers depending on task language and task.

The results of the statistical analysis indicate that there was a main effect of *task* ( $F(1,735)=5.52$ ,  $p<0.05$ ). Overall, speakers produced higher pitch variation when reading the stories compared to their sentence productions (mean PDQ for stories: 0.139; mean PDQ for sentences: 0.128). Also, the interaction of *native language* and *task language* was significant ( $F(1,735)=14.85$ ,  $p<0.0001$ ). A post-hoc planned comparison showed that for both language groups, the native productions were significantly higher in PDQ than the non-native productions (French L1:  $t(2)=-2.51$ ,  $p<0.05$ ; German L1:  $t(2)=-2.94$ ,  $p<0.01$ ).

No other factors or interactions reached significance. This first analysis suggests that irrespective of the L1,  $F_0$  variation is compressed in L2 productions. Also, the  $F_0$  range depends on the task in which participants were engaged. Reading

stories leads to a greater variation in pitch than reading short sentences, irrespective of native language or task language.

A second model was calculated to tap into the effect of proficiency. To this end, the first model was extended by the factor *proficiency* (Beginner/Advanced) and all its interactions. The results indicate that proficiency level was not a significant factor, nor were any of its interactions. Again, *task* ( $F(1,729)=5.66$ ,  $p<0.05$ ) and the interaction of *native language* and *task language* ( $F(1,729)=9.11$ ,  $p<0.01$ ) were significant. Figure 2 shows the PDQ for the different fluency levels. As can be seen, the overall results remain very similar. However, for the French advanced learners of German, the *story condition* shows a small trend into a different direction. Here, the PDQ for the stories in German is slightly higher than the PDQ for the stories read in French. Note that there were only two advanced speakers in both native speaker groups that read the two stories.

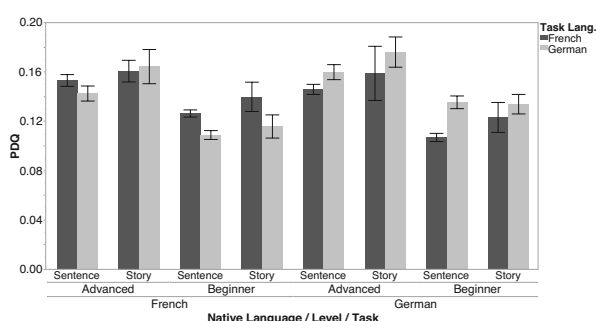


Figure 2: Mean PDQ for French and German speakers depending on task language, task and proficiency level of the learners.

Finally, we looked at individual speaker differences (Figure 3 for French native speakers and Figure 4 for German native speakers). The mean PDQ for French speakers speaking French shows a range from 0.102 to 0.162, whereas for the same speakers, the mean PDQ ranges in German from 0.078 to 0.171. For the German speakers, PDQ in German ranges from 0.083 to 0.194, whereas in the French production, the range lies between 0.069 and 0.164. This indicates that there is some individual variation, but that most speakers have a compressed  $F_0$  in their L2 productions.

However, it becomes also apparent that three speakers show a somewhat different behavior. In the French group, there were two speakers (503 and 501) where the L1 productions showed a smaller PDQ than the L2 productions. One of them (503) is an advanced learner, the other a beginner (501). In the German group, one speaker (006) had almost identical PDQ values for L1 and L2 productions. This speaker was an advanced learner of French. Due to the small number of speakers overall, further interpretation of these results is not possible.

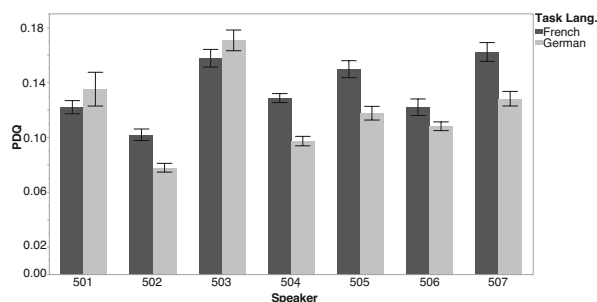


Figure 3: Mean PDQ of individual French speakers depending on task language. Speakers 503 and 505 are advanced learners.

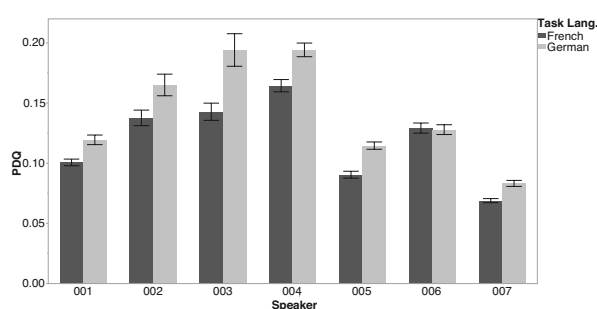


Figure 4: Mean PDQ of the individual German speakers depending on task language. Speakers 004 and 006 are advanced learners.

#### 4. Discussion & conclusions

The first question this article was set out to investigate is whether there is a general trend that French learners of German and German learners of French compress their pitch range when speaking L2. The results indicate that the pitch range (measured in PDQ) is indeed smaller in the L2 for both language groups, that is, irrespective of the native language. This means that the actual pitch range that is produced by native speakers when they speak their native language is larger than the pitch range in their L2. This result is compatible with results presented by other researchers who also found a compressed pitch range in L2 productions (e.g. [9], [10], [20], [29]).

At this point, we cannot give clear answers to the question why this is the case. However, we can give a speculative explanation. This explanation would entail a factor of insecurity, because learners are not completely confident how and where exactly the correct  $F_0$  targets have to be achieved. Furthermore, because they concentrate on other aspects, such as the correct pronunciation of segments, or the correct timing of stress, they could have a tendency to disregard the exact pitch range that is used by native speakers of a language. Future research needs to investigate the reasons for pitch range compression in more detail.

Furthermore, the results also replicate findings that there are differences in pitch range depending on the task (e.g. [1],

[21]). In this study, speakers showed a higher PDQ in the story condition compared to the sentence condition. This finding is not surprising: the single sentences that were read by participants, occurred without any context and arguably do not have a communicative function. Therefore, contrastiveness or givenness, for instance, do not play a role in contrast to the story condition. Especially for the story of the three little pigs, a lively, narrative production is more likely than for single sentences.

As for the second question whether the L2 proficiency level of L2 has an influence on the pitch range compression, the results presented here are not conclusive. They do not exclude the possibility of an overall trend for advanced speakers to produce greater pitch variation in L2 compared to L1 and are closer to native speakers (as results by [29] would suggest). Of the three speakers that did not compress pitch range in the L2 production, two are advanced learners. However, such interpretations have to be made with caution, because there were only two advanced learners for each native speaker group. Additionally, there was also one French beginner with a higher PDQ in the non-native productions.

For language teaching, the results suggest that students should be made aware of the lack in pitch variation in order to sound more native-like. Results indicate that a special teaching of pitch variation indeed reduces the foreign accent (e.g. [7], [15], [16]). Pitch range differences that occur across languages otherwise might be aggravated (see also [4] for a cross-linguistic comparison of such differences).

As indicated in the introduction, the results reported here are preliminary. For instance, the two language groups differed with respect to the number of male and female speakers. Furthermore, there were only 2 advanced learners, compared to 5 beginners. More generally, the number of 7 speakers is not very high. However, despite the rather small and somewhat unbalanced number of speakers, we were able to obtain results that are promising for further research. The results reported in this article are in need of replication with larger speaker groups before far-reaching claims can be made. But the preliminary results are a starting point for further research, with a corpus that has more speakers. Such a corpus is planned to include many more speakers (see [28]) balanced for gender and proficiency level. Future research could also concentrate on the effect of different speaking styles on pitch range in a L2, such as reading alone, versus speaking in front of an audience, or engaging in conversations. In such scenarios, it would also be interesting to investigate whether convergence of pitch range can be observed, in that one speaker changes his or her behavior in response to an interlocutor.

#### 5. Acknowledgements

This research was funded by ANR and DFG via the IFCASL project. We would like to thank our colleagues in Saarbrücken and Nancy who helped us with the acquisition of the data, and whose discussions improved this article tremendously as well as the three anonymous reviewers.

## 6. References

- [1] Abu-Al-Makarem, A., & Petrosino, L., "Reading and spontaneous speaking fundamental frequency of young Arabic men for Arabic and English languages: A comparative study", *Perceptual and Motor Skills*, 105:572-580, 2007.
- [2] Anderson-Hsieh, J., Johnson, R., & Koehler, K., "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure", *Language Learning*, 42(2):529-555, 1992.
- [3] Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., & Trouvain, J., "Comparison of pitch range and pitch variation in Slavic and German languages", *Proc. Speech Prosody*, Dublin, Ireland, 2014.
- [4] Aoyama, K., & Guion, S. G., "Prosody in second language acquisition: Acoustic analyses of duration and F0 range", in O.-S. Bohn & M. J. Munro [Eds.], *The role of language experience in second-language speech learning - In honor of James Emil Flege*, 281-297, Amsterdam: John Benjamins, 2007.
- [5] Barry, W., Andreeva, B., & Steiner, I., "The phonetic exponents of phrasal accentuation in French and German", *Proc. Interspeech 2007*, Antwerp, Belgium, 1010-1013, 2007.
- [6] Best, C. T., "A direct realist view of cross-language speech perception". In W. Strange (Ed.), *Cross-language studies of speech perception: A historical review*, 171-206, York: Timonium, 1995.
- [7] Bissiri, M. P., & Pfitzinger, H. R., "Italian speakers learn lexical stress of German morphologically complex words", *Speech Communication*, 51(10):933-947, 2009.
- [8] Boersma, P., & Weenink, D., "PRAAT: Doing phonetics by computer", (Version 5.3.59), retrieved from <http://www.praat.org/>, 2013.
- [9] Busà, M. G., & Stella, A., "Intonational variations in focus marking in the English spoken by North-East Italian speakers", In M. G. Busà & A. Stella [Eds.], *Methodological perspectives on second language prosody - Papers from ML2P 2012*, 31-35, 2012.
- [10] Busà, M. G., & Urbani, M., "A cross linguistic analysis of pitch range in English L1 and L2", *Proc. 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 380-383, 2011.
- [11] Caspers, J., & Kepinska, O., "The influence of word-level prosodic structure of the mother tongue on production of word stress in Dutch as a second language", *Proc. 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, 420-423, 2011.
- [12] Colotte, V., "Corpus Recorder", Nancy: LORIA, 2013.
- [13] Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S., "Persistent stress 'deafness': The case of French learners of Spanish", *Cognition*, 106:682-706, 2008.
- [14] Flege, J. E., Munro, M. J., & Fox, R. A., "Auditory and categorical effects on cross-language vowel perception", *Journal of the Acoustical Society of America*, 95(6):3623-3641, 1994.
- [15] Hincks, R., "Processing the prosody of oral presentations" *Proc. InSTIL/ICALL2004 NLP and Speech Technologies in Advanced Language Learning*, Venice (Italy), 63-66, 2004.
- [16] Hincks, R., & Edlund, J., "Promoting increased pitch variation in oral presentations with transient visual feedback", *Language Learning & Technology*, 13(3):32-50, 2009.
- [17] Keating, P., & Kuo, G., "Comparison of speaking fundamental frequency in English and Mandarin", *The Journal of the Acoustical Society of America*, 132(2):1050-1060, 2012.
- [18] Kingston, J., "Learning foreign vowels", *Language and Speech*, 46(2-3):295-349, 2003.
- [19] Koreman, J., Andreeva, B., & Barry, W., "Accentuation cues in French and German", *Proc. Speech Prosody 2008*, Campinas, Brazil, 613-616, 2008.
- [20] Mennen, I., "Can language learners ever acquire the intonation of a second language?", *Proc. STiLL*, Marholmen (Sweden), 17-20, 1998.
- [21] Mennen, I., Schaeffler, F., & Docherty, G., "Pitching it differently: A comparison of the pitch ranges of German and English speakers", *Proc. 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 1769-1972, 2007.
- [22] Mennen, I., Schaeffler, F., & Docherty, G., "Cross-language differences in fundamental frequency range: a comparison of English and German", *The Journal of the Acoustical Society of America*, 131(3):2249-2260, 2012.
- [23] SAS, "JMP" (Version 10), Cary (NC): SAS Institute, 2012.
- [24] Talkin, D., "A Robust Algorithm for Pitch Tracking (RAPT)", In Kleijn, W. B. and Paliwal, K. K. [Eds.], *Speech Coding and Synthesis*. New York: Elsevier, 1995.
- [25] Traunmüller, H., & Eriksson, A., "The frequency range of the voice fundamental in the speech of male and female adults", (manuscript [http://www2.ling.su.se/staff/hartmut/f0\\_m&f.pdf](http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf), last access: December 11, 2013), 1995.
- [26] Trofimovich, P., & Baker, W., "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech", *Studies in Second Language Acquisition*, 28:1-30, 2006.
- [27] Trouvain, J., & Gut, U. [Eds.], "Non-native prosody: Phonetic description and teaching practice", Berlin/New York: Mouton de Gruyter, 2007.
- [28] Trouvain, J., Laprie, Y., Möbius, B., Andreeva, B., Bonneau, A., Colotte, V., Fauth, C., Fohr, D., Juvet, D., Mella, O., Jügler, J., & Zimmerer, F., "Designing a bilingual speech corpus for French and German language learners", *Proc. Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Mélanges*, Strasbourg, 32-34, 2013.
- [29] Ullakonoja, R., "Comparison of pitch range in Finnish (L1) and Russian (L2)", *Proc. 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 1701-1704, 2007.
- [30] van Bezooijen, R., "Sociocultural aspects of pitch differences between Japanese and Dutch women", *Language and Speech*, 38(3):253-265, 1995.