# Comparison of Pitch Range and Pitch Variation in Slavic and Germanic Languages

*Bistra Andreeva*[1], *Grażyna Demenko*[2], *Magdalena Wolska*[3], *Bernd Möbius*[1], *Frank Zimmerer*[1], *Jeanin Jügler*[1], *Magdalena Oleskowicz-Popiel*[2], *Jürgen Trouvain*[1]

[1] Computational Linguistics & Phonetics, Saarland University, Germany
[2] Department of Linguistics, Adam Mickiewicz University, Poland
[3] LEAD, Eberhard Karls University Tübingen, Germany

[andreeva, moebius, zimmerer, juegler, trouvain]@coli.uni-saarland.de,
lin@amu.edu.pl, magdalena.jastrzebska@speechlabs.pl, magdalena.wolska@uni-tuebingen.de

## Abstract

This study presents the results of a large-scale comparison of various measures of pitch range and pitch variation in two Slavic (Bulgarian and Polish) and two Germanic (German and British English) languages. The productions of twenty-two speakers per language (eleven male and eleven female) in two different tasks (read passages and number sets) are compared. Significant differences between the language groups are found: German and English speakers use lower pitch maxima, narrower pitch span, and generally less variable pitch than Bulgarian and Polish speakers. These findings support the hypothesis that linguistic communities tend to be characterized by particular pitch profiles.

**Index Terms**: pitch range, pitch variation, cross-language differences, Bulgarian, Polish, German, British English

## 1. Introduction

Several studies over the past decades have shown that linguistic communities (different social groups within a single language or speakers of different languages) tend to be characterized by particular pitch profiles (pitch range and pitch variation, see [7] for a review). Luchsinger and Arnold [14] found that Puerto Rican girls in New York City and native American women use fundamental frequency (*f0)* differently. While Puerto Rican girls tend to speak on a rather high pitch, many American women prefer to speak on a low pitch level. Dialects of a language can also differ with respect to the use of *f0* (e.g. [6, 29]). Various cross-linguistic studies also indicate language specific differences with respect to *f0*. Comparing typologically different languages (English, Spanish, Japanese, Tagalog), Hanley et al. [9] and Hanley and Snidecor [10] found that the fundamental frequency of English males had the lowest median *f0*. Later studies compared Polish vs. English [17], Mandarin vs. English [4, 11], British English vs. German [18, 19], or Russian vs. German [20]. Some studies showed that bilingual speakers differ when speaking their two languages. For example, bilingual English/Japanese speakers used a higher pitch in Japanese than in English [8, 28, 30]. These findings demonstrate that such differences need not be due to physiological differences between speakers of different languages.

Ohala and Gilbert's [21] report on experiments in which listeners can identify their own language (Japanese, Cantonese and English) based solely on prosodic cues (*f0*, amplitude and timing characteristics). It has further been found that some languages are discriminable purely by their fundamental frequency ([23] for English and Japanese, [15] and [16] for English and French and [5] for English and Dutch).

Language specific components have also been found to be important in the perception and production of paralinguistic aspects ([13] for politeness in Japanese and English, [3] for 'confident', 'friendly', 'emphatic' and 'surprised' in British English and Dutch).

However, it is difficult to compare the data reported in these publications, because most studies have been limited to either male or female (mostly small numbers of) speakers, the analyses were based on different discourse types, or the methods for *f0* estimation were different.

The aim of this study is to lay a foundation for a large-scale quantitative analysis of the fundamental frequency (level and span) of speakers of two typologically different language groups (Slavic: Bulgarian and Polish, and Germanic: German and (British) English). The analysis presented here is based on the assumption that pitch range and pitch variation in linguistically homogeneous communities will cluster within each community, but might differ across communities.

## 2. Material and Methods

Two Slavic (Bulgarian and Polish) and two Germanic (German and English) languages are in the focus of this study. The material analyzed is continuous read speech taken from two comparable multi-lingual speech databases, EUROM-1 (for German and English) [2] and BABEL (for Bulgarian and Polish) [25, 26]. We used a subset of the data, consisting of 3 blocks of 20 numbers (from 0 to 9999) and 3 cognitively linked short passages, containing 5 thematically connected sentences, read by 22 speakers (11 male and 11 female) per language. The passages were based on identical, real-life topics for the different languages, freely translated and adapted for Bulgarian, German and Polish from the original English texts. The overall length of the analyzed material per language is about 60 minutes.

## 3. $f_0$ Measures

Pitch values were collected at 0.01 seconds time steps for the male and 0.005 seconds time steps for the female speakers using the RAPT algorithm [27] implemented in the program 'get_f0' from the ESPS software package. The automatically extracted *f0* values were verified and manually corrected, if necessary. Irregular voiced stretches of speech due to laryngealization were excluded from further analyses.

According to Ladd [12], *f0* values can be attributed to two partially related but distinct characteristics of a speaker's performance: (a) pitch level, i.e. the overall height of the speaker's voice, and (b) pitch span, i.e. the range of frequencies covered by the speaker. To analyze the cross-language differences in pitch range and variation, the following distributional measures were calculated: mean and median *f0* values for level and interquartile range (IQR) and the simple pitch excursion for span, whereas the latter was simply computed as the difference between maximum and minimum pitch values over a passage or number block. The obtained Hertz measurements for span were additionally converted to semitones by means of the formula [24]:

39.863 * log10(Maximum/Minimum).

The measures describing the variation and shape of the *f0* distribution were standard deviation (SD), kurtosis and skewness (in Hz).

## 4. Results

Means and standard deviations for each of the distributional measures for pitch level and span are presented in Table 1, organized by language, speaker sex, and task.

As a first step towards determining the differences, linear mixed models with the respective measure as dependent variable, speaker as random factor, and native language (Bulgarian/Polish/English/German), gender (male/female) and task (passage/number set) as fixed factors, as well as all their possible interactions, were computed for each dependent variable in separate analyses. Separate Tukey post-hoc tests were carried out per variable, if appropriate. The confidence level was set at $\alpha$=0.05.

### 4.1. Passages

Predictably, gender had a significant main effect on mean (F [1, 80] = 520.32, p<0.001) and median *f0* (F [1, 80] = 480.50, p<0.001), IQR (F [1, 80] = 70.47, p<0.001), minimum *f0* (F [1, 80] = 266.57, p<0.001), maximum *f0* (F [1, 80] = 341.84, p<0.001) and SD (F [1, 80] = 94.69, p<0.001), with females having significantly higher *f0* values (cf. Figure 1 for level measures). Gender did not differ in skewness, kurtosis and *f0* span measured in semitones.

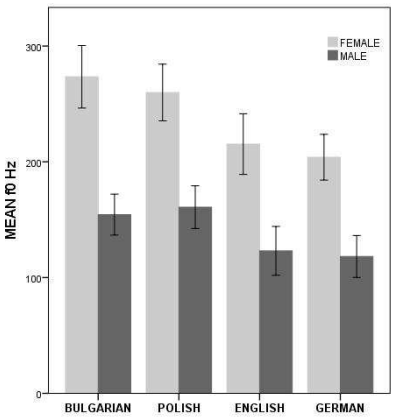


Figure 1: *Mean f0 pooled over all male and female speakers and all tasks.*

Table 1. Means and standard deviations for the distributional measures by language, speaker sex and task. The values for each measure are given in Hz except for the second span measure which is in semitones.

| | measure | Passages | | Numbers | |
|---|---|---|---|---|---|
| | | *M* | *F* | *M* | *F* |
| BG | mean | 154 (17) | 275 (26) | 155 (19) | 272 (28) |
| | median | 154 (17) | 273 (26) | 159 (20) | 278 (29) |
| | IQR | 43 (13) | 72 (17) | 28 (9) | 55 (16) |
| | minimum | 81 (11) | 152 (24) | 83 (17) | 165 (24) |
| | maximum | 231 (34) | 428 (42) | 195 (23) | 352 (40) |
| | span | 151 (33) | 276 (42) | 112 (21) | 187 (30) |
| | span (ST) | 18.3 (3.1) | 18.0 (2.7) | 14.9 (3.0) | 13.1 (1.8) |
| | SD | 29 (8) | 52 (9) | 22 (5) | 40 (8) |
| | skewness | -.02 (.39) | .18 (.24) | -.95 (.54) | -.50 (.38) |
| | kurtosis | -.27 (.71) | -.25 (.39) | .98 (1.5) | -.19 (.78) |
| PL | mean | 157 (18) | 259 (21) | 165 (19) | 260 (28) |
| | median | 156 (19) | 254 (21) | 166 (20) | 259 (31) |
| | IQR | 38 (11) | 62 (15) | 51 (18) | 73 (18) |
| | minimum | 82 (8) | 146 (23) | 92 (21) | 165 (16) |
| | maximum | 246 (34) | 437 (56) | 231 (27) | 382 (50) |
| | span | 165 (33) | 291 (59) | 139 (30) | 217 (43) |
| | span (ST) | 19.0 (2.7) | 19.1 (3.6) | 16.2 (4.0) | 14.5 (2.2) |
| | SD | 30 (8) | 50 (9) | 32 (10) | 46 (9) |
| | skewness | .10 (.51) | .49 (.43) | -.18 (.41) | .10 (.35) |
| | kurtosis | .28 (.57) | .52 (1.0) | -.77 (.40) | -.76 (.46) |
| DE | mean | 120 (18) | 206 (21) | 116 (18) | 202 (19) |
| | median | 119 (19) | 204 (22) | 117 (18) | 204 (18) |
| | IQR | 24 (8) | 44 (13) | 15 (5) | 28 (6) |
| | minimum | 82 (14) | 137 (30) | 85 (15) | 144 (24) |
| | maximum | 181 (33) | 298 (30) | 149 (22) | 264 (25) |
| | span | 100 (28) | 161 (39) | 64 (11) | 120 (24) |
| | span (ST) | 13.8 (2.9) | 13.9 (4.5) | 9.7 (1.4) | 10.6 (2.6) |
| | SD | 18 (6) | 30 (8) | 11 (3) | 20 (4) |
| | skewness | .44 (.49) | .26 (.29) | -.23 (.51) | -.04 (.48) |
| | kurtosis | .35 (1.1) | -.29 (.41) | .17 (.75) | .15 (1.4) |
| EN | mean | 127 (23) | 218 (23) | 119 (19) | 213 (29) |
| | median | 125 (22) | 214 (25) | 119 (20) | 212 (31) |
| | IQR | 30 (13) | 41 (12) | 27 (15) | 32 (14) |
| | minimum | 84 (14) | 155 (22) | 79 (10) | 160 (31) |
| | maximum | 205 (52) | 330 (44) | 180 (45) | 294 (42) |
| | span | 121 (44) | 175 (47) | 101 (43) | 134 (36) |
| | span (ST) | 15.1 (3.2) | 13.1 (3.4) | 14.0 (3.9) | 10.7 (3.1) |
| | SD | 23 (9) | 32 (8) | 19 (10) | 23 (8) |
| | skewness | .67 (.36) | .72 (.47) | .41 (.46) | .47 (.48) |
| | kurtosis | .50 (.91) | .73 (1.1) | .26 (1.7) | .23 (.90) |

However, over and above the expected gender effect, there was also a significant main effect of language on all measurements except on minimum *f0*, where the speakers are near the floor of their physiological *f0* range. Separate post-hoc tests showed that Bulgarian and Polish speakers had a significantly higher mean *f0* (F [3, 80] = 33.07, p<0.001), median *f0* (F [3, 80] = 32.60, p<0.001), IQR (F [3, 80] =

21.06, p<0.001), maximum *f0* (F [3, 80] = 33.29, p<0.001), *f0* span in semitones (F [3, 80] = 17.05, p<0.001) and SD (F [3, 80] = 26.96, p<0.001) than English and German speakers. We found a positively skewed *f0* distribution for the four languages. This implies that the most frequent *f0* observation occurs lower than the mean. The skewness values for English speakers were significantly higher than those for German, Polish and Bulgarian speakers (F [3, 80] = 11.51, p<0.001). English speakers also had a higher kurtosis than German and Bulgarian speakers, and Polish speakers had a higher kurtosis than Bulgarian speakers (F [3, 80] = 8.33, p<0.001). This reflects the fact that *f0* in Bulgarian and German is distributed over a narrower area (cf. Tables 1 and 2).

The statistical analysis revealed a significant interaction between language and gender for mean *f0* (F [3, 80] = 3.10, p<0.05), maximum *f0* (F [3, 80] = 6.15, p<0.001) and SD (F [3, 80] = 3.76, p<0.05). In the passages, the speakers of the Slavic group used higher average *f0* and higher maximum values and showed a larger SD (possibly indicating more liveliness) than the speakers in the Germanic group. The only exceptions to this finding are the English male speakers with respect to their maximum *f0* values and SD (cf. Figure 2a). Thus, the English male speakers used the same maximum *f0* values and had the same SD as the German male speakers and the male speakers from the Slavic group. Figures 2a and 2b display the mean and maximum pitch values as well as SD for male and female speakers in the four languages.

### 4.2. Number Blocks

In these analyses, as expected, we found again that women had a significantly higher mean *f0* (F [1, 80] = 424.13, p<0.001), median *f0* (F [1, 80] = 379.82, p<0.001), IQR (F [1, 80] = 43.09, p<0.001), maximum *f0* (F [1, 80] = 339.89, p<0.001), minimum *f0* (F [1, 80] = 315.98, p<0.001) and SD (F [1, 80] = 51.78, p<0.001) . In contrast to the findings for the passages, in the number task male speakers used a larger frequency range in semitones (F [1, 80] = 6.97, p<0.01) than females. This result may be partially attributed to the fact that speakers tend to use quite idiosyncratic intonation patterns for the number blocks: some speakers prefer continuation rises to separate the blocks, other speakers tend to use falling intonation to end a block. Gender was also significant for skewness (F [1, 80] = 8.32, p<0.05).

Table 2. *Language-group differences for the f0 measures on the basis of Tukey post-hoc comparisons.*

| *f0* measure | significant language-group differences | |
|---|---|---|
| | passages | number blocks |
| mean *f0* | BG = PL > EN = DE | BG = PL > EN = DE |
| median *f0* | BG = PL > EN = DE | BG = PL > EN = DE |
| min *f0* | N.S. | N.S. |
| IQR | BG = PL > EN = DE | PL > BG > EN = DE |
| max *f0* | PL = BG > EN = DE | PL > BG > EN > DE |
| span ST | PL = BG > EN = DE | PL = BG > BG = EN > DE |
| SD | BG = PL > EN = DE | PL > BG > EN = DE |
| skewness | EN > DE = PL = BG | EN > PL = DE > BG |
| kurtosis | EN = PL > PL = DE > DE = BG | BG = EN = DE > PL |

Again, a significant main effect of language was found in all measures except *f0* minimum: mean *f0* (F [3, 80] = 37.75, p<0.001), median *f0* (F [3, 80] = 36.73, p<0.001), IQR (F [3, 80] = 44.68, p<0.001), maximum *f0* (F [3, 80] = 35.52, p<0.001), *f0* span in semitones (F [3, 80] = 16.27, p<0.001), SD (F [3, 80] = 42.57, p<0.001), skewness (F [3, 80] = 31.87, p<0.001) and kurtosis (F [3, 80] = 9.82, p<0.001). However, the post-hoc tests yielded different language groupings (cf. Table 2).

Significant interactions between gender and language were found for IQR (F [3, 80] = 3.30, p<0.05), SD (F [3, 80] = 3.71, p<0.05) and kurtosis (F [3, 80] = 2.99, p<0.05).

### 4.3. Level vs. Span

The scatter plots in Figures 3 and 4 provide a visual representation of *f0* span (in Hz) and level (mean *f0* Hz) for all speakers of the four languages. The figures show that some speakers have a wide span but differ in level or vice versa, i.e. some speakers have a similar level but differ in span. The English and German speakers cluster in the lower left corner of the level/span plane, while the Bulgarian and Polish speakers cluster mostly in the higher right sector, which indicate that Slavic speakers may are more expressive than Germanic speakers in terms of span and level. The scatter plots also illustrate the different strategies of some speakers with respect to task type.

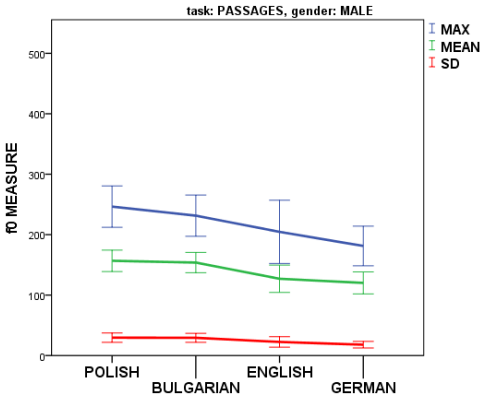


Figure 2a: *Mean and maximum f0 values and SD for male Bulgarian, Polish, English and German speakers.*

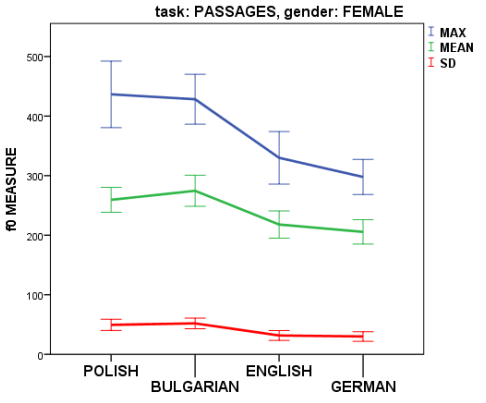


Figure 2b: *Mean and maximum f0 values and SD* for *female Bulgarian, Polish, English and German speakers.*

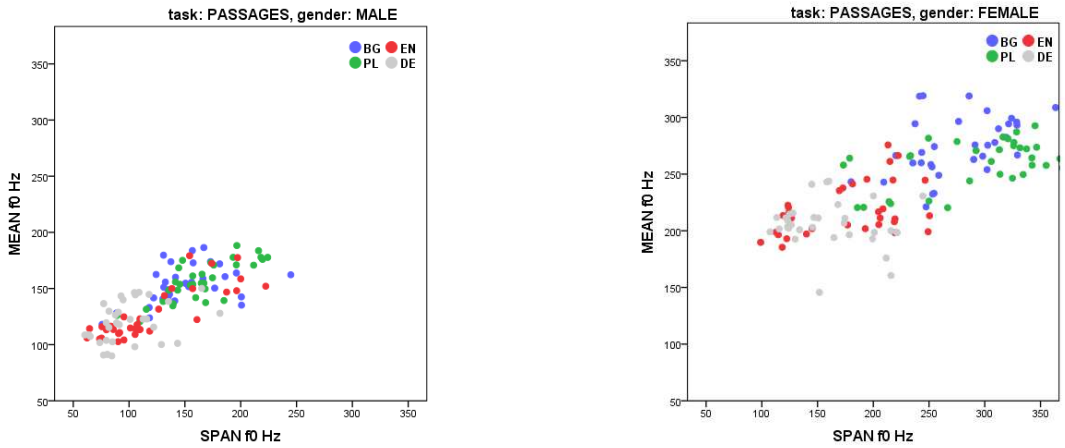


Figure 3: *Scatter plot showing span and level from the passages for male (left panel) and female (right panel) speakers.*

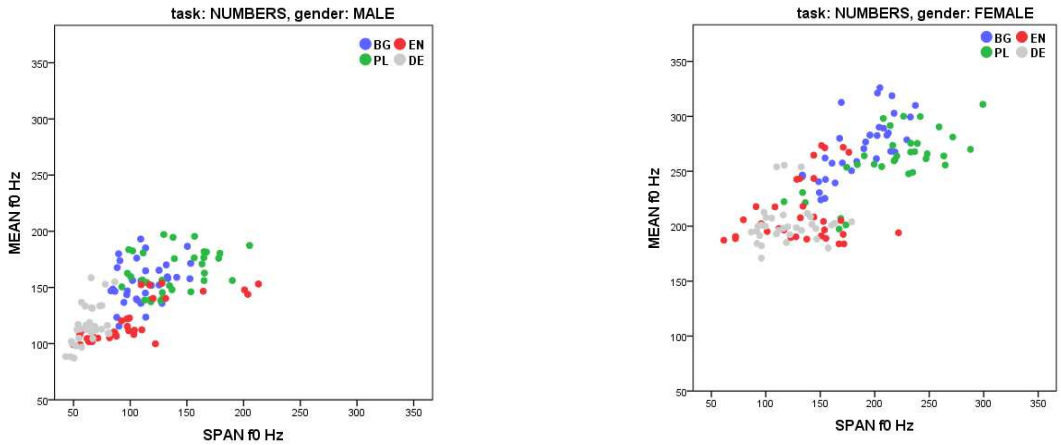


Figure 4: *Scatter plot showing span and level from the number blocks for male (left panel) and female (right panel) speakers.*

## 5. Discussion and Conclusions

This paper contributes to the growing number of studies on cross-language differences in pitch range and pitch variation. Our results confirm the hypothesis that linguistic communities tend to be characterized by particular pitch profiles. German and English speakers use a considerably lower level, narrower span, and generally less variable pitch than Bulgarian and Polish speakers. Gender also plays a significant role in *f0* variation. In the present study a distinctive frequency region as well as different mean frequencies were found for male and female speakers (about 139 Hz and 238 Hz respectively). However, the differences in mean frequencies were on the average less than one octave: 10.14/9.72 semitones for Bulgarian, 8.63/7.86 semitones for Polish, 9.32/9.64 semitones for German and 9.51/10.11 semitones for English (first number for passages, second one for number blocks).

Systematic differences between tasks were observed which appear to be attributable to differing strategies that speakers employ when reading short stories versus lists of numbers. Inter-speaker variability was considerably greater for the number lists. The syntactic-semantic structure of the story seems to constrain the speakers' prosodic options.

Our results do not corroborate the results reported by Mennen and colleagues [18, 19]; the female English and German speakers do not differ with respect to level and span. But distributional measures may in fact not be able to capture significant cross-linguistic differences (cf. [1, 19, 22]). In future work we expect to refine our measures of pitch range, by including linguistically based measures which were found to be better predictors of differences in pitch range and pitch variation across speakers and languages, and also by adding data from more speakers, including Bulgarian and Polish L2 speakers of English, more languages, as well as spontaneous speech data.

## 6. Acknowledgements

This research was partially supported by Research Grant UMO-2012/04/M/HS2/00551 from the NCN (Polish National Research Center).

We would like to thank Ryszard Gubrynowicz (Speech Acoustics Laboratory, Institute of Fundamental Technology Research, Polish Academy of Science) and Snezhina Dimitrova (English Department, Sofia University “St. Kliment Ohridski”) for kindly providing the Babel databases for Polish and Bulgarian, respectively.

## 7. References


[1] Campione, E. , and Véronis, J. (1998). A statistical study of pitch target points in five languages. *Proceedings of ICSLP'98*, 1391-1394.

[2] Chan, D. Fourcin, A.; Gibbon, D.; Granstrom, B.; Huckvale, M.; Kokkinakis, G.; Kvale, K.; Lamel, L.; Lindberg, B.; Moreno, A.; Mouropoulos, J.; Senia, F.; Trancoso, I.; Veld, C. and Zeiliger, J. (1995). Eurom - a spoken language resource for the EU. In *Eurospeech' 95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*, 1, Madrid., 18-21 September 1995, 867–870.

[3] Chen, A, Gussenhoven, C., and Rietveld, T. (2004). Language-specificity in the perception of paralinguistic intonational meaning, *Language & Speech* 47, 311–349.

[4] Chen, G. T. (1972). *A comparative study of pitch range of native speakers of Midwestern English and Mandarin Chinese: An acoustic study*, doctoral dissertation, University of Wisconsin-Madison, Madison.

[5] de Pijper, J. R. (1983). *Modelling British English intonation*, Dordrecht - Holland: Foris.

[6] Deutsch, D., Le, J., Shen, J., and Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *Journal of the Acoustical Society of America*, April, 125, EL208.

[7] Dolson, M. (1994). The pitch of speech as a function of linguistic community, *Music Perception* 11 (3), 321–331.

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

[9] Hanley, T.D., Snidecor, J.C., and Ringel, R. (1966). Some acoustic differences among languages, *Phonetica* 14, 97–107.

[10] Hanley, .D. and Snidecor, J.C. (1967). Some acoustic similarities among languages, *Phonetica* 17, 141–148.

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

[12] Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

[13] Loveday, L. (1981). Pitch, politeness and sexual role: an explanatory investigation into the pitch correlates of English and Japanese formula,. *Language and Speech* 24, 71–89.

[14] Luchsinger, R. and Arnold, G. (1965). *Voice-Speech-Language*. Constable&Co Ltd., London.

[15] Maidment, J. A. (1976). Voice fundamental frequency characteristics as language differentiators. *Speech and hearing: Work in progress*, University College London, 74–93.

[16] Maidment, J. A. (1983). Language recognition and prosody: further evidence, *Speech, hearing and language: Work in progress*, University College London 1, 133–141.

[17] Majewski, W., Hollien, H., and Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males, *Phonetica* 25, 119–125.

[18] Mennen, I., Schaeffler, F., & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. *16th International Congress of Phonetic Sciences* (ICPhS XVI), Saarbrücken,1769-1972.

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

[20] Nebert, Augustin Ulrich (2013). *Der Tonhöhenumfang der deutschen und russischen Sprechstimme. Vergleichende Untersuchung zur Sprechstimmlage.* Hallesche Schriften zur Sprechwissenschaft und Phonetik, Band 46. Frankfurt/M.

[21] Ohala, J. J., and Gilbert, J. B. (1979). Listeners' ability to identify languages by their prosody, in P. Léon and M. Rossi (eds.) *Problèmes de Prosodie*, Didier, Ottawa, 123–131.

[22] Patterson, D. (2000). *A Linguistic Approach to Pitch Range Modelling*. Ph.D. dissertation, University of Edinburgh.

[23] Ramus, F., and Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America* 105 (1), 512–521.

[24] Henning Reetz (1999): *Artikulatorische und akustische Phonetik*. Wissenschaftlicher Verlag, Trier.

[25] Roach, P.; Arnfield, S. and Hallum, E., (1996). BABEL: A multi-language speech database. In *Proceedings of SST-96: Speech and Science Technology Conference*, Adelaide, 351–4.

[26] Roach, P., Arnfield, S., Barry, W.J., Dimitrova, S., Boldea, M., Fourcin, A., Gonet, W., Gubrynowicz, R., Hallum, E., Lamel, L., Marasek, K., Marchal, A., Meister, E., Vicsi, K. (1998). Babel: a database of Central and Eastern European languages, *Proceedings of the First International Conference on Language Resources and Evaluation*, vol. 1, 28-30 May 1998, Granada, Spain, pp. 371–374.

[27] Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). In Kleijn, W. B. and Paliwal, K. K. (eds.), *Speech Coding and Synthesis*. New York: Elsevier.

[28] Todaka, Y. (1993). *A cross-language study of voice quality: bilingual Japanese and American speakers*, doctoral dissertation, University of California, Los Angeles, pp. 145–147.

[29] Torgerson, R. C. (2005*). A comparison of Beijing and Taiwan Mandarin tone register: An acoustic analysis of three native speech styles*, master's thesis, Brigham Young University, 73–82.

[30] Yamazawa, H., and Hollien, H. (1992 ). Speaking fundamental frequency pattern of Japanese women, *Phonetica* 49, 128–140.