Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Speaker Age Effects on Prosodic Patterns in Bulgarian

Snezhina Dimitrova[1], *Bistra Andreeva*[2], *Christoph Gabriel*[3], *Jonas Grünke*[3]

[1]Sofia University "St. Kliment Ohridski", Bulgaria
[2]Universität des Saarlandes, Germany
[3]Johannes Gutenberg University Mainz, Germany

snezhina@uni-sofia.bg, andreeva@coli.uni-saarland.de, christoph.gabriel@uni-mainz.de, jgruenke@uni-mainz.de

Abstract

We investigated prosodic variability attributable to age in Standard Bulgarian. In readings of *The North Wind and the Sun*, recorded by two groups of six female speakers aged between 19-23 and 79-88 years, we found significant differences in pitch span, minimum F0, syllable, intonation phrase and pause duration. The older speakers made more pauses, which were also of longer duration. They also realized longer syllables and intonation phrases than young speakers. Both groups used the same inventory of pitch accents and boundary tones, but there were significant differences in the frequency counts of some of the tones: young speakers used pre-nuclear rises with a post-tonic high target, while older speakers preferred rises with a high target within the stressed syllable; the nuclear pitch accent used most frequently by the young speakers was L*, whereas the one preferred by the elderly speakers was L+H*; younger speakers used more phrase accents (especially H-), while older speakers preferred boundary tones (H-% and L-%) and "level" (H-L% and HL-) pitch curves. Our findings suggest that the study of tonal repertoires and frequencies of use could offer interesting insights into age-related differences between speakers.

Index Terms: age-related variation, intonation, phrasing, pause duration, Bulgarian

1. Introduction

Age-related variation in speech has been studied extensively, primarily for age estimation or for clinical application purposes. Researchers have mainly used one of two methodological approaches – acoustic analysis or perception tests. A useful introduction along with a comprehensive state-of-the-art review of studies of the acoustic phonetic manifestations of ageing in the sound signal, intended primarily for the purpose of building reliable automatic classifiers of speakers according to their age, is provided by [1]. Other studies aiming at facilitating automatic age estimation of speakers' voices include [2], [3], [4] and [5], to name but a few.

On the other hand, studies such as [6] and [7] approach the question about the manifestations of ageing in the speech signal from the point of view of clinical research. An interesting further perspective on the relationship between vocal characteristics and perceived age is offered by [8], who investigated the possibility to affect age perception through vocal manipulation.

Most of the above research has been interested in identifying features of ageing both on the level of the sound segment and on the supra-segmental level. Some of the well-known and widely-investigated prosodic manifestations of adult speaker age in the speech signal include speech rate (segment and syllable duration, number of segments per unit of time, number and duration of pauses), sound pressure level, F0 (mean, level, range and standard deviation), spectral tilt, etc. ([1], [3], [5], [9], [10]).

However, the correlation between perceptual and acoustic cues to ageing is not always a straightforward one. Besides, some perceptual cues used by listeners to determine a speaker's age seem not to correspond to any measurable attributes of the acoustic signal. It is also important to note that some of the results hitherto reported in the literature are divergent or even occasionally contradictory.

A conspicuous feature which has not been previously analyzed, probably due to the primary orientation of existing research towards the technological or clinical domain, is the pattern of F0 change. Although some of the above-mentioned studies incorporated pitch curve information into the prosodic components of their models (e.g., [3] and [5]), pitch contour tracking was done automatically, and without any consideration of the linguistic importance of the respective change in pitch.

We investigate prosodic variability in the speech of young and mature speakers of Standard Bulgarian (SB) – a South Slavic language for which no longitudinal or cross-sectional research on age-related prosodic variability exists to date. Unlike most of the studies cited above, we approach the question of age-related variability in the speech signal primarily from a socio-phonetic point of view, and focus our attention in the present investigation on differences attributable to age which are found in the prosodic (temporal and intonational) domain.

2. Empirical study

2.1. Speakers and data

Our data consist of readings of the Bulgarian version of Aesop's fable *The North Wind and the Sun* by two groups of female speakers. The text was recorded by the speakers together with materials for other experiments not reported here. Two of the mature speaker recordings were made in September 2012, whereas all remaining recordings were made between September 2016 and May 2017.

The first group consists of six mature speakers of Standard Bulgarian who were between 79 and 88 years old at the time of recording (henceforth the "79-88 GROUP"). Two of them have lived in Sofia all their lives, while the other four moved to the Bulgarian capital city either in very early childhood, or

10.21437/SpeechProsody.2018-144

as young adolescents for study purposes. All mature female speakers hold an academic degree.

The second group consists of six young females who were aged between 19 and 23 at the time of recording (henceforth the "19-23 GROUP"). They were all born, grew up and live in Sofia. All of them were undergraduate university students.

The pronunciation of both groups of speakers displays the features typical of the capital Sofia.

2.2. Methodology

The temporal characteristics which were investigated in the present study were mean syllable duration, speech tempo (number of syllables per second), intonation phrase (IP) and pause duration.

Syllable boundaries were marked and prominent syllables were labelled manually in Praat [11] (for an example, see Figure 1). All temporal features were measured using Praat scripts.

We also analyzed pitch level (defined as the overall height of a speaker's voice) and pitch span (defined as the range of frequencies typically covered by a speaker). According to [12], the two are partially related but nevertheless distinct characteristics of a speaker's performance to which F0 values can be attributed.

The long-term distributional (LTD) measures which were calculated for the purposes of the present analysis were as follows: for pitch level - mean and median F0 values (in Hz), for span – pitch excursion (in semitones - ST), computed as the difference between the maximum and minimum pitch values obtained over a given IP. The measure used for describing F0 distribution variation is the standard deviation (SD, in Hz).

The obtained Hertz measurements for span were additionally converted to semitones by means of the formula in [13]:

39.863 * log10 (Maximum/Minimum). (1)

A Praat script was then used to calculate the LTD measures.

Finally, we used the ToBI labelling conventions outlined in [14], and also employed in recent Autosegmental-Metrical analyses of the intonation of the Sofia variety of Contemporay Standard Bulgarian ([15], [16], [17]) to mark pre-nuclear and nuclear pitch accents, phrase accents and boundary tones in the data (for an example, see Tier 1 "accents" in Figure 1).

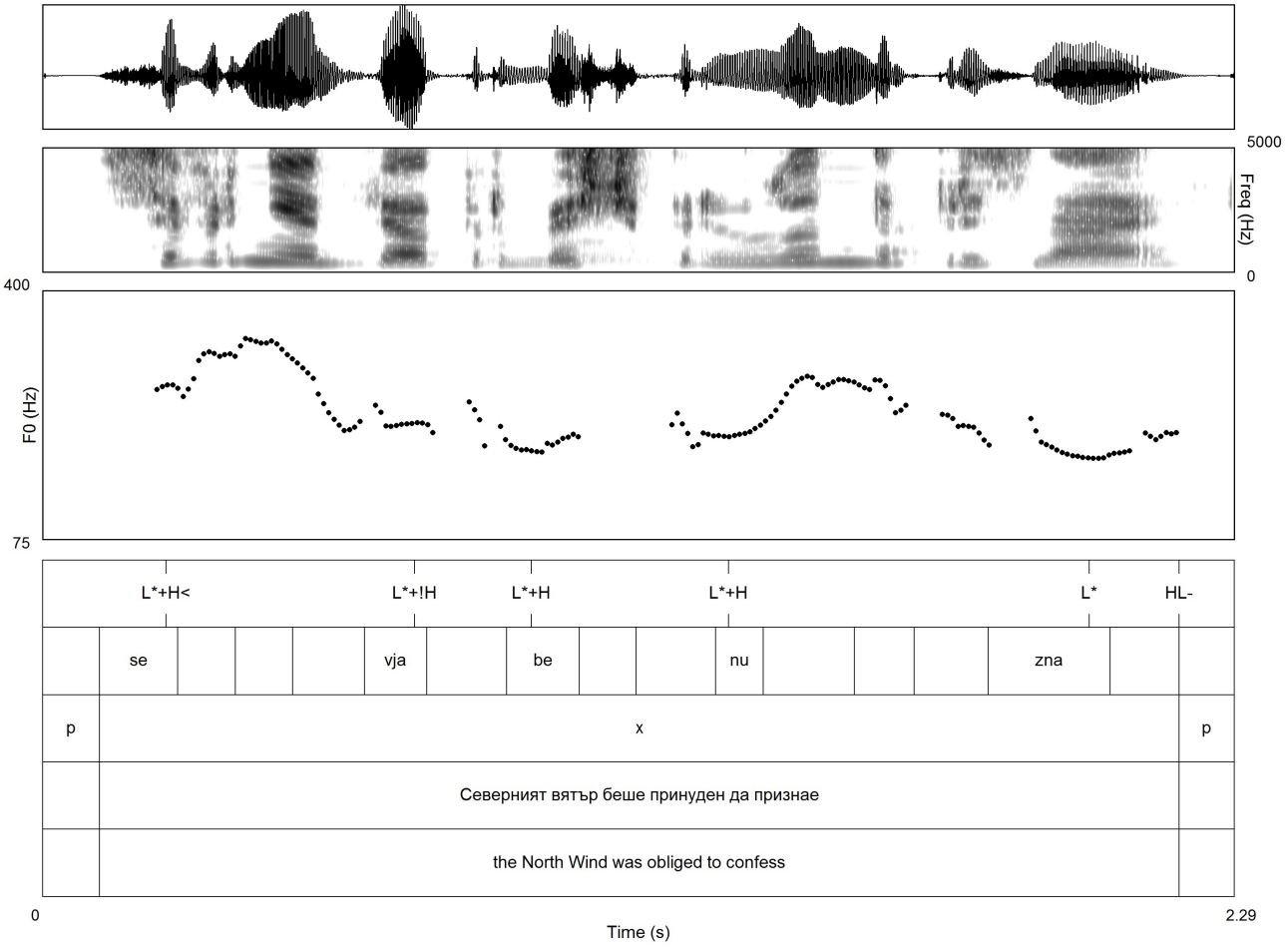

Figure 1: The utterance "Северният вятър беше принуден да признае" ('The North Wind was obliged to confess'), pronounced as a single IP by a young female Bulgarian speaker. Labelling of the data: tier 1 – ToBI labelling of pitch accents, phrasal accents and boundary tones; tier 2 – syllable boundaries and prominent syllable labels; tier 3 – speech intervals (x) and pauses (p); tier 4 – Bulgarian text; tier 5 – English translation.

2.3. Results and discussion

Linear mixed models (LMMs) with the respective measure as dependent variable, Speaker as random factor, and Group (old "79-88 GROUP", young "19-23 GROUP") as fixed factors were calculated, and post-hoc tests were carried out.

2.3.1 Fundamental frequency

Statistically significant differences were found for pitch range and minimum F0.

The minimum F0 value for the "79-88 group" was 135.8 Hz (St. error = 8.9), and 176.9 Hz (St. error = 8.9) for the "19-23 group", ($F [1, 10.06] = 10.7550$, $p<0.01$).

Pitch range for the "79-88" was 11.6 ST (St. error = 0.66), whereas for the "19-23 group" it was 9.1 ST (St. error = 0.66), ($F [1, 9.119] = 6.9627$, $p < 0.01$).

The wider pitch range used by the mature group of female speakers does not corroborate many of the previously reported findings in the literature which show the F0 range of older speakers to be narrower than that of younger ones. However, our present findings are in line with [18] who found a statistically significant main effect for age on minimum F0, span in Hertz and semitones, and SD: their "older" group Bulgarian speakers showed a significantly lower minimum F0, higher F0 span in Hertz and semitones, and higher SD than the "younger" Bulgarian speakers who participated in their study. Besides, results obtained e.g. for Hungarian by [19] found an insignificant effect of age on F0 range, emphasizing the fact that the pitch domain used by a speaker is very much an individual characteristic.

The lower F0 used by the mature speakers, on the other hand, is generally in conformity with findings for other languages (see overview in [19]).

The two groups of speakers use roughly the same number of pitch accents and boundary tones (Table 1).

Table 1: *Number of pre-nuclear pitch accents (PAs), nuclear pitch accents and phrase and boundary tones used by the two groups of speakers*

	pre-nuclear	**nuclear**	**Boundaries**
19-23 GROUP	250	142	146
79-88 GROUP	257	159	163

Our analysis of the types of accents shows that the speakers use very similar tonal repertoires. For frequency counts of the different pitch accents realized by the two groups, we used Chi square tests, the results of which showed statistically significant differences between the groups.

Pre-nuclear pitch accents

Frequency counts of the different pre-nuclear pitch accents used by the two groups of speakers were statistically significant:

χ^2 (6, N = 507) = 62,537, p <.001.

Both groups of speakers use pre-nuclear H* pitch accents (45.9% and 43.2% of all pre-nuclear accents realized by the "79-88" and "19-23" groups, respectively). But while for the "79-88 GROUP", L+H* is the second most frequently used tone which is found 31.1% of the time, the "19-23 GROUP" prefers to use L+<H* (21.2%) and L*+H (15.2%). For example, in the young speaker's utterance shown in Figure 1, all pre-nuclear pitch accents are of the L*+H type (with the H target of the first one aligned two syllables after the accent). On the other hand, in the mature speaker's pronunciation of the same utterance which is realized as two separate IPs, the pre-nuclear pitch accent in the first IP is H*, and the two pre-nuclear pitch accents in the second IP are both of the L+H* type.

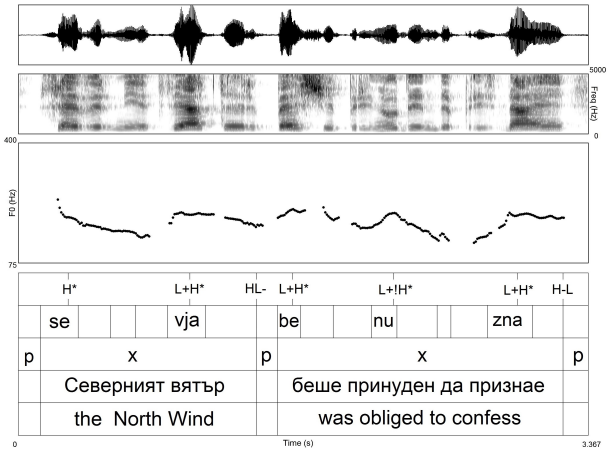

Figure 2. The utterance "Северният вятър беше принуден да признае" ('the North Wind was obliged to confess'), pronounced as two IPs by a mature female Bulgarian speaker.

If phonetic alignment details are disregarded, it appears that younger speakers use rising pre-nuclear tones with a low pitch target associated with the stressed syllable and a high pitch target reached in the post-tonic syllable (L*+<H and L*+H) much more often (20% of the time) than older speakers (only 4.3% of the time). On the other hand, the "79-88 GROUP" use pre-nuclear rises in which the H target is associated with the stressed syllable (L+<H* and L+H*) more often (42.8% of the time) than the young speakers (30.8%).

The distribution of pre-nuclear pitch accents is shown in Table 2.

Table 2: *Pre-nuclear pitch accent repertoires used by the two groups of speakers (in %).*

	H*	**H+L***	**L***	**L*+<H**	**L*+H**	**L+<H***	**L+H***
19-23	43.2	1.2	4.8	4.8	15.2	21.2	9.6
79-88	45.9	1.6	5.4	1.6	2.7	11.7	31.1

Nuclear pitch accents

Chi square tests for frequency counts of the different nuclear pitch accents realized by the two groups again showed statistically significant differences between the groups:

χ^2 (5, N = 301) = 25,533, p <.001.

Both groups use H* nuclear pitch accents (31.5% - by the "79-88 GROUP" vs. 26.8% by the "19-23 GROUP"). However, the most frequent tone which was used half of the time by the "19-23 GROUP" is actually L* (50%, vs. 25.2% for the "79-88 GROUP"). On the other hand, the older speakers use L+H* 33.3% of the time, while in the readings of the younger speakers it is found only 14.1% of the time. For example, the nuclear pitch accent used by the young speaker

in Figure 1 is L*, while the two nuclear pitch accents used by the mature speaker in Figure 2 are both L+H*.

The distribution of nuclear pitch accents is shown in Table 3, below.

Table 3. *Nuclear pitch accent repertoires used by the two groups of speakers (in %).*

	H*	**H+L***	**L***	**L*+H**	**L+<H***	**L+H***
19-23	26.8	6.3	50.0	1.4	1.4	14.1
79-88	31.5	6.9	25.2	0.6	2.5	33.3

Boundary tones

Chi square tests for frequency counts of the boundary tones realized by the two groups of female speakers again showed statistically significant differences between the groups: for boundary tones, $\chi 2$ (8, N = 309) = 50,291, p <.001.

Both groups of speakers use a low phrase accent followed by a low boundary tone L-% about 29% of the time. It is the most frequent boundary marker in the readings of the "79-88 GROUP", followed by H-% (21.5%). For the younger group, however, the most frequently occurring tone is H- (36.3%, usually preceded by a L* nuclear pitch accent).

Generally, in our data younger speakers tend to use more phrase accents (especially H-), while mature speakers seem to prefer boundary tones (H-%, L-%) and "level" (H-L%, HL-) pitch curves.

The distribution of boundary tones is shown in Table 4.

Table 4. *Boundary tones used by the two groups of speakers (in %).*

	H-	**H-%**	**HL-**	**H-L%**	**L-**	**L-%**	**LH-**	**L-H%**
19-23	36.3	9.6	6.9	6.9	2.7	28.8	2.0	4.1
79-88	7.4	21.5	11.0	17.2	0.6	29.5	1.8	8.6

We are unaware of similar research comparing the tonal inventories and frequency counts of the types of tones used by young and elderly speakers. The statistically significant results that we obtained seem to suggest that this line of research could offer interesting insights into the pitch pattern preferences of different age groups of speakers.

2.3.2 Temporal characteristics

The "79-88 GROUP" made almost twice as many pauses as the "19-23 GROUP" of speakers (116 vs. 68 pauses, respectively). The average duration of the pauses and the standard deviation for the two groups, however, was almost identical (M = 434.46 ms, SD = 243.46 for the elder group, and M = 434.75 ms, SD = 246.24 for the younger group). However, mean intonation phrase (IP) duration for the two groups was shown to be statistically significant: 150 IPs, M = 1471.86 ms , SD = 777.98 for the "79-88 GROUP" vs. 140 IPs, M = 1180.51 ms, SD = 517.46 for the "19-23 GROUP"; (F [1, 10.53] = 6.3802, p < 0.05) (Table 5).

Table 5. *Number of pauses and intonation phrases (IPs), means and SDs (in ms) used by the two groups of speakers.*

	PAUSES			**IPs**		
	N	**Mean**	**SD**	**N**	**Mean**	**SD**
19-23	68	434.75	246.24	140	1180.51	517.46
79-88	116	434.46	243.46	150	1471.86	777.98

Finally, mean syllable duration differences were also statistically significant: mature speakers from the "79-88 GROUP" used longer syllables (M = 183.07 ms, SD = 20.09) than younger speakers from the "19-13 GROUP" (M = 137.04 ms, SD = 6.06), (F [1, 10] = 28.8565, p < 0.001).

3. Conclusions

Our analyses of the temporal characteristics which distinguish younger from mature speech in our corpus are mostly in line with findings reported previously in the research literature for other languages. Our group of older female speakers made almost twice as many pauses as the young group, and the pauses were of longer duration compared to pauses made by the young speakers. The "79-88 GROUP" also realized intonation phrases which were on average 300 ms longer, and syllables which were on average 46 ms longer than those realized by the "19-23 GROUP" of young female speakers.

As far as the F0 characteristics which we analysed are concerned, our results are not in full conformity with published findings. Contrary to many previous studies, we found that the elderly speakers who took part in our investigation used a wider pitch range than the younger speakers. This finding, however, seems to corroborate doubts about the universal nature of F0 changes as cues to ageing ([9], [19].

Perhaps the most interesting of our findings concerns the distribution and use of pitch accent and boundary tone types. Although both groups made use of the same inventory of tones, they differed significantly in the frequency counts of some of the tones. Young speakers used pre-nuclear rises with a post-tonic high target (L*+<H and L*+H) much more often than older speakers, while older speakers used pre-nuclear rises with a high target reached within the stressed syllable (L+<H* and L+H*) somewhat more often than young speakers. The nuclear pitch accent which was used most frequently by the young speakers was L*, whereas the one which was used most frequently by the elderly speakers was L+H*. Generally, in our data younger speakers tended to use more phrase accents, while mature speakers seemed to prefer boundary tones and "level" (H-L% and HL-) pitch curves.

To our knowledge, no comparable data on tonal use by young vs. elderly speakers has been hitherto reported in the literature. The statistically significant results that we obtained seem to suggest that this line of research could offer interesting insights into the age-related differences between speakers.

4. References

[1] S. Schötz, "Acoustic Analysis of Adult Speaker Age," in C. Müller (ed), *Speaker classification I. Lecture notes in computer science* (vol. 1). Berlin: Springer, pp. 88–107, 2007.

[2] W. Spiegl, G. Stemmer, E. Lasarcyk, V. Kolhatkar, A. Cassidy, B. Potard, S. Shutn, Y. Song, P. Xu, P. Beyerlein, J. Harnsberger, E. Nöth, "Analyzing features for automatic age estimation on cross-sectional data," in *Proceedings of Interspeech 2009*, Brighton, United Kingdom, pp. 2923–2926, 2009.

[3] J. Volín, T. Tykalová, T. Bořil, "Stability of prosodic characteristics across age and gender groups," in *Proceedings of Interspeech 2017*, Stockholm, Sweden, pp. 3902–3906, 2017.

[4] C. Müller, "Automatic recognition of speakers' age and gender on the basis of empirical studies," in *Proceedings of Interspeech 2006*, Pittsburgh, PA, 2006.

[5] M. Li, K. J. Han, S. Narayanan, "Automatic speaker age and gender recognition using acoustic and prosodic level information fusion," *Computer Speech and Language* vol. 27, pp. 151–167, 2013.

[6] L. D. Shriberg, R. Paul, J. L. McSweeney, A. Klin, D. J. Cohen, F. R. Volkmar, "Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome," *Journal of Speech Language and Hearing Research* vol. 44, pp. 1097–1115, 2001.

[7] D. R. Barnes, "Age-related changes to the production of linguistic prosody," Open Access Theses, Paper 17, 2013.

[8] S. S. Waller and M. Eriksson, "Vocal age disguise: the role of fundamental frequency and speech rate and its perceived effects. *Front. Psychol.* 7, article 1814, pp. 1–10, 2016.

[9] S. E. Linville, "The sound of senescence," *Journal of Voice* vol. 10, pp. 190–200, 1996.

[10] J. D. Harnsberger, R. Shrivastav, W. S. Brown, Jr., H. Rothman, H. Hollien, "Speaking rate and fundamental frequency as speech cues to perceived age", *Journal of Voice* vol. 22, pp. 58–69, 2008.

[11] P. Boersma, D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.0.36, retrieved 11 November 2017 from http://www.praat.org/

[12] D. R. Ladd, *Intonational phonology*, Cambridge: Cambridge University Press, 1996.

[13] H. Reetz, *Artikulatorische und akustische Phonetik*, Wissenschaftlicher Verlag, Trier, 1999.

[14] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, J. Hirschberg, "ToBI: A standard for labelling English prosody," *Proceedings of the Second International Conference on Spoken Language Processing (ICSLP 92)*, Banff, Alberta, 1992, pp. 867–870, 1992.

[15] B. Andreeva, *Zur Phonetik und Phonologie der Intonation in der Sofioter Varietät des Bulgarischen*, PhD dissertation, Saarbrücken: Universität des Saarlandes, 2007.

[16] B. Andreeva, W. J. Barry, J. Koreman, "Local and global cues in the prosodic realization of broad and narrow focus in Bulgarian," in M. Zygis, Z. Malisz (eds), *Slavic perspectives on Prosody* (Special issue of *Phonetica* vol. 73, 2016, pp. 256–278), 2017.

[17] S. Dimitrova, S.-A. Jun, "Pitch accent variability in focus production and perception in Bulgarian declaratives," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, United Kingdom, 2015, Paper number 0832, retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0832.pdf

[18] G. Demenko, B. Möbius, and B. Andreeva, "Analysis of pitch profiles in Germanic and Slavic languages," *Forum Acusticum 2014*, 7–12 September, Kraków, Poland, 2014.

[19] A. Markó, J. Bóna. "Fundamental frequency patterns: The factors of age and speech type," in *Proceedings of the Workshop "Sociophonetics, at the crossroads of speech variation, processing and communication"*, Pisa, 14–15 December 2010, pp. 45–48, 2010.