

ACOUSTIC CORRELATES OF PROSODIC DIMENSIONS IN YOUNGER AND OLDER SPEAKERS OF CZECH

JAN VOLÍN AND TOMÁŠ BOŘIL

ABSTRACT

The present study reports phonetic data applicable for diagnostic purposes in voice related pathologies. However, apart from purely physiological concern, linguistic considerations are also acknowledged since the speech material consists of a continuous spoken text. Three age groups of speakers were recorded (young, middle-aged and old adults), each represented by 15 men and 15 women ($n = 90$). Several measures of fundamental frequency, together with variation in intensity and speech tempo were captured. An appreciably innovative metric, Cumulative Slope Index (CSI), was successfully employed to capture F0 variability in utterances. The results confirm differences between the age groups, but also between men and women, and contribute the normative mapping of the Czech population.

Key words: intonation, aging, diagnostics, articulation rate, fundamental frequency

1. Introduction

The speech modulation in the domains of the fundamental frequency, amplitude, timing and spectral setting has been shown throughout decades of linguistic research to encode multiple meanings in communication. The perceptual correlate of this acoustic aggregate of four dimensions is commonly termed *speech prosody* (or intonation in the broad sense of the word) and has been acknowledged to participate significantly in the sound structure of speech or sound patterns of languages (for detailed accounts see, e.g., Bollinger, 1978; Gussenhoven, 2004; Ladd, 2008; Büring, 2016). Various functions of prosodic events are classified into categories, such as *lexical* (phonemic function in the phonological composition of a single word), *grammatical* (signalling types of sentences and indicating their constituents of lower order), *affective* (revealing attitudes and interpersonal stances, moods, emotions), *pragmatic and discourse* (guiding the attention of the recipient and managing his consequential behaviour) and *indexical* (indicating the membership of the speaker in various social groups). It is the last one that relates most directly to the topic of our present study.

Language communities can be described in terms of groupings of the language users who share certain patterns of speech behaviour. These groups can be determined, for instance, by geographical region, education, socioeconomic class, gender or age. Each of such criteria provides interesting and eventually applicable information about the speakers' habits both in the area of speech production and speech perception. The age criterion, for instance, may inform the fields of cognitive and developmental psychology, language and speech acquisition, but also medical disciplines. The latter motivated our research. Our principal question was whether there is a possibility to capture differences between younger and older adults whose mother tongue is Czech, but also what the typical values describing Czech speakers are. In stipulating medical diagnoses, the parameters of speech can be often useful as long as the standards for healthy population are known. In other words, if normal age-related changes are reliably described, pathological conditions that affect voice can be identified.

The interest in the influence of age on voice performance can be traced back across centuries. R. Baken cites Shakespeare who commented on the voice of an elderly man as follows: "... turning again toward childish treble, pipes and whistles in his sound" (Shakespeare, *As You Like It*, Act 2, Scene 7, quoted in Baken, 2005). Yet our attention is attracted quite naturally by less poetic and more systematic approach to the issue: the empirical research of the modern era, i.e., more or less the twentieth century. Here, again, the evidence of keen interest can be observed, but, in addition to that, the methodologically plausible routines dating back many decades (e.g., Bach, Lederer & Dinolt, 1941; Macklin & Macklin, 1942; Birren, 1956 or Mysak, 1959). However, despite the achievements of the past, the available literature also documents that the search for new data and new models is inevitable: R. Baken, one of the leading researchers in the field, expanded and reinterpreted some of his earlier views (cf. Baken, 1987 and Baken, 2005), or S. Linville pointed out serious methodological problems in earlier measurements (Linville, 2000: 364) In addition to that, the research in prosody shows that individual languages may impose specific constraints on physiological mechanisms. Pitch range, for instance, is systematically narrower in Czech than in English in comparable spoken texts (Volín, Poesová & Weingartová, 2015). Therefore, the studies that were carried out abroad and mapped certain important trends cannot be just translated for domestic purposes. It is necessary to investigate the population norms across the linguistically diverse communities.

The age of a speaker has its "signature" in his or her speech. It is known that listeners can differentiate between older and younger voices with remarkable confidence. In research where respondents listened to continuous texts (which is the material of the present study as well), the correlations between perceived and actual age reached the magnitude of about 0.9, which signals a very strong deterministic link (Shipp & Hollien, 1969; Ryan & Capadano, 1978; Hartman, 1979). Hartman also found that female listeners could assess the age of the speaker better than male judges. Yet, if asked about characteristics of older voices, the lay listeners do not necessarily know what really guides their decisions. Certain controversies can be noted in a table compiled by Linville (2000: 361). The listeners, for instance, claim, that elderly persons speak with lower voices regardless their gender. Empirical evidence indicates that this is true only for women, whereas for men the trend is usually ascertained as inverse. Similarly, it is generally believed that aging

causes vocal tremor (*ibid.*). However, many studies have demonstrated that instability in phonation is linked to physical health more strongly than to age. Linville actually suggests that larger frequency and amplitude variation might be better predictors of aging than small fluctuations known as tremor (also jitter and shimmer), not only but also because of methodological problems with measurements (Linville, 2000: 364).

The question arises then as to what methodology is suitable and reliable if we want to quantify the differences that humans distinguish instinctively (or rather implicitly). In addition, it would be useful to know what the contributing factors are and which of them are purely physiological, and which phonological, i.e., connected with different linguistic norms used by younger and older speakers. Our present study contributes modestly to this area of inquiry by testing methodology and by producing descriptive parameters for three age groups of Czech population.

2. Method

2.1 Recordings

Three age groups of native speakers of Czech were recruited. We required healthy individuals without neurological ailments or speech/larynx therapy, with sufficient eyesight and hearing typical for the given age (no special treatment or hearing aids). We also ascertained smoking habits of the subjects as smoking is known to have substantial effect on voice – it generally enhances the aging effects (Gilbert & Weismer, 1974; Braun & Rietveld, 1995; Tarafder, Datta & Tariq, 2012). Hence, smokers were not included in the present sample. The age divisions were: 20–39 years of age for young adults, 40–59 years for middle-aged adults, and 60–79 years for old adults. Each group was represented by 30 subjects (15 male + 15 female). The recording took place in a quiet, comfortably furnished office with a short natural reverberation. High-quality microphone was plugged directly to a portable recorder with uncompressed signal capturing, even though for the parameters investigated in our study the high quality of the recording is not as crucial as it would be, for instance, in the case of jitter, shimmer or breathiness measurements.

The respondents were asked to read out a short extract from a book by a well-known Czech author Karel Čapek (retrieved from selected writings published in Čapek, 1983). The text comprising 137 words does not contain any unusual lexical items or syntactic structures and the speakers were given time to get acquainted with it prior the reading. No special instruction concerning the style of reading was provided. For easier manipulation and better analytical insight, the recordings were divided into 12 units (hypothetical breath-groups, see Table 1).

Table 1. The division of the text into ideal breath-groups for easier analysis.

1	I na tom, že člověk si opatří psa, aby nebyl sám, je mnoho pravdy.
2	Pes opravdu nechce být sám.
3	Jen jednou jsem nechal Mindu o samotě v předsíni.
4	Na znamení protestu sežrala všechno, co našla, a bylo jí pak poněkud nedobře.
5	Po druhé jsem ji zavřel do sklepa s tím výsledkem, že rozkousala dveře.
6	Od té doby nezůstala sama ani po jedinou minutu.
7	Když píše, chce, abych si s ní hrál.
8	Když si lehnu, považuje to za znamení, že si mně smí lehnout na prsa a kousat mě do nosu.
9	Přesně o půlnoci s ní musím provádět velkou hru, při níž se s velkým hlukem honíme, koušeme a kutálíme po zemi.
10	Když se urtí, jde si lehnout.
11	Pak si smím lehnout i já, ovšem s tou podmínkou,
12	že nechám dveře do ložnice otevřené, aby se Mindě nestýskalo.

2.2 Analyses

The sound supervision and measurements were performed predominantly in Praat v6.0.14 which, apart from computations themselves, allows for labelling the sounds (Boersma & Weenink, 2016). Intensity contours were obtained with 10-millisecond time step by cubic interpolation from intensity objects (minimum pitch 50 Hz, time step auto). F0 contours and stylized F0 contours based on the tonal perception model (Mertens & d'Alessandro, 1995) were computed in Prosogram v2.13 (Mertens, 2004) with the following settings: Calculate intermediate data files (no graphics files), Time range all, F0 detection range 0–450 Hz, parameter calculation Full (saved in file), Frame period 0.005 sec, Segmentation method Automatic: acoustic syllables, Thresholds $G = 0.16 - 0.32/T^2$ (adaptive), $DG = 30$, $dmin = 0.050$. All frequency values were subsequently converted to semitones (ST) with the reference value of 100 Hz.

The processing of the extracted measurements and analyses of the data were performed in R (R Core Team, 2016) with the application of rPraat (Bořil & Skarnitzl, 2016). To focus on F0 contour variations that is not under the speakers' control, we have also created *alternative F0 contours* by subtracting linearly stylized contours from the raw F0 tracks represented by pitchtiers. Apart from the arithmetic mean we also calculated F0 range (from 5th to 95th percentile). One data point in correlation scatterplots and calculations represents a speaker. (We wanted to avoid data inflation resulting from regarding each breath-group a data point.) Also, due to the log-normal distribution of the measured phenomena, we had to perform logarithmic transformation of the data.

Apart from conventional measurements we also measured the Cumulative Slope Index (Hruška, 2016; for detailed evaluation see also Hruška & Bořil, 2017 – this volume). It is a metric that allows for quantification of the amount of variation in contours (e.g., F0 or intensity contours in relation to their duration). We have adapted this measure for speech in that instead of physical time we used the number of syllables as a timing unit. Cumulative Slope Index (CSI) relative to the number of syllables is computed as follows:

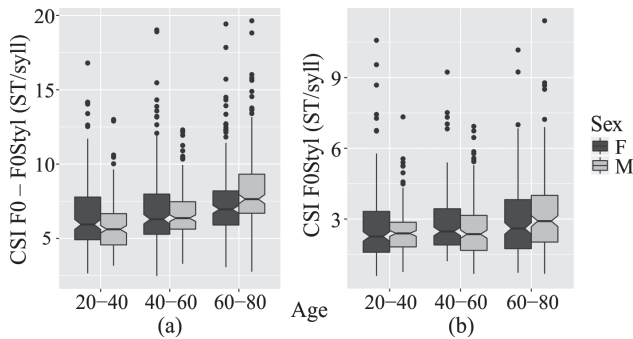
$$CSI = \frac{1}{N_{syll}} \sum_{n=2}^N |x(n) - x(n-1)|$$

where N_{syll} is the number of syllables in the utterance, N is the number of discrete points in the analysed contour and $x(n)$ is the value of the n -th point (either in semitones for pitch contours or in decibels for intensity contours).

3. Results

As our primary interest was in the Cumulative Slope Index (CSI), which was an innovative (or less commonly used) metric designed to reflect variation in contours, we report it first. Figure 1 comprises two panels. Panel (a) on the left captures the variation in alternative F0 contours (see above), whereas panel (b) on the right relates to the stylized contours of F0. The alternative contours suggest that with growing age the variation grows as well and the trend is more salient in male than in female production. The stylized contours (panel b) do not display the trend. This means that the crude melodic course which reflects the phonology of the language is not differentiating between older and younger speakers in terms of CSI, while the melodic movements that are outside phonology (we might think of them as being outside the speakers' control) have discriminative power.

Figure 1. Cumulative Slope Index (in semitones per syllable) (a) for the alternative contours (i.e., stylized contour subtracted from the raw F0 track), (b) for the plain stylized contours for three age groups of speakers with gender differentiated. (Notice the difference in the scale of the y axis between panels a and b.)



Variation of F0 above pertains to intonation in the narrow sense of the word. Figure 2 complements the previous account with data that map the tempo and loudness modulations (i.e. speech rate and intensity, respectively). It is obvious that the articulation rate decreases with age. A visible exception is the group of middle-aged males, but considering the confidence intervals, this exception does not seem to be very important.

Cumulative Slope Index for intensity contours seems to grow with age for female speakers, but young males go against this trend in the male sample. It has to be emphasised, though, that the relationship between intensity and loudness is extremely complex so attempts to explain this trend should remain very cautious.

Figure 2. Panel (a) – speech rates in syllables per second and Panel (b) – Cumulative Slope Index for intensity in decibels per syllable for three age groups of speakers with gender differentiated.

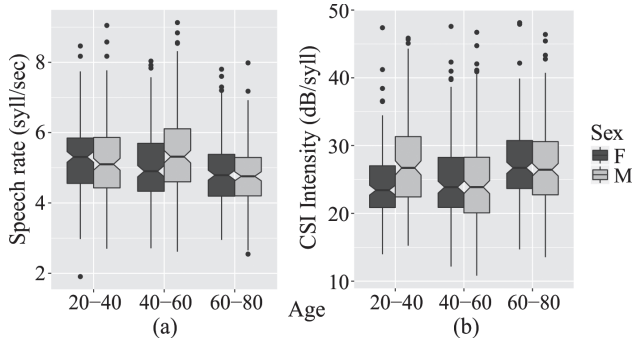


Figure 3. Panel (a) – mean fundamental frequency in semitones (re 100 Hz) and panel (b) – F0 range between the 5th and 95th percentile for three age groups of speakers with gender differentiated.

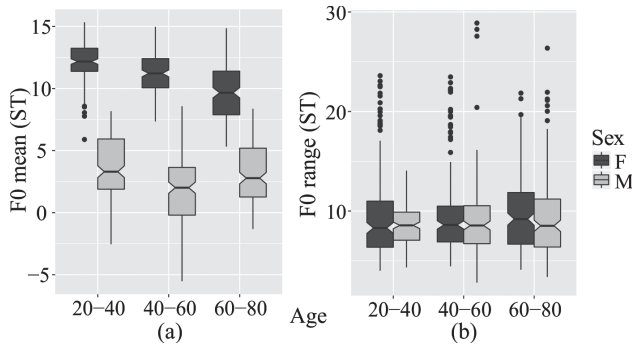


Figure 3 confirms to some extent the antagonistic behaviour of F0 level for male and female population found elsewhere. Women speak with lower voices as they grow older, men produce a scooping trend with the middle-aged group using the lowest values. The question is whether this scoop relates to the intensity variation (Figure 2, panel b), which captures a similar trend. The 90%-range of F0 values did not reveal any age related trend, although the male samples visibly increase the variance in the range values while keeping the median more or less equal. In other words, as to the F0 range, young males produced a compact set of values, while old males displayed a notable within-group dispersion of values.

Figure 4 captures correlations of speech rate with our two measures of contour variation. Both men and women behave in a very similar manner. On the left, the intensity is inversely correlated with the tempo at $r = -0.71$ (Pearson corr. coefficient). This suggests that the faster the speech, the less varied in intensity it is. A similar, but weaker trend is displayed in panel *b* of Figure 4. The alternative F0 contours (i.e., those where stylization is subtracted from the F0 track – see above) also vary less at higher tempos. The Pearson correlation coefficient $r = -0.44$. As already stated, these trends should be only interpreted after further experiments (but see Discussion below).

Figure 4. Scatterplots with trendlines (and 95%-confidence bands): panel (a) – speech rate against variation in intensity, panel (b) – speech rate against variation in F0 (alternative contours). Female data points and trendline are black, male are grey.

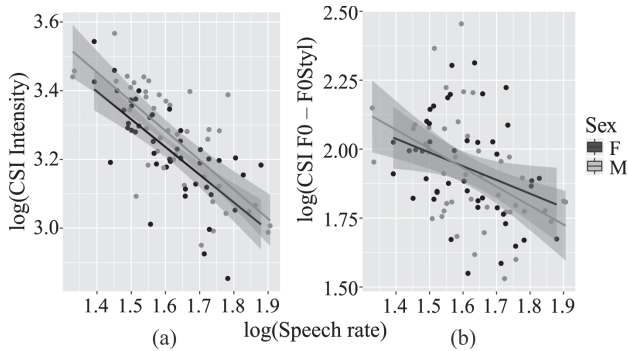
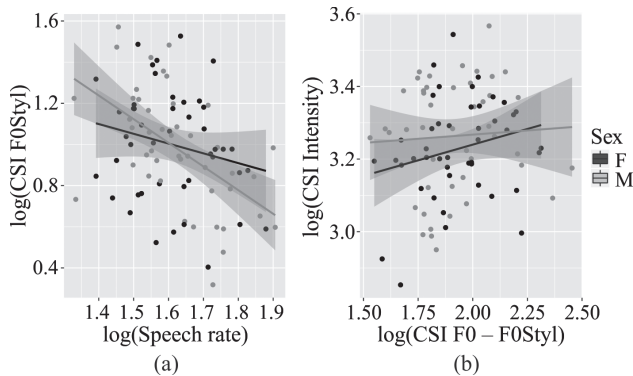


Figure 5. Scatterplots with trendlines (and 95%-confidence bands): panel (a) – speech rate against variation in stylized F0 contours, panel (b) – variation in F0 (alternative contours) against variation in intensity contours.



For the sake of completeness we also correlated speech rate with the stylized F0 contours which are believed to capture basic phonological properties of the intonation (Figure 5, panel *a*). The trend is not very strong: Pearson correlation coefficient $r = -0.38$. Even weaker is the link between variation in intensity and F0 in alternative contours: Pearson correlation coefficient $r = 0.19$. Here we might speculate about the common denominator, the speech rate, provided it really influences the variation in the two plotted domains.

4. Discussion

Prosodic variations in spoken texts fulfil important communicative tasks, but they also accommodate for individual and group variation. Our task was to map the values typical of selected age groups in Czech population. In general, we confirmed a trend found elsewhere that with aging, men increase their mean F0 (\approx pitch level), while women

speak lower (see Ryalls et al., 1994 for overview). In addition, the decrease in articulation rates as a function of age was confirmed in our sample. It has to be stressed, though, that our primary goal was to obtain values typical of Czech population, rather than replicate findings from other studies.

What we find noteworthy is that our material also suggests an increase in variability of F0 and intensity contours with age. On the one hand, this might be a manifestation of the fact that fast articulation rates do not allow for proper prosodic variation: fast speakers make fewer prosodic boundaries and fewer prominences. On the other hand, the informal auditory inspection of the recordings suggested a pragmatic factor: it sounded as if the older speakers enjoyed the performance and wanted to show how lively or engaged their reading can be. This is also confirmed by the behaviour of some of the elderly subjects who wanted to recite poems they learnt by heart at school when they were students. Contrary to that, younger readers displayed mild anxiety not to make reading errors, or to read in a flawless manner. They sounded as if inexperienced with the use of their voice for loud reading. It seems that perceptual testing and development of some more linguistically sensitive metrics will be inevitable if we aspire at solving this and similar dilemmas.

The previous sentence has some bearing on another issue raised in the present study. We noticed that alternative F0 contours capture the age differences while the stylized F0 contours do not. (As explained in the section *Method* above, alternative contours are residual contours after subtraction of stylized contours from the raw extracted F0 tracks.) It is widely believed that stylized contours capture the intonational phonology of a language more clearly than raw contours (see Hermes, 2006 for overview). However, our CSI metric is not sensitive to the alignment of the melodic movements with the syllables of the utterances. Therefore, no definite statement about intonational phonology of younger and older Czech speakers can be made at this stage.

The age as a factor of variation in prosodic features is traditionally considered extralinguistic, and as such has been mostly studied in medical research, recently also by sociolinguistics. “Extralinguistic” is an immensely unfortunate label. Linguists should not light-heartedly relinquish anything that is connected with speech communication for others to study. If they do (and they have done formerly), many interesting facts will be discovered outside linguistics, whilst grammar itself will not guarantee understanding the competences of language users and their speech behaviour.

ACKNOWLEDGEMENT

This research was supported by the Czech Science Foundation project No. GA16-19975S “Age-related changes in acoustic characteristics of adult speech”.

REFERENCES

- Bach, A. C., Lederer, F. L. & Dinolt, R. (1941). Senile changes in the laryngeal musculature. *Arch. Otolaryng.* 34, pp. 47–56.
- Baken, R. J. (2005). The aged voice: A new hypothesis. *Journal of Voice* 19/3, pp. 317–325.

- Baken, R. J. (1987). *Clinical Measurement of Speech and Voice*. Boston: Allyn & Bacon.
- Birren, J. E. (1956). The significance of age changes in speed of perception and psychomotor skills. In: J. E. Anderson (Ed.) *Psychological Aspects of Aging*. Washington: Amer. Psychol. Association.
- Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.14, retrieved 11 February 2016 from <http://www.praat.org/>.
- Bollinger, D. (1978). Intonation across Languages. In: J. Greenberg (Ed.). *Universals of Human Language*, Vol. 2. pp. 471–524. Stanford: Stanford University Press.
- Bořil, T. & Skarnitzl, R. (2016). Tools rPraat and mPraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech, and Dialogue*. Springer International Publishing, 367–374.
- Braun, A. & Rietveld, T. (1995). The influence of smoking habits on perceived age. In: *Proceedings of 13th International Congress of Phonetic Sciences*, Vol. 2, pp. 294–297. Stockholm: IPA.
- Büring, D. (2016). *Intonation and Meaning*. Oxford: Oxford University Press.
- Čapek, K. (1983). *Spisy: Zahradníkův rok / Měl jsem psa a kočku / Kalendář*. Praha: Československý spisovatel.
- Gilbert, H. R. & Weismer, G. G. (1974). The effect of smoking on the speaking fundamental frequency of adult women. *Journal of Psycholinguistic Research* 3. pp. 225–231.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge Univ. Press.
- Hartman, D. E. (1979). The perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders* 12, pp. 53–61.
- Hermes, D. J. (2006). Stylization of pitch contours. In: S. Sudhoff et al. (Eds.) *Methods in empirical prosody research*. Berlin: Walter de Gruyter, pp. 29–61.
- Hruška, R. (2016). Properties of fundamental frequency contours in segmental contexts. *Prague: Institute of Phonetics, Charles University, Faculty of Arts, bachelor thesis*.
- Hruška, R. & Bořil, T. (2017). Temporal variability of fundamental frequency contours. *AUC – Philologica, Phonetica Pragensia*, pp. 35–44.
- Macklin, C. & Macklin, M. (1942). Respiratory system. In: E. V. Cowdry (Ed.) *Problems of aging* (2nd ed., pp. 185–253). Baltimore: Williams & Wilkins.
- Mertens, P. & d'Alessandro, C. (1995). Pitch contour stylization using a tonal perception model. In: *Proceedings of 13th International Congress of Phonetic Sciences*, Vol. 4, pp. 228–231.
- Mertens, P. (2004). The Prosogram: Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In: B. Bel & I. Marlien (Eds.) *Proceedings of Speech Prosody 2004*, Nara (Japan): ISCA.
- Mysak, E. (1959). Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research* 2, pp. 46–54.
- Ladd, D. R. (2008). *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.
- Linville, S. E. (2000). The aging voice. In: R. D. Kent & M. J. Ball (Eds.) *Voice Quality Measurement*, pp. 361–376.
- R Core Team (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. URL <https://www.R-project.org/>.
- Ryalls, J., Le Dorze, G., Lever, N., Ouellet, L. & Larfeuil, C. (1994). The effects of age and sex on speech intonation and duration for matched statements and questions in French. *JASA* 95/4, pp. 2274–2276.
- Ryan, E. B. & Capadano, H. I. (1978). Age perceptions and evaluative reactions toward adult speakers. *Journal of Gerontology* 33, pp. 98–102.
- Shipp, T. & Hollien, H. (1969). Perceptions of the aging male voice. *Journal of Speech and Hearing Research* 12, pp. 703–710.
- Tarafder, K. H., Datta, P. G. & Tariq, A. (2012). The aging voice. *BSM Medical University Journal* 5/1. pp. 83–86
- Volín, J., Poesová, K. & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language* 13(1), pp. 107–123.

RESUMÉ

Studie přináší sadu referenčních fonetických deskriptorů, které je možno využívat pro diagnostické účely nejen ve foniatrické oblasti, ale i všude tam, kde se patologické změny odrážejí i v hlase pacienta. Ovšem kromě čistě fyziologických aspektů jsou pojednány i aspekty lingvistické, neboť řečovým materiálem je souvislý mluvený text, nikoli izolované jednotky nebo logatomy. Studie pracuje s nahrávkami tří věkových skupin dospělých mluvčích, z nichž každá je reprezentována 15 muži a 15 ženami (tj. celkový počet subjektů = 90). Měření byla prováděna ve třech základních akustických doménách: frekvenční, temporální a intenzitní, a to různými metodami. Jednu z nich, kumulativní index změny (Cumulative Slope Index – CSI) lze považovat za inovativní. Tento index v sebraných datech prokázal svou užitečnost. Celkově potvrdily výsledky studie rozdíly nejen mezi věkovými skupinami, ale i mezi muži a ženami v interakci s věkem. Tyto výsledky jsou příspěvkem ke stanovení chybějících populačních norem pro české jazykové prostředí.

Jan Volín, Tomáš Bořil
Institute of Phonetics
Faculty of Arts, Charles University
jan.volín@ff.cuni.cz