INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Stability of prosodic characteristics across age and gender groups

Jan Volín[1], *Tereza Tykalová*[2], *Tomáš Bořil*[1]

[1]Institute of Phonetics, Faculty of Arts, Charles University in Prague, Czech Republic
[2]Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

jan.volin@ff.cuni.cz, tykalova.tereza@gmail.com, tomas.boril@ff.cuni.cz

Abstract

The indexical function of speech prosody signals the membership of a speaker in a social group. The factors of age and gender are relatively easy to establish but their reflection in speech characteristics can be less straightforward as they interact with other social aspects. Therefore, diverse speaker communities should be investigated with the aim of their subsequent comparison. Our study provides data for the population of adult speakers of Czech – a West Slavic language of Central Europe. The sample consists of six age groups (20 to 80 years of age) with balanced representation of gender. The search for age and gender related attributes covered both global acoustic descriptors and linguistically informed prosodic feature extraction. Apart from commonly used measures and methods we also exploited Legendre polynomials, k-means clustering and a newly designed Cumulative Slope Index (CSI). The results specify general deceleration of articulation rate with age and lowering of F0 in aging Czech women, and reveal an increase in CSI of both F0 tracks and intensity curves with age. Furthermore, various melodic shapes were found to be distributed unequally across the age groups.

Index Terms: speech prosody, intonation, aging, gender, cluster analysis, Legendre polynomials

1. Introduction

The prosody of speech fulfills a number of important functions in human verbal communication (e.g., [1], [2], [3], [4]) and one of them concerns the membership of a speaker in a socially delineated group. This function is commonly labelled as *sociophonetic* or *indexical* [5]. The groupings of language users who manifest common patterns of speech production reflect both physiological and socially conditioned factors and point to the users' association with certain geographical region, socioeconomic class, education, ethnicity, age or gender. The information coded in the speech signal about the age and gender is relatively salient judging from the confidence with which listeners usually determine specifications of these two attributes [6]. It seems that estimates of, for instance, socioeconomic class or education are on average less accurate. As to geographically conditioned variation, there often exist 'regionless' forms in many language communities that have the status of a standard. Contrary to that, it is difficult to imagine speech without indices of age and gender even if exceptional individuals may generate deceiving cues.

The interest in age and gender specifics is very old and even the empirical research that we consider 'modern' dates back many decades (e.g., [7], [8], [9], [10]). Over the years it has been shown that with age the articulation rates decelerate, intensity decreases and fundamental frequency lowers in female speakers, while the male population generally displays increase or stable trends [11]. However, as to the male F0 aging trends, contrary findings can be encountered as well [12]. For this and many other reasons there are also warnings against self-complacent conclusions. The methodology of research is sometimes criticised [13, p. 361] and the idiosyncratic influence of the speech material on the results is pointed out [14, p. 9].

Given the interaction of language-specific and culture-specific constraints with the physiological universals, we stipulated three objectives of our study. First, we committed ourselves to provide reliable data concerning healthy Czech population so that subsequent comparisons with other samples are made possible. Second, we decided to test a newly designed measure – the Cumulative Slope Index (CSI – see below) – to establish its usefulness. Third, we wanted to complement a purely acoustic approach with linguistically informed analyses to see if they reveal any tendencies in age and gender variation. In this final goal, we are not trying to contribute to models that would predict the speaker's age, but to establish which features of Czech speech prosody are stable across generations of males and females and which are not.

2. Method

The following paragraphs describe the data collection for the study in terms of participants and recording procedure, and the methods of data analysis.

2.1. Participants

We recruited 200 speakers of Czech as their mother tongue for the study to cover the age span from 20 to 80 years. Both genders were represented more or less evenly: *n* females = 106 (mean age = 49.9 yrs; s.d. = 18.1 yrs) and *n* males = 94 (mean age = 52.6 yrs; s.d. = 18.2 yrs). As shown in Table 1 the 10-year age bands were not entirely balanced, but this was taken into account during the analytical procedures and interpretation of the outcomes. (The collection of the data currently continues to reach the balance and the target *n* = 300.)

Table 1: *Composition of the speaker sample.*

Age Band	Age Interval (years)	*n* Females	*n* Males
20s	(19,29]	19	17
30s	(29,39]	16	18
40s	(39,49]	8	14
50s	(49,59]	13	12
60s	(59,69]	25	14
70s	(69,79]	25	19

http://dx.doi.org/10.21437/Interspeech.2017-1503

All participants were Czech native speakers born or permanently living in or around the capital city of the Czech Republic (the region called Central Bohemia).

The exclusion criteria for speakers were the history of neurological or communication disorders, serious problems with respiration or hearing, suspicion of memory deficits and/or the active usage of antidepressants, antipsychotics or other drugs with direct effect on speech or mood. We also excluded heavy smokers as it is known that smoking generally enhances the aging effects [15].

The research project was approved by the local ethics committee and each participant provided written, informed consent for the recording procedure.

2.2. Recording procedure

The speech samples were recorded in a quiet, comfortably furnished office with a low level of ambient noise and short natural reverberation. A head-mounted condenser microphone (Bayerdynamic Opus 55) was plugged directly into a pocket recorder set to uncompressed 48 kHz 16-bit mode.

The speakers were asked to read out an extract from a book of narratives by a well-known Czech author [16]. The passage comprised 137 words and special care was taken to exclude low frequency lexical items or unusual syntactic structures. The speakers were allowed to get acquainted with the text prior the reading, but no special instruction concerning the style of reading was provided. The recordings were divided into 12 ideal breath-groups (i.e., hypothetical units delimited by spots of the highest probability of breath intakes).

2.3. Material processing and analyses

The speech sample processing was carried out mostly in the software Praat [17]. All recordings were carefully manually segmented on levels of phones, words and intonation phrases. Fundamental frequency (F0) tracks were extracted and manually corrected to eliminate errors typical of current F0 extractors (octave jumps, false periodicity, creaky voice artefacts). Intensity was measured in 10ms steps with cubic interpolation.

Apart from common statistical descriptors such as arithmetic means and standard deviation, several techniques that need to be specified were used. First, we calculated the Cumulative Slope Index (CSI) to capture variation in contours of F0 and intensity [18]. The following formula was used:

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

where N_{syll} is the number of syllables in the speech unit, 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 contour or in decibels for intensity contours).

Furthermore, Pearson correlations of dependent variables with age were calculated. In these correlations and related regression analyses each data point represented one speaker rather than all his/her utterances, which would inflate the values of statistical significance). In analyses of F0 contours we used their approximation by Legendre polynomials [19] implemented in rPraat [20], followed by k-mean clustering of the Legendre coefficients or, alternatively, of several equidistant points from the contour.

3. Results

3.1. Temporal domain

Even though the general trend of deceleration in speech rate with age has been confirmed by many studies, the detail of this relationship for various populations of speakers is always useful to provide. Figure 1 displays the trend found for Czech men and women. It shows that the articulation rate (AR) drops over the selected 60 years by slightly more than 1 syll/s in female population ($r = -0.520$) while in males it is only three quarters of a syllable per second ($r = -0.339$).

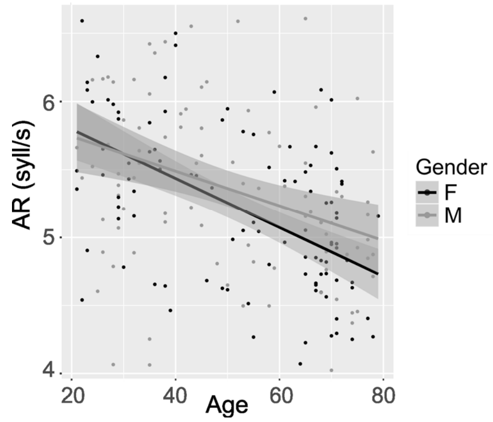

Figure 1: *Drop in speech tempo over years of life as reflected in scatterplot of articulation rate (syll/s) and age with regression lines for both genders and 95% confidence bands (6 outliers not depicted).*

Although AR is in inverse relationship with duration of speech units, due to disparate temporal behaviour of various speech units it is possible to gain insight into AR trends by measuring duration of individual classes of units. This is shown in Figure 2, which presents measurements of vowel durations against the ages of the speakers. While the female speakers follow a near-linear trend, there is a curious drop in vowel durations in middle-aged men compared with younger male adults. This drop also explains the lower r for linear fit in Figure 1.

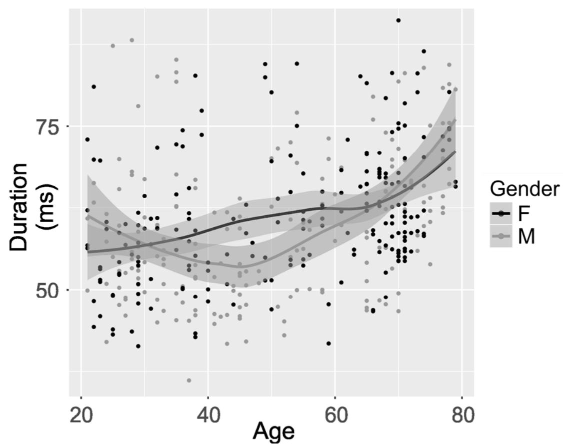

Figure 2: *Changes in vowel durations over the years of life as reflected in scatterplot of vowel durations (ms) and age with regression curves for both genders. LOESS regression (3 outliers not depicted).*

We further decided to investigate the stability of the temporal structure. For that purpose we took two subsequent stretches of speech and calculated the difference between the mean duration of phones in them. The stretches were about one intonation phrase long with the exclusion of the phrase-final word (with the phrase-final lengthening). Figure 3 shows that women reach comparable span of values across three 20-year age bands, while men are most stable in the middle band (40 – 60 years of age) and in younger and older age produce variation that is about twice as large.

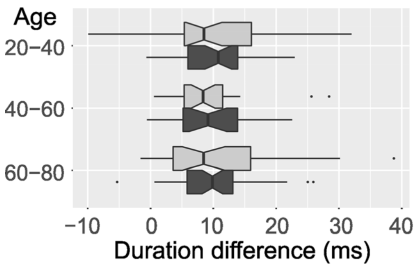

Figure 3: *Variation in phone duration differences between two neighbouring parts of the spoken text. (1 outlier not depicted).*

3.2. Variation in intensity contours

Cumulative Slope Index (CSI – see Section 2.3 above) also changes over the lifespan, yet in our sample only in female speakers. Figure 4 documents the increase and the associated Pearson coefficients are $r = -0.019$ for men and $r = 0.456$ for women.

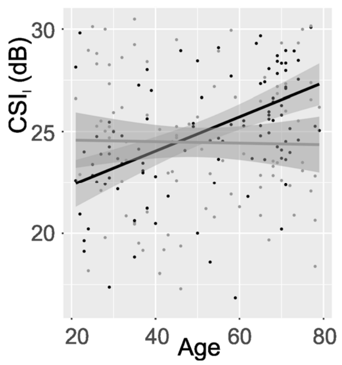

Figure 4: *Changes in intensity variations over the lifespan as reflected in scatterplot of CSI (dB) and age with regression lines for both genders (4 outliers not depicted).*

3.3. Fundamental frequency domain

First, our sample confirmed the trend reported elsewhere in literature (see Figure 5a): female voices grow lower with age ($r = -0.365$). However, often reported increase in male voices proved quite subtle ($r = 0.026$).

With regard to variation, both genders produced larger 90% range (measured from 5th to 95th percentile) at later age. For women the regression coefficient r was 0.348, while for men only 0.168. Since the measurements were carried out in semitones (ST), the difference is not an artefact of exponential Hertz scale.

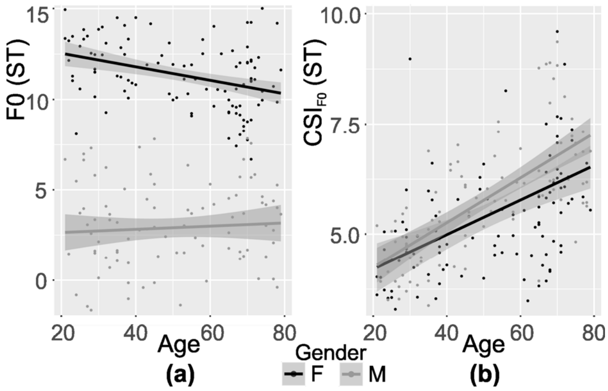

Figure 5: (a) *Changes in mean F0 over the lifespan as reflected in scatterplot of F0 (ST) and age with regression lines for both genders (6 outliers not depicted). (b) Changes in F0 variations over the lifespan as reflected in scatterplot of F0 CSI (ST re 100 Hz) and age with regression lines for both genders, (3 outliers not depicted).*

In parallel with intensity curves we also calculated Cumulative Slope Indices (CSI – see Section 2.3 above) for manually corrected F0 tracks converted to semitones. Figure 5b displays the situation. The regression coefficients were relatively high, especially for men ($r = 0.671$). This finding contrasts with the fact that the 90% range of male speakers increased only moderately with age, as reported in the previous paragraph.

3.4. More regressions and a linear model

Apart from correlations with age, we were also interested in mutual correlations of our dependent variables. We were particularly curious about the regressions of CSIs over the articulation rate. It turned out that for both genders the variation increased with deceleration. For intensity the male speakers produced $r = -0.631$, whereas female speakers returned $r = -0.736$. For F0 men produced $r = -0.381$, while women $r = -0.295$. Slower articulation rates are connected with higher cumulative slopes and the trend is stronger for intensity contours. Naturally, we have to avoid speculations about causality here. The changes are clearly interrelated, but our measurements do not allow for statements concerning causes and consequences.

Several multiple linear models were also tested of which the following makes most sense:

$$age_M = 75.0 + 7.8CSI_{F0} - 7.2AR_{syll} - 1.3CSI_I, \quad (2)$$

$$age_F = 55.7 + 3.7CSI_{F0} - 8.8AR_{syll} + 0.9CSI_I \quad (3)$$

where age_M and age_F are predicted ages for both genders, CSI_{F0} and CSI_I are CSI of F0 and intensity contours, and AR_{syll} is AR in syll/s. Although there is a great interpersonal variability, 50 % of age_M residuals lie between −9.6 to 9.3 yrs and 50 % of age_F residuals between −11.2 to 10.0 years.

3.5. Fundamental frequency contours in stress-groups

Several four-syllable stress-groups were selected in the text and their F0 profiles were subjected to further examination. (Four-syllable length was chosen since shorter stress-groups are more influenced by their surroundings and longer stress-groups are much less stable and frequent.)

First, Legendre polynomials were fitted onto the F0 contours of 4-syllable stress-groups (SGs). Afterwards, the Legendre coefficients were fed into a k-means cluster analysis with the requirement to find three types (Method 1). K-means cluster analysis was also performed with seven values taken directly from the F0 shape equidistantly with the fourth one centred and the edge ones 10 milliseconds from the onset and offset of the SG (Method 2). An example of the outcome can be observed in Figure 6. Clearly, both methods returned similar outcomes (Fig. 6, top).

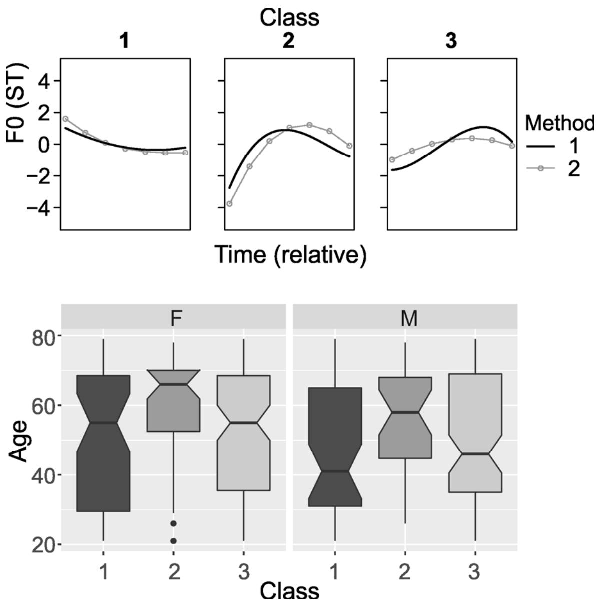

Figure 6: *The resulting F0 courses from cluster analysis (top) and preferences of the contour classes by age groups.*

The lower part of Figure 6 presents the choices by men and women of the three contours classes relative to their age. Contour of the class 2 (steep rise to early peak) is apparently preferred by older speakers. Moderate rise with late peak and slightly declining peak-less contours are more typical for younger speaker. However, the preferences are far from categorical.

4. Discussion

A relatively ambitious project of data collection (several hundred speakers with balanced representation of age groups and genders) provides opportunities to various types of segmental and suprasegmental analyses. Our present study dealt with several acoustically based measures to map prosodic variation in spoken texts produced by six selected age groups of 200 Czech speakers.

In general, we confirmed (but quantitatively specified for Czech population) trends found elsewhere. The deceleration in articulation rates as a function of age was confirmed in our sample, but it was also demonstrated that the trends need not be linear. Our Figures 2 and 3 show that middle-aged men are the fastest and most stable across utterances. Therefore, reanalysis of linear fits with non-linear algorithms is recommended.

It is no surprise that Czech women speak lower at older age, but it should be taken into consideration that men speak only negligibly higher. Together with the finding in [12] it begs for further expansion of the sample, perhaps with regard to speakers sociolinguistic descriptors. As to F0 variation, it clearly rises with age, but when measured by variation range, women appear more "varying" than men, while the opposite is suggested by the CSI measure. This methodological issue again should serve as an inspiration for future research.

We also found increased variation in intensity contours with age. On the one hand, this might be a manifestation of the fact that fast articulation rates tend to iron out prosodic variation: fast speakers make fewer prosodic boundaries and fewer prominences, on the other hand, individual local patterns of variation should be checked. We did that for F0 contours over four-syllable stress-groups and found that preferences for contours are also related to age.

The age should not be reduced to mere physiological deterioration. Aging is also a mental process that influences the attitudes of an individual to the surrounding world. Our elderly subjects seemed more self-confident and more talkative. In comparison with the youngest subjects they were also less anxious about making errors in the reading task.

The research presented here continues and apart from enlargement of the sample it will expand the analyses of the correlates of prosodic structure. It might be useful to mention that there is also a group of researchers who perform segmental analyses on the same material.

5. Conclusions

Two hundred Czech speakers aged 20 to 80 years provided read-aloud speech material for the study of influences of aging on speech production patterns. Findings in temporal, intensity and fundamental frequency domains confirmed differences not only among various age groups, but also between the genders. General trends known from previous research on other populations of speakers were corroborated, although not entirely. For instance, the male rise in F0 with older age was only very subtle. Also, some non-linearities were found in the material. Importantly, a quantification of the trends for native speakers of Czech living in the most populated area of the Czech Republic was made available.

6. Acknowledgements

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

7. References

[1] D. Bollinger, "Intonation across languages," in: J. Greenberg (Ed.). *Universals of Human Language*, vol. 2. pp. 471–524. Stanford: Stanford University Press, 1978.

[2] C. Gussenhoven, *The phonology of tone and intonation*. Cambridge: Cambridge Univ. Press, 2004.

[3] D. R. Ladd, *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press, 2008.

[4] D. Büring, *Intonation and Meaning*. Oxford: Oxford University Press, 2016.

[5] G. Docherty and P. Faukes, "The social life of phonetics and phonology," *Journal of Phonetics*, vol. 34, pp. 409–438, 2006.

[6] J. D. Harnsberger, R. Shrivastav, W. S. Brown, H. Rothman and H. Hollien, "Speaking rate and fundamental frequency as speech cues to perceived age," *Journal of Voice*, vol. 22, no. 1, pp. 58–69, 2008.

[7] A. C. Bach, F. L. Lederer and Dinolt, R. “Senile changes in the laryngeal musculature,” *Arch. Otolaryng. 34*, pp. 47–56, 1941.
[8] C. Macklin and M. Macklin, “Respiratory system,” in: E. V. Cowdry (Ed.) *Problems of aging* (2nd ed., pp. 185–253). Baltimore: Williams & Wilkins, 1942.
[9] J. E. Birren, “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, 1956.
[10] E. Mysak, “Pitch and duration characteristics of older males,” *Journal of Speech and Hearing Research 2*, pp. 46–54, 1959.
[11] S. E. Linville, “The sound of senescence,” *Journal of Voice*, vol. 10, pp. 190–200, 1996.
[12] W. Decoster and F. Debruyne, “Longitudinal voice changes: facts and interpretations,” *Journal of Voice*, vol. 14, no. 2, pp. 184–193, 2000.
[13] S. E. Linville, “The aging voice,” in: R. D. Kent & M. J. Ball (Eds.) *Voice Quality Measurement*, pp. 361–376, 2000.
[14] S. Schötz and C. Müller, “A study of acoustic correlates of speaker age,” in C. Müller (Ed.) *Speaker Classification II, LNAI 4441*, pp. 1–9. Berlin: Springer Verlag, 2007.
[15] A. Braun and T. Rietveld, “The influence of smoking habits on perceived age,” in *Proceedings of 13th International Congress of Phonetic Sciences*, vol. 2, pp. 294–297. Stockholm: IPA, 1995.
[16] K. Čapek, *Spisy: Zahradníkův rok – Měl jsem psa a kočku – Kalendář*. Praha: Československý spisovatel. 1983.
[17] P. Boersma and D. Weenink, “Praat – doing phonetics by computer,” version 6.0.26, retrieved from www.praat.org, 2017.
[18] R. Hruška, *Properties of fundamental frequency contours in segmental contexts*. (Bachelor thesis – unpublished) Prague: Institute of Phonetics, Charles University, Faculty of Arts, 2016.
[19] M. Abramowitz and I. A. Stegun, “Legendre Functions” and “Orthogonal Polynomials.” Chapters 8 & 22 in *Handbook of Mathematical Functions*, 9th printing. New York: Dover, pp. 331–339 and 771–802, 1972.
[20] T. Bořil and R. Skarnitzl, “Tools rPraat and mPraat,” in *P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), Text, Speech, and Dialogue*. Springer International Publishing, pp. 367–374, 2016.