Izvorni znanstveni rad
Rukopis primljen 2. 3. 2021.
Prihvaćen za tisak 10. 2. 2022.
https://doi.org/10.22210/govor.2021.38.09

Anna Huszár, Valéria Krepsz
huszar.anna@nytud.hu, krepsz.valeria@nytud.hu
Research Institute for Linguistics, Budapest
Hungary

The development of variability in pausing and articulation rate in Hungarian speakers ten years apart

Summary

In recent years, attention given to speech-related disciplines has shifted to analyzing extra- and paralinguistic information occurring in speech. Moreover, there is a growing public interest in ageing, which, in turn, may encourage the conduct of this type of research. Earlier studies have analyzed older speakers, and, to date, we have little information about the changes taking place in the voices of healthy, young speakers growing older. Our research examines how certain temporal patterns in speech change in two different types of speech (reading and spontaneous speech) over a period of 10 years. The analysis included speech samples from 13 healthy, young adult male speakers recorded under the same conditions ten years apart. The study focuses on the timing differences within and between the recordings. The development of temporal and pausing patterns were analyzed depending on the type of speech, position, and the time of the recording (10-year difference). The results showed that not only the type of speech but the differences within and between the recordings affected the pausing strategies. In addition, the data indicated that (1) a slower articulation rate was observed in spontaneous speech and (2) higher variability between sample types i.e., a more varied articulation rate was observed in spontaneous speech than in reading. In the case of the articulation rate the individual differences dominated.

Keywords: temporal patterns, longitudinal study, articulation rate, age, type of speech

1. INTRODUCTION

In aging Western societies, discerning the age of speakers and related stereotypes are key to successful intergenerational communication. However, most of the previous studies analyzed the relationship between speech prosody and age primarily in language-acquisition or in terms of speech pathology, therefore our knowledge on changes in healthy, speakers' speech characteristics can greatly contribute to our knowledge on this topic.

The interest in the process of aging in speech (whether young or old) dates back to the 1960s when Ptacek and Sander (1966) found that 78% of listeners were able to correctly identify whether vowels were produced by old or young speakers. When having to make judgments based on reading, this rate was 99%. Another study reports similar results about both production and perception: the articulation rate slows down with age, thus speakers with slower articulation rate are judged to be older. However, this trend was noticed only in reading, no such result was found in spontaneous speech comparing articulation rate, actual and estimated age (Brückl & Sendlmeier, 2003).

Precise assessment of human age judgment raises several questions: e.g., what type of parameters are needed to determine speakers' age? What kind of acoustic information is used by listeners to determine the biological age of speakers? The answers to these questions may be interesting and applicable primarily in medical context, for an example: the slower articulation rate may also be due to changing laryngeal functions and a decreased activity of the oral muscles and lips (Linville, 2001).

Research suggests that prosodic characteristics (and their changes) play a central role in age judgement (Winkler, 2007). Assessment, based on acoustic features, has been described mainly in relation to the following concepts: a relationship was found between fundamental frequency, formant values, modulation of vocal cord vibration, intonation, spectral energy distribution, and age (Brückl, 2007).

The most frequently studied parameter among the temporal parameters is the articulation rate. The rate depends on the age of speakers as well as on several other factors (Jacewicz, Fox, O'Neill, & Salmons, 2009), including temporality within a text. Timing is also determined by the type of speech: several studies have confirmed that the articulation rate of reading is faster than that of spontaneous speech (Bóna, 2014; Duchin & Mysak, 1987; Ramig, 1983). However, there are some contradictory findings. In a study by Jacewicz and his colleagues (2009), spontaneous speech proved

to be faster, while Kohári (2018), for an example, did not find significant differences between the two types of speech. In addition to tempo, pause may vary between the two speech types; spontaneous speech is characterized by higher pause rates, more frequent pauses, and consequently, shorter speech units compared to reading (cf. e.g., Bóna, 2013; Váradi, 2010; Walker 1988).

Regarding the articulation rate, several studies have shown that the spontaneous speech of elderly people was slower compared to the speech of young people (Brown, Morris, & Michel, 1989; Oyer & Deal, 1985; Smith, Wasowicz, & Preston, 1987; Yuan, Liberman, & Cieri, 2006; Zellner-Keller, 2006). Based on word and sentence repetition at normal and accelerated articulation rates produced by 10 younger (aged between 24 and 27 years) and 10 elderly (aged between 66 and 75 years) speakers, Smith et al. (1987) found that older speakers were able to produce a 20–25% slower speech rate than younger ones. The reasons for such a slowdown include longer syllable duration and longer pauses, compared to the production of younger speakers. The increased number and duration of pauses is associated with age-specific physiological changes that may explain the change in production as well as the cognitive differences between older and younger speakers, such as in discourse planning (Linville, 2001). In addition, the appearance of a larger number of pauses in older speakers may reduce the duration of speech segments (speech segments are defined as speech units bounded by two silent or filled pauses) probably due to a decreased elasticity of lung tissue (Linville, 1996). However, other studies show inconclusive results: some studies do not find a relationship between age increase and articulation rate decrease (e.g., Hoit, Watson, Hixon, McMahon, & Johnson, 1994 – based on a study of 80 individuals aged 20–30, 40–50, 60–70, and 80+; Linville, 2000 – based on a study of 80 people, 40 of whom were speakers between the ages of 19 and 24, and 40 speakers between the ages of 62 and 79). Using the data from the LangAge corpus, which includes the audio material of 48 male and female speakers (over the age of 70), speech duration was negatively correlated with the age of speakers, while there was no significant correlation between pauses and age (Gerstenberg, Fuchs, Kairet, Frankenberg, & Schröder, 2011). Other research suggests that the rate of articulation in the elderly is characterized by a greater degree of variability than in young people, so no unified conclusion can be drawn regarding the output of speakers based solely on their age (Linville, 2001).

Previous research used cross-sectional methodology: the effect of age, based on a comparison of the results obtained from different age groups, was compared. In

contrast, the number of studies using longitudinal methodology is low, and, in most cases, they have been conducted only with a small number of participants and covered a short period of time (cf. e.g., Brückl, 2007). Some of the difficulties lie in methodological reasons: data collection and ensuring the same conditions for the recordings are complex methodological tasks, as well as searching for the informants after a long period of time. However, the longitudinal method has a number of advantages: for example, it can be used to filter out the cohort effect, whereas differences in cross-sectional studies due to different career paths may make comparisons difficult, or the different biological predispositions or different medical histories. Longitudinal procedures make it easier to deal with heterogeneity, the effect of which has been demonstrated in the field of prosodic features: one of the most defining results in aging research is extreme individual variation (Ringel & Chodzko-Zajko, 1987). In our opinion, the most important aspects in making the recordings are to ensure the same conditions and to keep as many participants during the recording period as possible, who form a homogeneous group according to certain aspects (for example, professional speakers or not, lifestyle, monolingual or multilingual speakers etc.). The choice of the methodology of a given study (cross-sectional or longitudinal) is primarily determined by the study question.

The relationship between temporal features and age has been examined in childhood or in pathological cases using longitudinal methodology and only a few studies have used healthy adult speakers. Gersternberg and colleagues (2011) examined the concept of age-related deceleration based on older German and French speakers (70–80 years old) in the context of different prosodic parameters of articulation rate, duration of speech units, and complex interaction between pauses 10 years apart. They found significant individual differences related to age in all factors examined, such as the articulation rate, the duration of speech segments/units, and the number of syllables per speech segment, which increased and decreased to varying degrees. Moreover, some language-specific features were discovered: the articulation rate increased for French speakers, while it slowed down for German speakers. The interpretation of the results is partly in line with international trends: temporal features are explained by changes in the cardiovascular system associated with shorter respiratory times and lower lung capacity. In addition to physiological changes, individual lifestyles (e.g., smoking, noise exposure) can also affect the results.

In his dissertation, Shum (2008) tangentially elaborated on how the rate of articulation had changed in the speech of Queen Elizabeth II, based on a Christmas

broadcast. Although a detailed description of the methodology and specific values in the results part of the paper are not provided, it can be inferred only from the 8th figure in Shum's dissertation that the Queen's articulation rate gradually increased from the age of 20 to approximately 55, followed by a decline until the late 1980s.

Another study (Quené, 2013) examined the temporal features in the speech of Queen Beatrix, using longitudinal methodology, based on her 'Troonrede' ('throne speeches') recorded between 1980 and 2012 (from when the Queen was 42 until she was 74, a total of 9 recordings were analyzed). The results corroborated the age-related deceleration: although articulation rate values declined slightly in the first few decades, an acceleration was observed in later decades. A change in tempo was also observed in speeches: while earlier recordings (between 1980 and 1996) were characterized by a slower-faster-slower pattern, later recordings (between 1996 and 2012) were characterized by gradual acceleration. (Unfortunately, specific values are not available in this case either.)

1.1 Hypothesis

Previous studies have mostly analyzed older speakers, and, to date, we have little information about the changes taking place in the voices of healthy, young speakers as they get older. From the age of 30, up to 5–10% per decade the function of certain organs may deteriorate (primary aging, cf. Lalley, 2013). All this is closely related to lifestyle, external influences, possible diseases (secondary aging, cf. Busse, 2002). Thus, it can be assumed that even in the case of a 10-year study period, a change can be detected in the speakers' utterances along certain acoustic parameters.

Therefore, our study examines how certain timing characteristics of speech change in two different types of speech over a period of about 10 years. Our hypotheses are the following: there are differences in a) articulation rate and b) pausing (pause rate and frequency) based on the utterances of healthy, young adult men.

1. According to each type of speech for the reading task i. faster articulation rate and ii. more frequent pauses were expected.
2. According to the time elapsed between the two recordings i. a slowdown in the articulation rate 10 years later, ii. higher rates and more frequent pauses were expected.
3. Within each recording, also i. deceleration and ii. more frequent pauses were assumed to take place.

2. MATERIALS AND METHODS

Speech samples of 13 healthy, young adult male speakers were analyzed. The first recordings (in the figures hereafter Recording 1 (R1)) were selected from the BEA spontaneous speech database (see Neuberger et al., 2014 about the conditions of the recordings), with participants aged between 19 and 40 years (mean: 27 years, SD: 5 years). The second recordings collected about 10 years later (in the figures hereafter Recording 2 (R2)) were selected from a longitudinal database (see Gráczi et al., 2020 about the conditions of the recordings), aged between 29 and 50 years (mean: 38 years, SD: 5 years). In the case of the second recordings the technical equipment was the same as in the case of the first recordings. The participants were native Hungarian speakers and had no hearing or speech impairments. Speech samples in each timepoint i.e., from both databases included two types of speech: the spontaneous speech and the reading task. In the spontaneous speech part of the recording, the interviewer asked the speaker about their job, hobbies, family, and other free time activities. This recording unit is a quasi-monologue part of the task, as the interviewer speaks for the most part, with the aim of helping the speaker speak continuously by asking questions and making comments. In the reading part, the task was to read an informative text (consisting of 12 declarative sentences, 234 words in total). The participants were allowed to read the text to themselves before reading it aloud. There was no limit to the preparation time of the participants: they could read the text as many times as they wanted.

The recordings were annotated manually at the speech segment level, using Praat software (Boersma & Weenink, 2019). Interpausal units refer to the unit of speech that is bounded by pauses (silent pause or hesitation). Speech material was also analyzed with Praat software using the following method: the articulation rate was calculated using a script for each speech unit, in syll/s. Pause patterns were analyzed by the duration and frequency of pauses. For the analysis of possible temporal patterns within each recording, the recordings were divided into five equal-duration sections (hereafter divided into 20% sections), then, the articulation rate values of the speech sections we assigned to them. To compare the timing characteristics of the given parts more accurately, the articulation rate values were normalized (z-transform: $Z = (X-\mu) / \sigma$, where, in this case, Z is the normalized articulation rate, X is the current articulation rate, μ is the average articulation rate measured in the sample, and σ is the standard deviation of the articulation rate measured in the sample, per speaker +

by speech type). Another method of splitting was used to examine the timing locally: the total speaking time was divided into half-minute sections, and then the average articulation rate value was calculated for each half-minute section.

The duration ratio of the pauses was calculated by comparing the duration of pauses to the total speaking time of each speaker ($t_{SP}/(t_{SP}+t)$, where t_{SP} is the sum of the duration of the silent pauses and $t_{SEGMENT}$ is the sum of the durations of the speech segments). The frequency of pauses was analyzed in terms of the total speech time ($n_{SP}/(t_{SP}+t_{SEGMENT})$, where n_{SP} is the number of pauses, t_{SP} is the sum of the duration of the pauses, and $t_{SEGMENT}$ is the sum of the duration of the speech segments). In addition to the duration ratio of the pauses, it was important to examine the frequency because it was likely to be different for the same pause ratio (e.g., for fewer but longer pauses or more but shorter pauses, the pause ratio may be similar, but the frequency is higher in the second case). The change of the pause ratio and the frequency of the pauses were also observed in the 20% breakdown of the recordings, and a Praat script was used to extract the pattern of pause for each 20% part.

The development of articulation rate and pauses were analyzed by linear mixed models in the R program (R Core Team, 2018) with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), p-values were obtained using Satterthwaite approximation (lmerTest package, ANOVA function, Kuznetsova, Brockhoff, & Christensen, 2017). Fixed effects (intercept) were the recording time (recording 1 (BEA database), recording 2 (Longitudinal)), speech type (spontaneous speech or reading task) and parts divided by 20%. Random effects were given to speakers, and the dependent variables were the articulation rate, pause ratio, and pause frequency. For each parameter, we also constructed a random intercept and a random slope model (with the speaker as a random factor for each variable) and compared the two models (with the ANOVA function available in the lmerTest package, Kuznetsova et al., 2017). There was no significant difference between the models, so in the results section that follows, we present the random slope values giving a lower AIC number (Akaike, 1973). In case the significant differences occurred in the 20% breakdown, the groups were compared by Tukey post hoc test.

3. RESULTS

3.1 Articulation rate

There was a difference between the two types of speech in the articulation rate: when reading, the participants spoke faster (average rate: 6.24 syll/s) than in spontaneous

speech (average rate: 5.60 syll/s). Similar trend was found in Czech for the speech rate: 2–4.8 syll/s for spontaneous speech and 3.3–6.7 syll/s for reading (Palková, 1997, as cited in Tivadar, 2017). Similar trend can be seen in English: the speech rate was higher in reading (4.69 syll/s) than in spontaneous speech (3.94 syll/s) (Duchin-Mysak, 1987). Further difference was observed in the variance of the two speech types, the standard deviation is almost 1 syll/s greater for spontaneous speech (SD: 1.83 syll/s) than in the reading task (SD: 0.97 syll/s).

There is a smaller difference in the articulation rate values regarding the timepoint of the recording (10-year difference) than by speech type. In spontaneous speech, the difference was less than 0.05 syll/s between the averages (average $rate_1$ = 5.59 syll/s; SD_1 = 0.68 syll/s; average $rate_2$ = 5.61 syll/s; SD_2 = 1.91 syll/s). In the reading task, a slightly larger difference between the values of the two recordings was found, the difference between the averages did not exceed 0.3 syll/s, so it can be regarded as negligible (average $rate_1$ = 6.39 syll/s; SD_1 = 0.91 syll/s; average $rate_2$ = 6.1 syll/s; SD_2 = 1.01 syll/s).

Figure 1 gives a detailed picture of the development of the articulation rate by speaker.

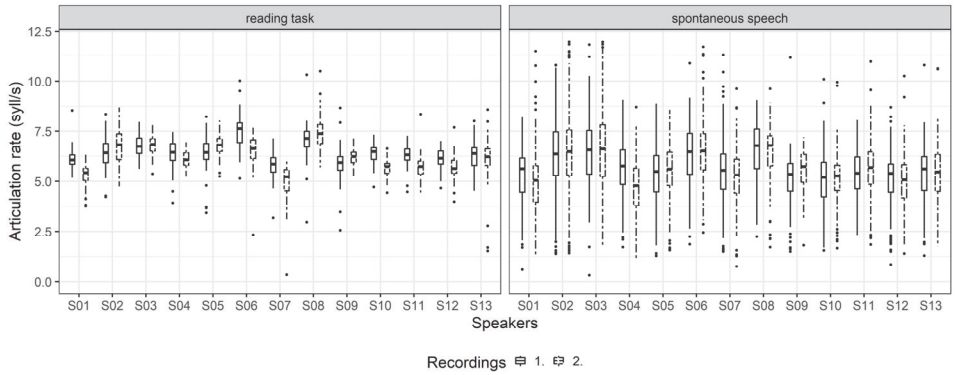

Figure 1. Articulation rate values by time of recording and breakdown of speakers in interview and reading task (median and interquartile range; S01, S02, …refer to the speakers)

Slika 1. Vrijednosti tempa artikulacije u različitim točkama snimanja i raščlamba prema govornicima u intervjuima i zadatku čitanja (medijan i interkvartilni raspon; S01, S02… odnose se na govornike)

The breakdown by speaker points out that the variance was not only smaller in the reading task, but also lower in all cases compared to spontaneous speech. For most speakers, the interquartile range was at least double in spontaneous speech compared to the reading task. For most speakers (10 speakers), the same tendency is evident: when reading aloud, the articulation rate was faster compared to spontaneous speech. For the other three speakers, the difference between the averages of the two speech types did not exceed 0.2 syll/s, therefore the difference can be considered negligible in their case.

For some speakers, the articulation rate values differed between the two recordings, but differences in both directions did not show a consistent trend: when the medians of articulation rate were compared, it was found that in spontaneous speech, the values for five speakers were lower while for other five speakers were higher in the second than in the first recording. Deviations of less than 0.1 syll/s were considered negligible and were measured for three speakers. In the reading task, the median values for eight speakers were lower while those for four speakers were higher in the second compared to the first recording. For one speaker, the difference between the two recordings was negligible. The difference between speech types was significant: $F(1, 6025) = 23.858$; $p < 0.001$).

As shown in Figure 2, there was no significant difference in the articulation rate depending on the section of the analyzed recording. In spontaneous speech, the large overlap in the interquartile ranges also suggests that the difference in the articulation rate between the different speech parts is negligible. This is also supported by the fact that the differences between the average values of the articulation rate do not exceed the value of 0.3 syll/s. According to the 20% breakdown, the average articulation rate of the spontaneous speech in the first recording was 5.79 syll/s (SD: 2.21 syll/s), while in the second recording, it was 5.80 syll/s (SD: 2.13 syll/s). The difference between the highest and lowest value in the first and second recordings, according to the 20% breakdown, was only 0.36 syll/s.

The difference in the same values in reading was already greater than in spontaneous speech. In both the first and second readings, the articulation rate of the first and the last 20% of sections were the highest, whereas the lowest was that of the middle (40–60%) sections. The difference between the highest and lowest values was 0.42 syll/s in the first recording and 0.65 syll/s in the second recording. According to the statistical analysis, there was no significant difference in the articulation rate according to the 20% breakdown.

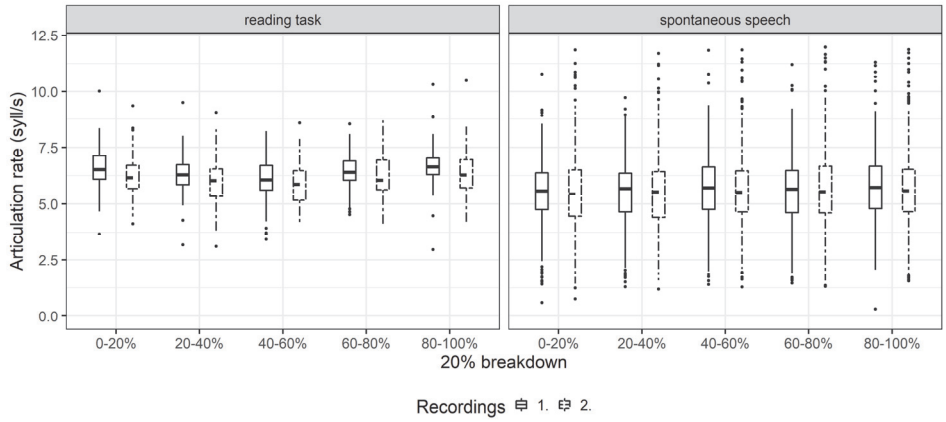

Figure 2. Development of articulation rate in 20% breakdown (median and interquartile range)

Slika 2. Razvoj tempa artikulacije u raščlambama od 20 % (medijan i interkvartilni raspon)

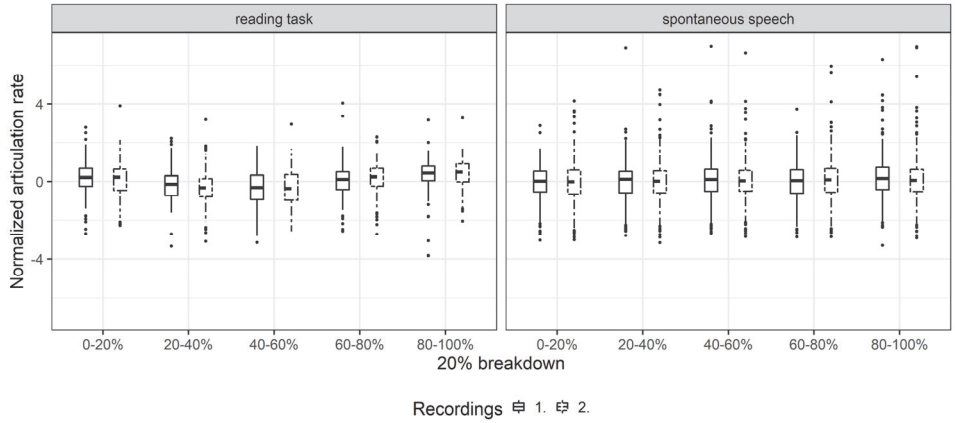

Figure 3. Change in normalized articulation rate at 20% breakdown (median and interquartile range)

Slika 3. Promjene u tempu artikulacije poslije normalizacije uzoraka u raščlambama od 20 % (medijan i interkvartilni raspon)

For a more accurate comparison, the articulation tempo values were normalized by z-transform (Figure 3) and the results confirmed the trends shown earlier: (1) The values in spontaneous speech showed a negligible difference between each part (the difference between the largest and smallest for the first recording was 0.06; for the second it was 0.05); (2) The highest values in the reading were measured at the beginning and the end of the recordings, whereas the smallest values were measured in the middle of the recordings. Thus, we can identify a tendency for reading faster at the beginning of the reading task, then gradually slowing down and reading faster again towards the end of the task (this is the case for recordings at both timepoints).

A 20% breakdown of the speech samples allows a comparison of all recordings in equal proportions. However, this method masks the time differences between recordings. Thus, this calculation method does not show any differences between shorter and longer shots. To solve this methodological difficulty, the reports were divided into 30-second sections (Figure 4).

The following trends are reflected in the results, based on the reading task: for most speakers, the first or second recording is faster, but this difference is systematic. This can be confirmed for each half-minute period, and the differences between the two recordings are not only reflected in the difference in mean values.

Although when a speaker accelerates or decelerates in a particular section varies, the change in the reading speed in both recordings reveals individual trends i.e., a speaker read the same text with nearly the same rate pattern 10 years apart. The figure showing the half-minute breakdown (Figure 4) clearly demonstrates the remarkable individual differences and the specific, individual temporal characteristics of the given speakers. The trend seen in the 20% breakdown of participants shows that reading is slower in the middle of the sample and it can be observed in the half-minute breakdown for almost half of the speakers. For the other participants, we do not find a significant difference between the tempo values in the individual parts and a different trend was noted for one speaker only (S02).

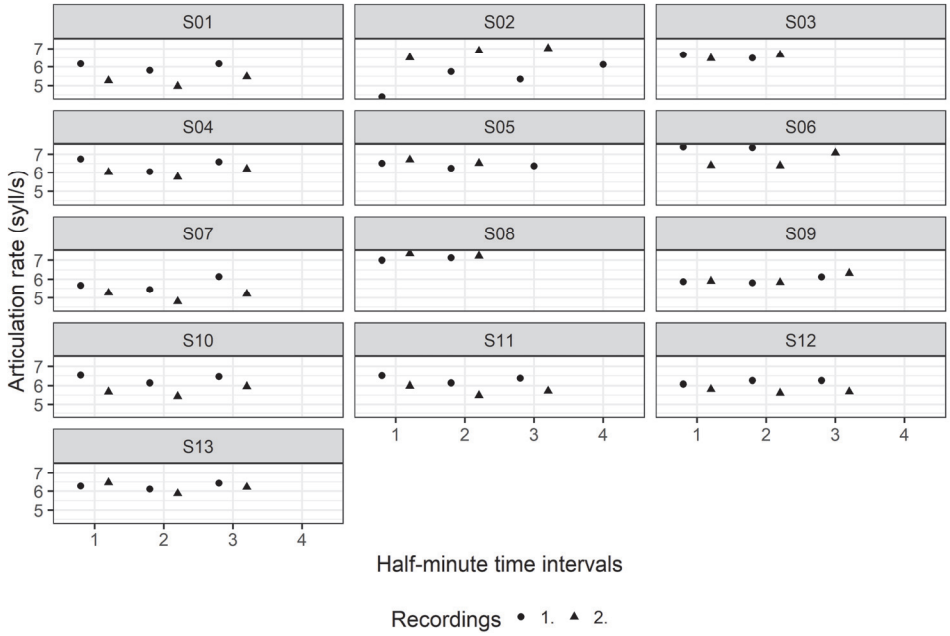

Figure 4. Average articulation tempo values calculated in half-minute time intervals in the reading texts (S01, S02, …refer to the speakers)

Slika 4. Prosječne vrijednosti tempa artikulacije u poluminutnim vremenskim intervalima tijekom čitanja (S01, S02… odnose se na govornike)

Processing the values of spontaneous speech is difficult because the length of the recordings differs significantly between the first and second recordings as seen from the extent and location of the circles and triangles. In addition, the values of spontaneous speech show greater variance in temporal patterns than in the readings (Figure 5). The articulation tempo values showed a significant overlap for all participants between data collection timepoints, with the exception of speaker S04 (for whom a clear deceleration was detected). However, no trends pointing towards differences between the half-minute sections were found, as opposed to 20% breakdown.

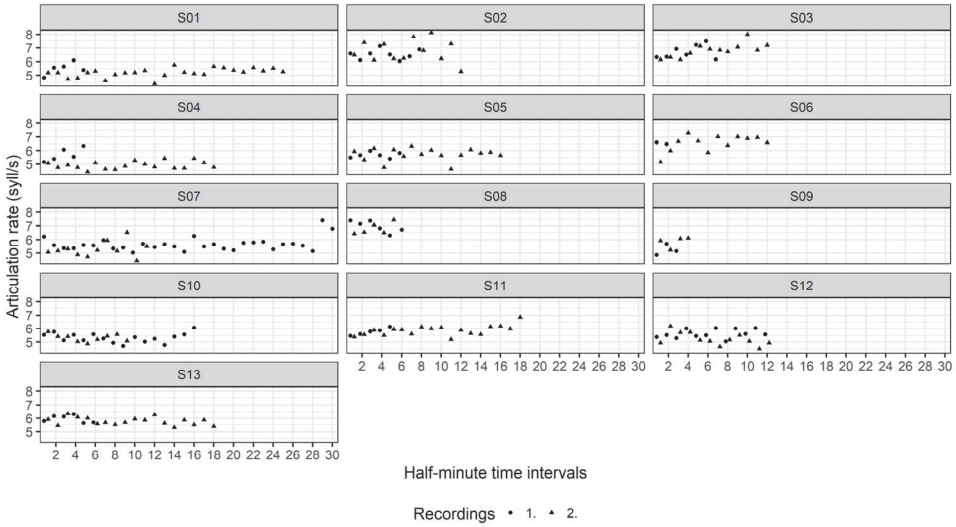

Figure 5. Articulation tempo values calculated in half-minute time intervals occurring in spontaneous speech by speaker

Slika 5. Vrijednosti tempa artikulacije u poluminutnim vremenskim intervalima tijekom spontanoga govora

3.2 Pauses

The temporal pattern of speech samples is determined by the frequency and duration of pauses in the conversations. The change in the utterances of individual speakers in terms of pause patterns 10 years between the first and last recordings for each type of speech was examined. The mean duration ratio of the pauses in spontaneous speech was 0.24 (SD: 0.06) for recording 1 and 0.23 (SD: 0.06) for recording 2. In the reading task, it was 0.18 (SD: 0.03) in the first recording and 0.19 (SD: 0.03) in the second recording. Based on the averages, the difference between the recordings seems to be negligible.

The speech types display a difference in the duration ratio of the pauses: spontaneous speech had higher average ratio of pauses. The breakdown by speaker (Figure 6) shows individual characteristics of pause realisations.

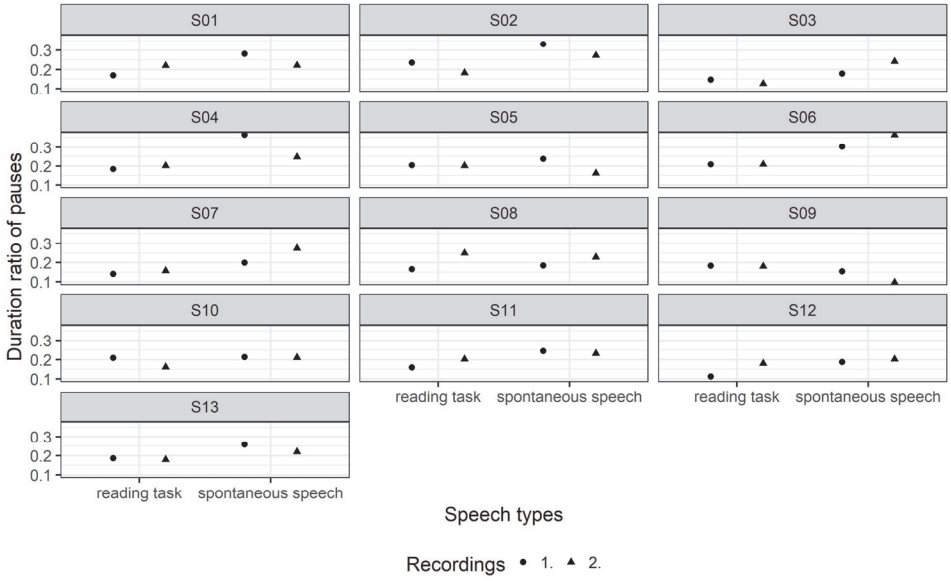

Figure 6. Duration ratio of pauses over total speaking time in the two speech types by recording and speaker

Slika 6. Odnos trajanja stanki i ukupnoga vremena govorenja u različitim vrstama govornoga uzorka i vremenskim točkama za pojedinoga govornika

The results showed remarkable individual differences, similar to the aspects studied previously. In the reading task, four speakers displayed a higher ratio of pauses whereas three speakers produced less pauses, 10 years after the first recording. No considerable differences were found in the values of the two recordings in almost half of the participants (i.e., six speakers). In spontaneous speech, no significant differences were found in the values for three speakers. The ratio of pauses decreased in six cases, whereas it increased in four cases in the recordings done 10 years later. The results, thus, justify very strong individual tendencies with few general characteristics. If we analyze the ratios regardless of the speech types, we see an obvious trend in only two speakers: the frequency of pauses clearly decreased in participant S02, while it increased in participant S08. The other participants showed either a mixed trend (e.g., S01: an increase in reading task and a decrease in spontaneous speech) or a change in one of the speech types. For most participants (eight speakers), both recordings clearly show a trend of spontaneous speech having a higher proportion of pause duration.

For other speakers, this trend is usually seen in one of the recordings, while the difference between the two speech types is negligible in the other recording. In one speaker (S09), an inverse trend can be observed in both recordings, with a higher ratio of pauses in reading.

Figure 7 shows the change of the pause rate within the recordings in a 20% breakdown.

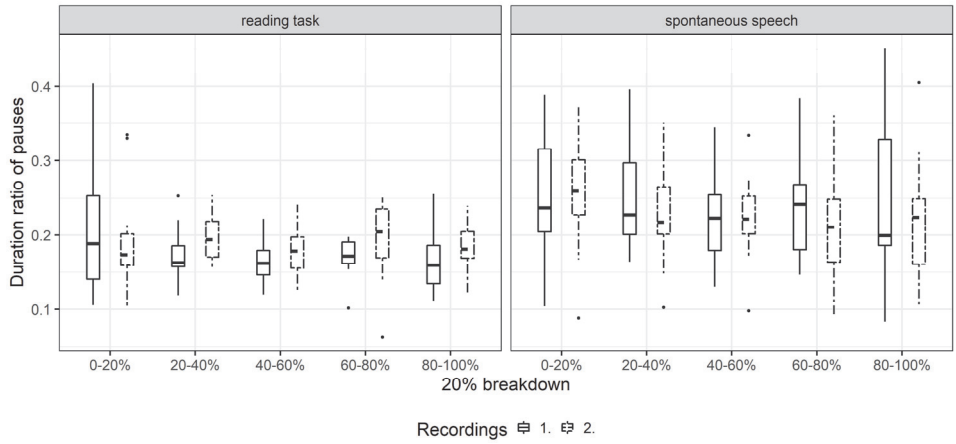

Figure 7. Duration ratio of pauses in 20% breakdown (median and interquartile range)

Slika 7. Odnos trajanja pauza u raščlambama od 20 % (medijan i interkvartilni raspon)

No trend can be detected between different speech type or recording time samples. The difference between the means of the 20% sections does not exceed 0.05 in the first and second (10 years later) recording either in spontaneous speech or in the reading text, so the difference is negligible (SD_{read} = 0.05; SD_{spsp} = 0.07).

The statistical analysis showed a significant difference between reading task and spontaneous speech [$F(1, 52) = 61.839$; $p < 0.001$], while there was no statistical difference between the 20% sections for the first and second recordings.

The frequency of pauses in the reading task is 0.3 1/s, on average, during the first recording (SD: 0.06 1/s), and 0.28 1/s during the second recording (SD: 0.05 1/s). In spontaneous speech, the average was 0.4 1/s (SD: 0.07 1/s) during the first recording and 0.39 1/s (SD: 0.06 1/s) during the second recording.

Regarding frequency, there is a difference between the types of speech, similar to the pause ratio: the pauses are more frequent in the spontaneous speech – a deviation of approximately 0.1 1/s is not negligible with a standard deviation of 0.06 1/s.

There are no significant differences between the first and the second recording, similarly to what we have seen for the pause ratio. If we compare the mean values, there are no significant changes, although the values change in the same direction (in both speech types, the average of the frequencies in the second recording is lower). These changes are only 0.01–0.02 1/s, so they are negligible with a standard deviation of 0.06 1/s.

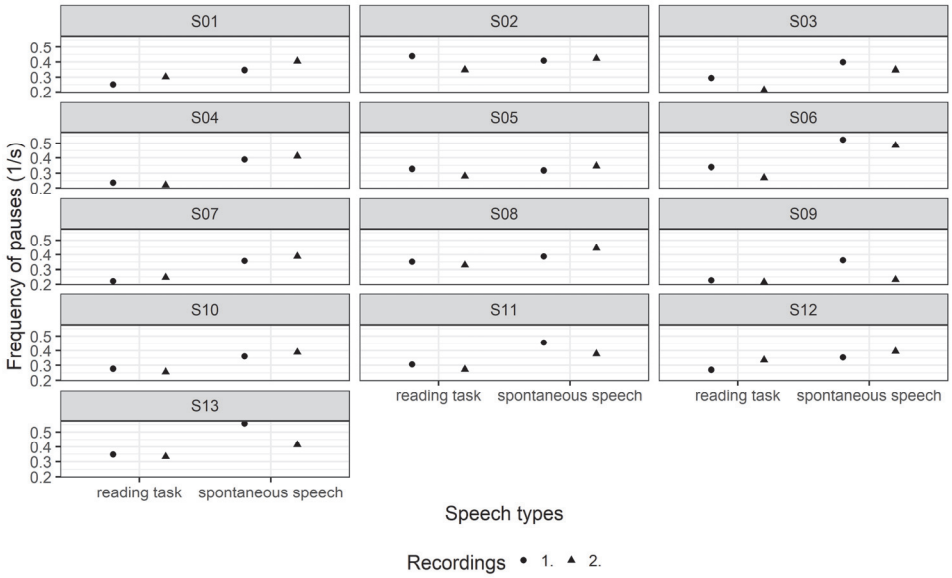

Figure 8. Frequency of pauses in the two speech types by recording and speaker
Slika 8. Učestalost pauza u različitim vrstama govornoga uzorka i vremenskim točkama za pojedinoga govornika

Figure 8 also shows some individual differences with a similar mixed trend. In two cases, the pauses became more frequent in reading during the second recording than in the first case, in four cases they became less frequent, however, in as many as seven cases, and there were no considerable differences between the two recordings. In the spontaneous speech the frequency of pauses increased in three cases, decreased in five cases, and in five cases no considerable change occurred between the two

recordings. For two speakers the frequency of pauses increased regardless of speech type, for other two speakers, it decreased, however, nine cases showed mixed or negligible changes during the period of 10 years. Thus, the change in the frequency of pauses is mixed for different timepoints, while for different speech types, a clear trend is showing: the pauses are more frequent in spontaneous speech than in the reading for the majority of speakers (10 speakers). The same tendency is seen in the results of the remaining three participants, but only in one of the recordings. In the other two recordings, the difference in the pause frequency between the two speech types does not exceed 0.03 1/s, so it can be considered negligible.

Figure 9 shows the frequency of pauses within recordings in a 20% breakdown. The results for spontaneous speech, showed no trend in the different parts of the sample. The differences between the averages do not exceed the value of 0.03 1/s, so they were negligible (SD_1 = 0.09 1/s, SD_2 = 0.08 1/s). However, when reading (as in the case of the articulation rate) the participants showed tendency of making less breaks in the middle of the text and more towards the end during the first recording. This tendency was also noticeable during the second recording, although the pauses were more frequent in the last section in both recordings. In the first recording, the difference between the lowest and the highest average value is 0.08 1/s, and in the second, it is 0.1 1/s (SD_1 = 0.08 1/s, SD_2 = 0.07 1/s).

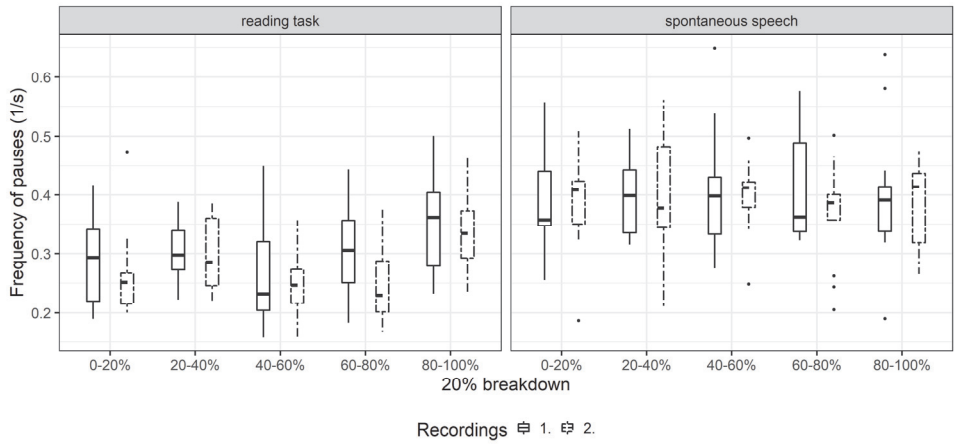

Figure 9. Frequency of pauses in 20% breakdown (median and interquartile range)
Slika 9. Učestalost pauza u raščlambama od 20 % (medijan i interkvartilni raspon)

Statistical analysis confirmed that the speech type ($F(1, 260) = 159.450$; $p < 0.001$), the time between the two recordings ($F(1, 260) = 4.312$; $p = 0.039$), and the 20% breakdowns ($F(4, 260) = 3.629$; $p = 0.007$) also have an effect on the frequency of pauses. According to Tukey post hoc, the last section (80–100%) of the reading text differed from all other sections. In this section, several pauses were reported per unit of time (0–20%: $p = 0.021$; 20–40%: $p = 0.024$; 40–60%: $p = 0.007$; 60–80%: $p = 0.027$).

The correlation between the change in the ratio of pauses and the frequency of pauses between speech types (Table 1) was examined.

Table 1. Direction of change in pause ratio and pause frequency in reading and spontaneous speech ('↑' sign means increasing tendency (min. 10%), '↓' sign means decreasing tendency (min. 10%). 'x' sign shows less than 10% alteration between the two recordings.)

Tablica 1. Promjene u odnosu i učestalosti pauza u čitanju i spontanome govoru (↑ označava rast mjerene varijable (za min. 10 %), a ↓ pad mjerene varijable (za min. 10 %). Znak x označava da nema zamjetnih promjena u uzorcima, tj. manje su od 10 %.)

Speakers / Govornici	Ratio of pauses / Odnos pauza		Frequency of pauses / Učestalost pauza	
	Reading task / Zadatak čitanja	Spontaneous speech / Spontani govor	Reading task / Zadatak čitanja	Spontaneous speech / Spontani govor
S01	↑	↓	↑	↑
S02	↓	↓	↓	x
S03	x	↑	↓	↓
S04	x	↓	x	x
S05	x	↓	↓	x
S06	x	↑	↓	↓
S07	x	↑	x	x
S08	↑	↑	↓	+
S09	x	↓	x	↓
S10	↓	x	x	x
S11	↑	x	x	↓
S12	↑	x	↑	↑
S13	x	↓	x	↓

If we compare Figure 6 and 8, similar differences can be noticed per speaker between spontaneous speech and reading in terms of pause rates and pause frequencies – the results show which type of speech has more pauses in proportion to time, but, of course, there are individual differences in pausing. There is no uniformity regarding the change of the examined parameters of the speakers in the reading task. For four speakers, neither the ratio nor the frequency of pauses changed over a period of 10 years. For three participants, even though the ratio of pauses did not change, their frequency decreased, so longer pauses were produced during the second recording. For two speakers, both the frequency and the proportion also increased, so pauses were implemented more often, and they were similar in length: the tendencies showed great variability.

In spontaneous speech, although the pause ratio for three speakers decreased, their frequency did not change, so shorter pauses were implemented in the second recording than in the first. Individual differences dominated in all other cases.

4. DISCUSSION

In recent years, attention given to speech-related disciplines has shifted to analysis of extra- and paralinguistic information occurring in speech. This is mainly since they significantly influence the evaluation of linguistic information and provide an additional incentive to improve technological possibilities. Moreover, public interest in aging is growing which, in turn, may encourage increase of similar studies.

Responding to this need, the present research examined certain timing characteristics to indicate changes in speech over a period of 10 years, using two different types of speech: spontaneous speech in the form of interviews and reading tasks. The sample consisted of 13 young, adult males.

Our results showed that (1) a slower articulation rate was observed in spontaneous speech and (2) the two types of speech differed more in variance: a more varied articulation rate was observed in spontaneous speech than in reading. Therefore, the first part of Hypothesis 1 was confirmed by the data. A possible explanation for this may be that spontaneous speech and reading present different cognitive loads due to different degrees of preparation and, therefore, different design and execution processes (cf. summary Howell & Kadi-Hanifi, 1991; Silverman, Blaauw, Spitz, & Pitrelli, 1992).

It is known that planning processes of different difficulty levels are seen, among other things, in the temporal characteristics of speech (Bóna, 2013; Fletcher, 2010; Markó, 2014). During the production of spontaneous speech, speech planning and execution take place simultaneously. The speaker plans the content and grammatical form of the message at the moment of speaking while there is no need to operate higher planning processes in reading because the grammatical form is given (Krepsz, 2016; Váradi, 2010). This is critical in terms of the magnitude of the articulation rate and its variability. In addition to preparation time, the length of the texts also influences the speed: the reading time was shorter, it took 1–1.5 minutes on average while the spontaneous speech was sometimes as long as 15 minutes, depending on the individual habitus, physical and mental condition.

The 10 years between the recordings did not have a significant effect on the articulation rate of the utterances: no typical patterns for the speakers could be noticed in different timepoints. The same number of speakers showed an increasing and decreasing trend and no changes were noted in three cases. Thus, the first part of Hypothesis 2 was not confirmed by the data. The results partially confirm that fewer notable changes are expected in young, healthy adults following a longer period of time, even a decade. The results are also in agreement with those of Gersternberg et al. (2011), who reported non-unified trends, but significant individual characteristics in older speakers (regardless of the languages spoken, see the introduction for more details). The process of aging (regardless of the studied stage of life) strongly depends on individual characteristics, inherited and environmental factors.

Changes in the articulation rate in spontaneous speech were found to be "stable" in both recordings, so the point in which the tempo values were measured in a particular sample was not decisive. In contrast, a small but typically parabolic curve was seen in the reading task, i.e., higher tempo values at the beginning and at the end of the reading and lower ones in the middle. These results are not in accordance with the cone-shaped arrangement observed in the speech of Queen Beatrix – the average tempo value in the three-minute sections gradually increased and then decreased from approximately half of the utterances. A possible reason for this is that the Queen's speech may have been practiced and had a longer preparation time, while other speakers may have had a shorter preparation time (they may have read the text in a few seconds). Alternatively, our participants were not professional speakers.

The results for the two speech types have shown trends supported by normalized values. Thus, the first part of Hypothesis 3 was not supported by the data. This is

probably because in the reading task, the speaker can see the length of the text, and consequently, has control of and plan their rate, while in spontaneous speech, questions or switching to new topics allow the speaker to adjust the speech rate. In addition, all this may result from the structure of the text, although the trend was not the same for all speakers.

The breakdown in smaller units of time showed significant individual differences in the reading task, however, it showed patterns in temporal features when compared to the differences in the 10-year period. The same trend was not detectable in spontaneous speech, which can be explained by the planning and implementation of reasons listed earlier. The pausing patterns showed a mixed picture. There was a difference between the two types of speech in terms of both pause ratio and frequency of pauses, thus the second part of Hypothesis 1 was also confirmed by the data.

The rate of change in the 10-year period was very small for most speakers. In spontaneous speech, the differences between recordings were negligible. In the reading task, the trend-level difference in the pausing ratio was commented previously. For the second series of recordings, the rate of pausing was higher than in the first one. Regarding the frequency of pauses, we found a significant difference between the two recordings: in the reading task, pauses were more frequent in the first recording than in the second. The second half of Hypothesis 2 was, thus, partially confirmed (in terms of frequency of pauses). The recordings of spontaneous speech showed no noticeable trend in the changes over time, either in terms of the ratio of pauses or the frequency of pauses. In the reading task, there was no typical trend noticed in the ratio of pauses between the parts. In terms of frequency, the last 20% of the recordings differ significantly from the values in the other parts: in the last section, the frequency of pauses was higher. Thus, the second part of Hypothesis 3 was also partially confirmed (in terms of frequency of pauses). The comparison of temporal features shows that a higher articulation rate, a smaller pause ratio, fewer pauses and, thus, longer speech segments characterized the text participants had read compared to spontaneous speech. In the recordings of the reading task, the correlation between the characteristics in the different parts, suggests that the articulation rate is faster at the beginning and at the end of the readings with pauses being more frequent at the end, resulting in longer speech segments. These temporal features may result from the structure and segmentability of the reading task. In addition, several other factors may affect the timing characteristics of the utterances (cf. Krepsz, 2016).

A better understanding of the aging process within and between the speakers is highly relevant to applied linguistics, human-machine communication and speech recognition. It is closely related to the topic of elderspeak, which has recently been receiving an increasing amount of attention, as well as to the development of tools designed to help and improve the quality of life for older people. The results of the study show well that not only the emerging trends, but also individual differences should be taken into account when describing changes along the age. Although the analysed samples were produced by young male speakers, the study should be replicated for both female speakers and informants of the older age group. In this way, we can collect more information about the characteristics of the aging process.

ACKNOWLEDGMENTS

The research was supported by the project NKFI-FK 128814 Research Scholarship.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csáki (Eds.), *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267–281). Budapest: Akadémiai Kiadó, Republished in S. Kotz, & N. L. Johnson (Eds.) (1992). *Breakthroughs in Statistics I.* (pp. 610–624). Springer-Verlag.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Version 6.0.19) [Computer program]. Retrieved August, 12, 2020, from https://www.praat.org/

Bóna, J. (2013). A beszédszünetek fonetikai sajátosságai a beszédtípus függvényében [Phonetic characteristics of speech silences as a function of speech type]. *Beszédkutatás*, *21*(1), 60–75.

Bóna, J. (2014). Temporal characteristics of speech: The effect of age and speech style. *Journal of the Acoustical Society of America*, *136*(2), EL116–EL121.

Brown, W., Morris, R. J., & Michel, J. F. (1989). Vocal jitter in young adult and aged female voices. *Journal of Voice*, *3*(2), 113–119.

Brückl, M. (2007). Women's vocal aging: A longitudinal approach. In R. Bock, F. Bonin, N. Campbell, & R. Poppe (Eds.), *Interspeech (2007 – 8th Annual*

Conference of the International Speech Communication Association (pp. 1170–1173). International Speech Communication Association.

Brückl, M., & **Sendlmeier, W.** (2003). Aging female voices: An acoustic and perceptive analysis. *Proceedings of VOQUAL'03* (pp. 163–168). Geneva.

Busse, E. W. (2002). General theories of aging. In J. R. M. Copeland, M. T. Abou-Saleh, & D. G. Blazer (Eds.), *Principles and Practice of Geriatric Psychiatry* (2nd ed.). Oxford: Willey-Blackwell.

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, *20*(3), 245–257.

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 521–602). Oxford: Willey-Blackwell.

Gerstenberg, A., **Fuchs, S., Kairet, J. M., Frankenberg, C., & Schröder, J.** (2011). A cross-linguistic, longitudinal case study of pauses and interpausal units in spontaneous speech corpora of older speakers of German and French. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proc. 9th International Conference on Speech Prosody* (pp. 211–215). Poznan: Adam Mickiewicz University.

Gráczi, T. E., Huszár, A., Krepsz, V., Száraz, B., Damásdi, N., & Markó, A. (2020). Longitudinális korpusz magyar felnőtt adatközlőkről [Longitudinal corpus of Hungarian adult speakers]. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *XVI. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 103–114). Szegedi Tudományegyetem, Informatikai Intézet, Szeged.

Hoit, J. D., Watson, P. J., Hixon, K. E., McMahon, P., & Johnson, C. L. (1994). Age and velopharyngeal function during speech production. *Journal of Speech, and Hearing Research*, *37*(2), 295–302.

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*(2), 163–169.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256.

Kohári, A. (2018). *Időzítési mintázatok a magyar beszédben* [*Timing patterns in Hungarian speech*]. ELTE Eötvös Kiadó, Budapest.

Krepsz, V. (2016). Fonetikai hasonlóságok és különbözőségek a beszédtípusokban [Phonetic similarities and differences in speech types]. In J. Bóna (Ed.), *Fonetikai*

olvasókönyv [*Phonetic reading book*] (pp. 175–188). ELTE Fonetikai Tanszék, Budapest.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(1), 1–26.

Lalley, P. M. (2013). The aging respiratory system – Pulmonary structure, function and neural control. *Respiratory Physiology & Neurobiology*, *187*(3), 199–210.

Linville, S. E. (1996). The sound of senescence. *Journal of Voice*, *10*(2), 190–200.

Linville, S. E. (2000). The aging voice. In R. D. Kent, & M. J. Ball (Eds.), *Voice Quality Measurement* (pp. 265–270). San Diego: Singular Thomson Learning.

Linville, S. E. (2001). *Vocal aging*. San Diego: Singular Thomson Learning.

Markó, A. (2014). A beszéd temporális szerkezete a beszédmód és a beszédhelyzet függvényében [Temporal structure of speech as a function of speech type and speech situation]. In S. Bátyi, J. Navracsics, & M. Vígh-Szabó (Eds.), *Nyelvelsajátítási, nyelvtanulási és beszédkutatások* [*Language acquisition, language learning and speech research*] (pp. 33–45). Pszicholingvisztikai tanulmányok IV. Gondolat Kiadó – Pannon Egyetem MFTK, Budapest – Veszprém.

Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., & Beke, A. (2014). Development of a large spontaneous speech database of agglutinative Hungarian language. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Proceedings of TSD* (pp. 424–431). Heidelberg: Springer.

Oyer, E., & Deal, L. (1985). Temporal aspects of speech and the aging process. *Folia Phoniatrica (Basel)*, *37*(3–4), 109–112.

Ptacek, P., & Sander, E. S. (1966). Age recognition from voice. *Journal of Speech and Hearing Research*, *9*(2), 273–277.

Quené, H. (2013). Longitudinal trends in speech tempo: The case of Queen Beatrix. *The Journal of the Acoustical Society of America*, *133*(6), 452–457.

R Core Team. (2018). R: A language and environment of statistical computing. Vienna: R Foundation for Computing. Retrieved May, 10, 2019, from https://www.r-project.org/

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders*, *16*(3), 217–226.

Ringel, R. L., & Chodzko-Zajko, W. J. (1987). Vocal indices of biological age. *Journal of Voice, 1*(1), 31–37.

Shum, S. (2008). *The Language of an Aging Queen. Linguistics 110 Final Paper, December 2008*. Retrieved from http://people.csail.mit.edu/sshum/ucb_papers/aging_queen.pdf

Silverman, K. E. A., Blaauw, E., Spitz, J., & Pitrelli, J. (1992). Towards using prosody in speech recognition/understanding systems: Differences between read and spontaneous speech. *Proceedings DARPA Speech and Natural Language Workshop* (pp. 435–440).

Smith, B. L., Wasowicz, J., & Preston, J. (1987). Temporal characteristics of the speech of normal elderly adults. *Journal of Speech and Hearing Research*, *30*(4), 522–529.

Tivadar, H. (2017). Speech rate in phonetic-phonological analysis of public speech (using the example of political and media speech). *Journal of Linguistics*, *68*(1), 37–56.

Váradi, V. (2010). A felolvasás és a spontán beszéd temporális sajátosságainak összehasonlítása [Comparison of temporal features of reading and spontaneous speech]. *Beszédkutatás*, *20*, 100–109.

Walker, V. G. (1988). Durational characteristics of young adults during speaking and reading tasks. *Folia Phoniatrica et Logopaedica*, *40*(1), 12–20.

Winkler, R. (2007). Influences of pitch and speech rate on the perception of age from voice. In J. Trouvain, & W. J. Barry (Eds.), *16th International Congress of Phonetic Sciences* (pp. 1849–1852). Saarbrücken, Germany.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. *Interspeech – ICSLP, 9th International Conference on Spoken Language Processing* (pp. 541–544). Pittsburgh (PA), USA.

Zellner-Keller, B. (2006). Ageing and speech prosody. In R. Hoffmann, & H. Mixdorff (Eds.), *Speech Prosody* (pp. 696–701). Retrieved from http://sprosig.org/sp2006/contents/papers/OS5-1_0001.pdf

Anna Huszár, Valéria Krepsz
huszar.anna@nytud.hu, krepsz.valeria@nytud.hu
Istraživački institut za lingvistiku, Budimpešta
Mađarska

Razvoj varijabilnosti pauza i tempa artikulacije kod mađarskih govornika tijekom desetogodišnjega razdoblja

Sažetak

U posljednje vrijeme istraživanja u govornim znanostima sve više uključuju izvanlingvističke i paralingvističke karakteristike govora. Štoviše, raste zanimanje i za starenje općenito, što potiče ovakva istraživanja. Prethodna su istraživanja većinom uključivala starije govornike te nam do danas nije dostupno puno informacija o promjenama u glasu kod mlađih, zdravih govornika tijekom vremena. U ovome se radu istražuju promjene vremenskih karakteristika govora u spontanome govoru i čitanju u razdoblju od deset godina. Analizirani su govorni uzorci 13 zdravih, odraslih govornika snimljenih u istim uvjetima u razmaku od deset godina te razlike u vremenskim karakteristikama unutar i između snimki. Rezultati pokazuju razlike u ostvarenju stanki ne samo u različitim govornim uzorcima već i razlike unutar i između snimki. Dodatno, podatci pokazuju da se u spontanome govoru javlja sporiji tempo artikulacije i da se dvije vrste govornoga uzorka razlikuju s obzirom na varijabilnost: veća se varijabilnost u tempu artikulacije javlja u spontanome govoru nego u čitanju. Individualne razlike dominiraju u rezultatima dobivenima za tempo artikulacije.

Ključne riječi: vremenske karakteristike, longitudinalno istraživanje, tempo artikulacije, dob, vrsta govora