ÁKOS GOCSÁL[1,2] – SVETLANA MARUSENKO [3] –
MÓNIKA GALAMBOSNÉ TISZBERGER[4]

University of Pécs, Faculty of Music and Visual Arts, Institute of Music[1]
Hungarian Academy of Sciences, Institute of Linguistics, Dept. of Phonetics[2]
gocsal.akos@pte.hu
University of Pécs, Faculty of Sciences[3]
University of Pécs, Faculty of Business and Economics, Institute of Economics and Econometrics[4]

Ákos Gocsál – Svetlana Marusenko – Mónika Galambosné Tiszberger: Listener differences in speaker age estimation
Alkalmazott Nyelvtudomány, XIX. évfolyam, 2019/1. szám
doi:http://dx.doi.org/10.18460/ANY.2019.1.002

Listener differences in speaker age estimation

A beszéd akusztikai minőségét befolyásoló nyelven kívüli tényezők közül a beszélő életkora az, amelyiknek az egyik legjelentősebb a hatása. Számos korábbi kutatás igazolta, hogy a hallgató képes bizonyos pontossággal következtetni a beszélő életkorára a hangja alapján. A hallgatók közötti különbségekkel kapcsolatban is állnak rendelkezésünkre adatok. A korábbi kutatások elsősorban a hallgató neme, életkora, illetve a hallgató és a beszélő anyanyelve vagy akcentusa alapján vizsgálták az életkorbecslések pontosságát. Nincs azonban ismeretünk arról, hogy a hallgatók tipizálhatók-e az általuk adott életkorbecslések pontossága alapján, azaz a becslési adatok kirajzolnak-e olyan, jól elkülönülő mintázatokat, amelyek eltérő életkorbecslési mechanizmusokat tükröznek. A jelen kutatásban 85 hallgató 24 férfi beszélő életkorát becsülte meg hangja alapján. Az adatokat a *k*-közép klaszteranalízis módszerével dolgoztuk fel, 4 illetve 3 klaszteres megoldással. A klaszteranalízis a várakozásoknak megfelelően az életkorbecslés pontossága alapján jól elkülöníthető csoportokat határozott meg, amelyek felvetik az eltérő stratégiák létezésének lehetőségét. Mivel itt csak a jelenséget, azaz az eltérő mintázatokat mutattuk ki, további kutatás feladata az észlelési mechanizmusok közötti különbségek lehetséges okainak feltérképezése.

Introduction

The purpose of voice-based age estimation experiments is to reveal the nature of judgments that listeners make when they hear an unseen speaker. As Pettorino and Giannini (2011) summarized based on previous literature, the major changes in voice that are in relation with age as follows:

- lowering of breathing functions
- muscle relaxation, hardening of vocal folds
- progressive tonal lowering
- lowering of speech rate
- increase of jitter and shimmer
- lowering of formant frequencies
- longer vowels and stop consonants
- increased standard deviation of f_0

Certain parameters or combinations of parameters in the acoustic structure of speech, such as tempo (Stölten–Engstrand, 2003; Skoog Waller et al., 2015; Gocsál, 2017), duration (Schötz, 2004), f_0 and F1 (Reubold et al., 2010), or spectral information (Schötz, 2004) are used as markers of age by listeners who thus infer the speaker's age.

Although the earliest experiments, such as the one carried out by Allport and Cantril (1934) already found that listeners' age judgments fairly well matched the chronological age of the speakers, more reliable results have been produced only since the 1960s. The most commonly used parameter to describe association between chronological and perceived age of speakers is Pearson's correlation coefficient. A number of studies has demonstrated significant correlation between the chronological age and mean age estimates (Table 1.)

Table 1. Correlation coefficients between calendar and perceived age from previous literature, $p<0.05$ in all cases

Paper	***r***	**remarks**
Braun–Cerrato (1999)	.300–.790	speakers: German/Italian males listeners: German/Italian college students, no knowledge of the other language
Stölten–Engstrand (2003)	.96 .86	unmanipulated samples speech rate and f_0 manipulations speakers: young (20-30) and older (50-70) males and females from SWEDIA 2000 dialect database listeners: Stockholm area male and female students (age 20-29)
Schötz (2004)	.944 .825	Swedish male and female speakers listeners: male and female university students (age 18-36)
Bóna (2013:126, 128)	.907 .809	Hungarian male and female speakers listeners: male and female university students (age 18-25)
Huckwale–Webb (2015)	.759	native English speakers and listeners (age 20-69)
Gnevsheva–Bürkle (2019)	.37–.64	English/Japanese speakers and English (age 19-57) and Japanese (age 19-70) listeners

Correlation, however, does not imply that listeners' estimations are correct. Many studies have demonstrated that younger speakers, typically below the age of approximately 35 years are believed to be older than their calendar age, while those older than 35 are usually perceived younger (Pettorino–Giannini, 2011; Kasuya, 2006). Other researchers found a somewhat higher age (over 40 years) that separated over- and underestimation (Huckwale–Webb, 2015), but in that case younger and older listeners were also involved. Overestimation of young speakers' age and underestimation of that of the older ones have been demonstrated by many other researchers too (Schipp et al., 1992; Hughes–Rhodes 2010; Moyse et al., 2014; Sandman et al., 2014; Krepsz–Gósy, 2016; Hunter–Ferguson, 2017).

Although the primary focus of research is usually speaker variability in perceived and chronological age, listener variability with regard to the accuracy of speaker age estimation has also been tested by many researchers and some listener attributes also seem to be of importance. Such listener attributes include:

- gender
- age
- familiarity with the speaker's language

As to listener gender, no significant differences between male and female listeners' age estimations were found in general (Hartman, 1979; Eriksson et al. 2004; Pettorino–Giannini, 2011; Moyse, 2014; Huckwale–Webb, 2015), but some conflicting results have been reported. Hartman (1979) found that female listeners were better at estimating male speakers' age if the speaker was over 50 years of age. In another research, male listeners performed non-significantly better than females (Braun & Cerrato, 1999).

Listener age may be of importance as well, but again, results are conflicting at some points. In a review article, Moyse (2014) stated that younger listeners are more accurate than older participants irrespective of the age of stimuli. This statement is confirmed by a number of studies. In an experiment by Huntley et al. (1987) four listener groups were used: adolescents, young adults, middle-aged, and older participants. While the age of the older speakers was judged very similarly by the four groups, significant differences were found in with the 20- and 30-year olds: the adolescents and the older individuals significantly overestimated their ages, while the young and middle-aged participants were more accurate. Moyse et al. (2014) also found that the age of older speakers was underestimated both by the younger and older listeners to the same extent. The age of the younger speakers was relatively well estimated by the younger participants, while the older participants made larger errors, i.e. overestimated them. Huckwale and Webb (2015) found that listeners in age bands 40-49 and 60-69 gave significantly worse age predictions than those in the 20-29 band.

In contrast, different results have been provided by Eppley and Mueller (2001), who played voice samples of elderly speakers in two listener groups. The group of young listeners included subjects between 18-22 years of age, while those in the old listeners' group were between 61 and 84. The older listeners were somewhat more accurate in estimating the speakers' age than the young listeners, but the difference was not statistically significant. In another research by Eriksson et al. (2004), two groups of listeners (mean ages: 31.1 and 20.7 years) were employed. The participants were consistent in ranking the younger speakers by age, however, the younger participants failed to order correctly the two oldest speakers by age which suggests that younger listeners were better at estimating the ages of those speakers who were closer to them in age. Hughes and Rhodes (2010) found that listeners, divided into four age groups, differed in estimating the age of the oldest speakers, i.e. those belonging to the oldest listeners' group were significantly better at estimating the oldest speakers' age than the others.

Listener or speaker accent is another factor that may influence the accuracy of age estimates. German and Italian listeners were played German and Italian speakers' voice samples. Although the listener groups performed almost equally with respect to the Italian stimuli, the Italian listeners performed slightly worse on the German stimuli than did the Germans. The difference, however, did not reach statistical significance (Braun–Cerrato, 1999). Significant differences were, however found between the age estimations of German, Finnish and Swedish listeners when they heard native English speakers born in different English-speaking countries (Sullivan et al., 2000), but no significant difference was found between native and non-native speakers of English in general.

Jiao et al. (2019) proved the significant the main effect of linguistic familiarity, i.e. native Korean and Mandarin listeners estimated the ages of speakers of their own native languages significantly more accurately than native Arabic speakers' ages and vice versa. In their experiment, all speakers and listeners were learners of English and the speech stimuli were recorded in English. In another experiment Gnevsheva and Bürkle (2019) also proved the effect of L1. Native English listeners perceived English speaking Japanese speakers' ages as younger than their native English counterparts, while English- and Japanese- accented speech did not affect Japanese listeners' age estimation.

A mention should be made about the methodology of speaker age estimation. In general, two main approaches are applied. In one group of the experiments, the researchers ask the listeners to provide the accurate calendar age of the speakers as they perceive (Braun–Cerrato, 1999; Eppley–Mueller, 2001; Schötz, 2004; Pettorino–Giannini, 2011; Huckwale–Webb, 2015; Jiao et al., 2019) and correlation coefficients or linear models are used to establish conclusions. In other experiments, the researchers define age groups and the listeners' task is to find which age groups the speakers belong to. The researchers then usually calculate the percentage of correct group assignments (Hummert et al, 1999; Amir et al., 2012; Tatár, 2013) The mixture of the two approaches is also used in several papers (Hughes–Rhodes, 2010; Pettorino–Giannini, 2011; Huckwale–Webb, 2015; Gnevsheva–Bürkle, 2019) These authors used estimated calendar ages, however, when they processed the data, they created speaker age ranges and determined the percentage of correct answers or mean errors of prediction with regard to the age ranges.

Both the correlation based and the age range based approaches have their own benefits. While the correlation based approach can provide general tendencies over a wider range of speaker age, the age range based method can demonstrate possible deviations from general tendencies that may occur in different speaker age groups. For example, Hughes and Rhodes (2010) found that mean difference of estimated ages from actual ages was smaller with male speakers over 55 than with speakers between 46-55, which means that it was not the speakers over 55 whose age was actually most underestimated, but the middle-aged speakers. It is therefore important to highlight that although underestimation and overestimation, as a function of speaker age, are in general demonstrated by many researchers, conflicting results make further investigations necessary.

While listener age, gender and accent have been researched in the context of age estimation accuracy, little is known, however, about possible types of listeners, irrespective of these differences. Can we say that certain listeners are better than the others? Are there listeners who systematically overestimate or underestimate speaker age, while others are more accurate? The main objective of this paper is to find answers to these questions.

In our experiment, we also wished to test if listeners' musical experience influences age estimations. Previous results suggest that musicians have enhanced auditory perceptual skills in the perception of a variety of acoustic skills, compared with non-musicians. More accurate identification of changes of pure tone frequencies (Liang et al., 2016), enhanced performance on frequency discrimination (Micheyl et al., 2006), enhanced sensitivity to discriminating and identifying subtle temporal and timbre differences in speech (Sadakata–Sekiyama, 2011) are just a few examples where musicianship proved to be an advantage. It seems therefore reasonable to examine if musicianship is an advantage in speaker age estimation as well, resulting more accurate estimations.

Research objectives and hypotheses

In the present paper, our purpose is to demonstrate listener variability in speaker age estimations. For the calculations presented here, we use the same dataset as in Gocsál (2018), however, in the present study we raise different questions and apply different statistical methods. The previous experiment focussed on three areas: (1) we analysed correlation coefficients between mean estimated ages and calendar ages for musicians and non-musicians, and although some differences were found, musicianship and listener gender had no significant effect. (2) We also analysed musician and non-musician listeners' age estimations in three separate age groups of speakers but no statistically significant differences were found, and (3) no statistically significant differences were found either between male and female listeners, however, there was a non-significant tendency that male musicians were more accurate in the age estimation of younger speakers, while female musicians were slightly better at estimating the age of the older speakers.

In our present paper, we focus on the issue of individual differences in speaker age estimation that may be revealed in the form of different patterns. In this paper we use the term 'pattern' to refer to response types of listeners that systematically differ in age estimation. One pattern may be that of the "overestimators", who in general believe that speakers are older than their chronological age. Similarly, there may exist "underestimators", whose age estimation pattern follow an opposite tendency. We expect that our research will either reveal these patterns or other ones that can be defined in different way because of the different behaviour of the listeners.

Most of the previous results suggest that there is no significant difference between male and female performance in age estimation (Moyse, 2014) and we also expect this result to be confirmed also by the *k*-means cluster method. However, because of the data processing methods applied here are very different from those used in previous literature, the possibility of obtaining different outcomes here cannot be ruled out. Third, in a similar way, although musicianship did not prove to be an advantage in our previous calculations (Gocsál, 2018), we expect that this other methodology used for data processing may reveal some areas where musicians' performance is better.

We have thus developed the following hypotheses:

H1 We hypothesize that well definable patterns in speaker age estimation exist, i.e. listeners can be grouped according to the accuracy of age estimates.
H2 We hypothesize that male and female listeners do not demonstrate different patterns of age estimation.
H3 We hypothesize that musicians outperform non-musicians in age estimation, i.e. their estimations are more accurate than those of the non-musicians.

In this research, grouping of listeners is a statistic evidence-based categorization.

Materials, participants, procedure

Before the experiment, voice recordings from 24 non-smoker male speakers (aged 20-72) were selected from the BEA database (Gósy et al., 2012). When ordered by age, age difference between two adjacent speakers was not more than 4 years. We used 20-30 s long spontaneous speech samples from the "interview" or "argument" parts of the BEA recordings, in which the speakers were talking about an everyday topic (e.g. job, or school experiences, hobby etc.) in a neutral emotional state. No verbal information was included that may have given a hint about the age of the speaker. The interviewer's voice was not included in the recordings and no previous conditions were set to other parameters, such as those of the pauses, although care was taken to choose samples so that they do not include long pauses.

Listeners were normal hearing university students (n = 85, age range: 19-37, median: 22 years) of the Faculty of Music and Visual Arts, and also the Faculty of Humanities of the University of Pécs. 42 students (14 male, 28 female) have studied music for at least 8 years and are students of musical performance. 43 students (14 male, 29 female) have not received any kind of musical education other than the Singing and music school subject and have no experience as players of any musical instrument.

The students listened to the voice samples in groups of 5-10, in a silent seminar room of the Zsolnay campus of the University of Pécs, Hungary. For carrying out the listening task, a built-in multimedia system was used with professional loudspeakers. Before the listening task, the experimenter played three speech samples to the group to familiarize them with the task and also to make sure that all participants can properly hear the recordings. For collecting estimation data, printed table-like forms were used. Each participant was asked to write down the estimated age of the speaker in years. Each speech sample was played only once, in a randomized order, and the experimenter played the next sample only when all participants have written down their answers.

For data processing we used SPSS 25 and Microsoft Excel 2016 software. First, we used Pearson's correlation coefficient to compare results with previous findings. We then calculated the estimation error, i.e. the difference (D) between the perceived and chronological age of the speakers (Amilon et al., 2007) with regard to each listener and also standardized these difference values (ZscoreD). We performed *k*-means cluster analyses on the standardized data to organise listeners into groups who provided similar age estimates. We tested a 4-cluster and a 3-cluster solution. We used the 4-cluster solution because we expected the existence of groups of "underestimator", "accurate estimator", "overestimator" listeners and a fourth group whose members may not fit into any of these three. We tested the 3-cluster solution as well, expecting only the existence of the first three groups. We also applied Chi-square tests of association to establish if the proportion of male/female and musician/non-musician listeners differs across the clusters.

Results

In a previous study we already reported Pearson's correlation coefficients between the mean age estimates and chronological age found on these data (Gocsál, 2018) with regard to all listeners as one group and some sub-groups as well. Here we present further data. Figure 1 shows the scatterplot of calendar age and mean age estimates with all listeners included.

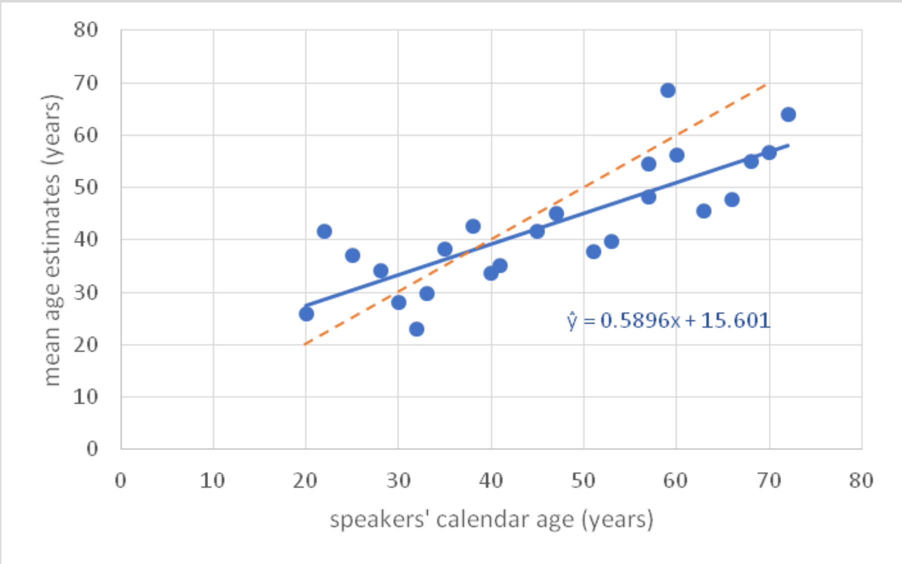


Fig. 1. Scattered plot of 24 voices' calendar age and mean of age estimates given by the 85 students

The brown coloured dashed line is the *y* = *x* line. If the listeners had given accurate estimations, the dots would be on this line. The solid line is the regression line that fits on the data points. With the exception of one speaker, all dots over the calendar age of 40 are below the solid line. This reflects the underestimation of older speakers' age, while a slight overestimation of younger speakers' age can also be observed.

We have calculated the correlation coefficients between the calendar age and mean age estimates for the whole group and several sub-groups of listeners (Table 2), asterisk indicating previously published data (Gocsál, 2018).

Table 2. Correlation coefficients between calendar and perceived age, *p*<0.05 in all cases

listener groups	*r*
all listeners* (n=85)	.806
males (n=28)	.825
females (n=57)	.795
musicians (n=42)	.808
non-musicians (n=43)	.800
male musicians* (n=14)	.803
female musicians* (n=28)	.806
male non-musicians* (n=14)	.839
female non-musicians* (n=29)	.777

These data suggest that the highest correlation coefficient was achieved by the male non-musicians, and the weakest, but still significant coefficient was that of the female non-musicians. In all cases, correlation coefficients are above .7 therefore the association is strong and obviously positive.

Next, *k*-means cluster analyses were administered with D values for the 24 voice samples as separate variables and individual listeners as cases in order to identify different patterns of age estimations whose existence is hypothesized. We standardized the D values (computed Z-scores) and tested four and three cluster solutions to analyse if listeners belonging to these clusters differ in age estimation accuracy.

Table 3 contains the number of listeners in each cluster and Table I (Appendix) shows the cluster centres for the four-cluster solution. Since the group sizes are similar, there are no individual listeners that behave very differently from the rest of the listeners.

Table 3. Number of listeners in each cluster (4-cluster solution)

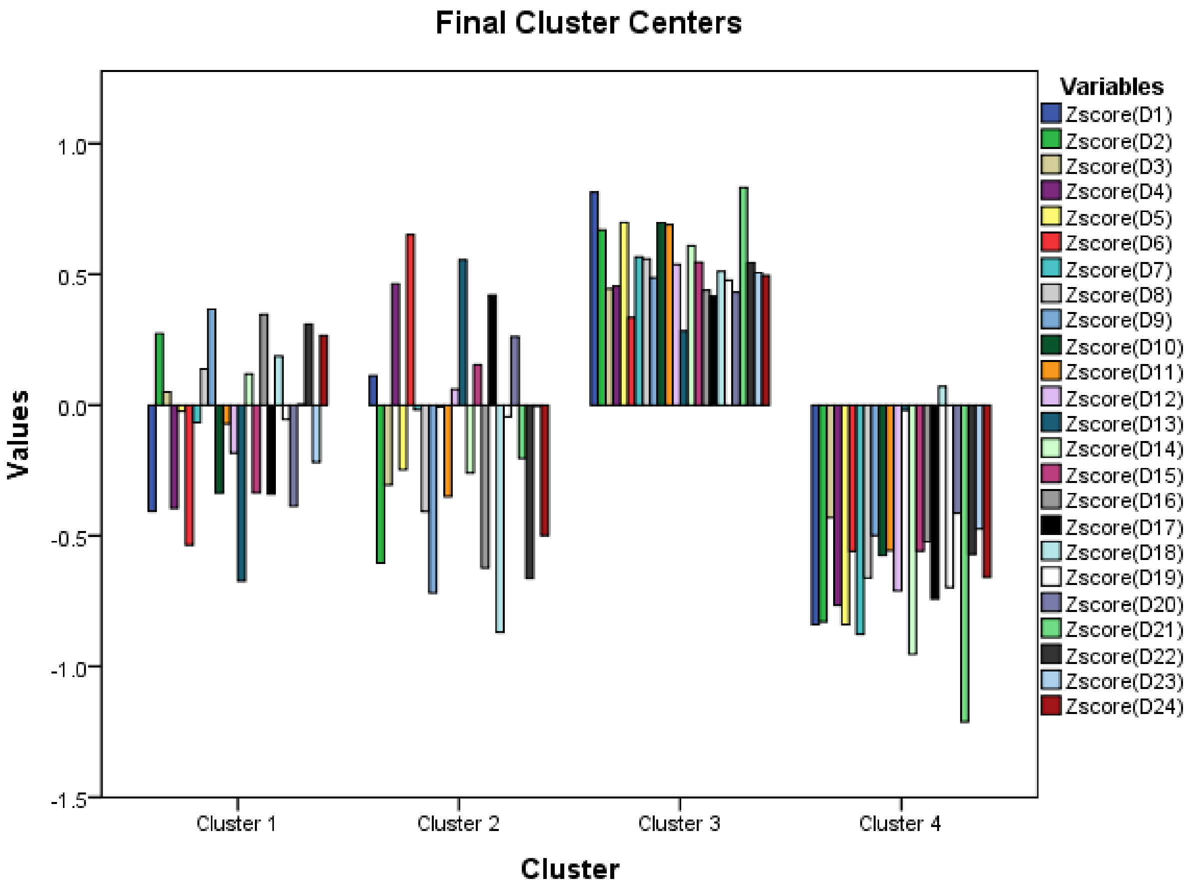
Cluster	Listeners
1	27
2	21
3	24
4	13
Total	**85**

An analysis of variance was carried out (df=3 and 81) to find out if the individual variables have a significant contribution to the formation of the clusters. The last two columns of Table I show the *F* and *p* values. The results suggest that all variables have a significant impact on the formation of the clusters.

To examine the nature of the differences between the individual clusters, data were displayed on a bar chart (Fig. 2.) The chart shows clear differences. The positive values in Cluster 3 and the negative values in Cluster 4 show that listeners who belong to these two groups differ in age estimation accuracy. Zero standardized values represent the overall mean. Positive values indicate higher, while negative values indicate lower cluster means compared to the overall mean.

Cluster 1 and 2 display similar distribution of data: positive and negative values also occur. Despite their similarity, a closer inspection of the charts reveals that individuals belonging to these clusters usually give different estimations. The sign of the standardized D values differs in 18 cases of the 24 acoustic stimuli, i.e. where positive values are found in Cluster 1, negative values occur with Cluster 2, and vice versa.

Figure 2. Bar chart of final cluster centres (4-cluster solution)



To determine the composition of the clusters, we created contingency tables to find the proportion of male and female, and musician and non-musician participants therein. Table 4 is a crosstabulation table showing the number of male and female participants in each cluster.

Table 4. The number of male and female listeners in each cluster

Cluster\Gender	Male	Female	Total
1	8	19	**27**
2	10	11	**21**
3	7	17	**24**
4	3	10	**13**
Total	**28**	**57**	**85**

The Chi-square test of association yields $\chi^2(3) = 2.91$, $p = .406$. This means that there is no statistically significant association between cluster number and listener gender; that is, the proportion of males and females does not differ significantly across the clusters. Similarly, the next crosstabulation table summarizes the number of musicians and non-musicians in the clusters (Table 5).

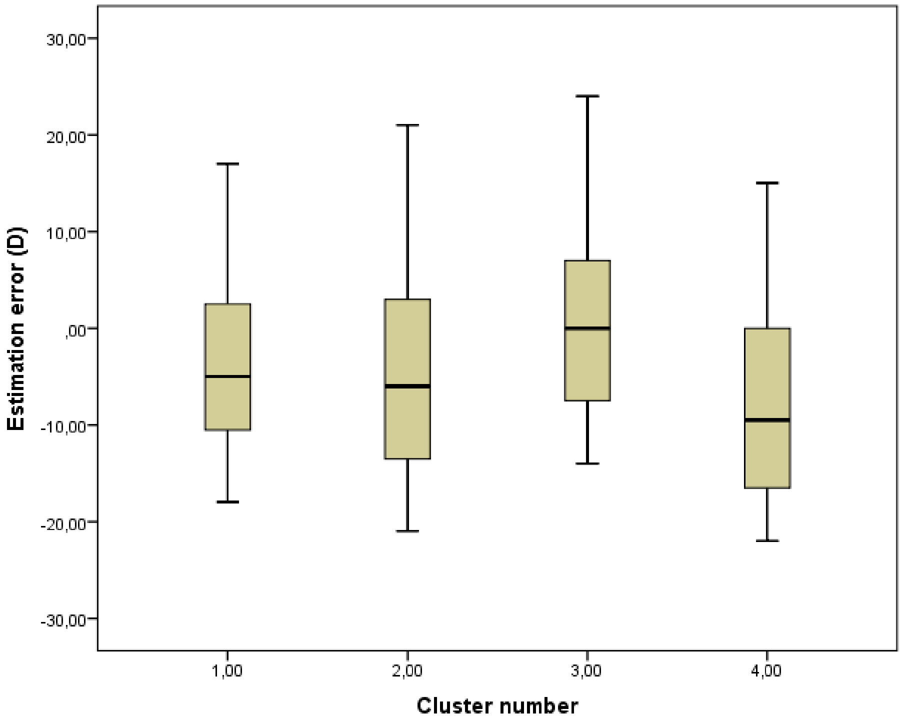
Table 5. The number of musicians and non-musician listeners in each cluster

Cluster\Musician	non-musician	musician	Total
1	7	20	**27**
2	14	7	**21**
3	14	10	**24**
4	8	5	**13**
Total	**43**	**42**	**85**

Again, a Chi-square test of association was run yielding $\chi^2(3) = 9.941$, $p < .05$. This means that clusters and musicianship have a significant stochastic relationship. We calculated the adjusted residuals for each cell. We found 3.1 for musicians and -3.1 for non-musicians in Cluster 1 which shows that musicians are overrepresented in this cluster. In the other clusters, the adjusted residuals were between -1.96 and 1.96, suggesting that musicians and non-musicians are not overrepresented or underrepresented in those clusters.

Finally, we determined if the *z*-scores, used previously, really mean underestimation or overestimation. Table II (Appendix) shows the unstandardized D values for each speaker and cluster, and Fig 3. shows the boxplot diagrams of those data.

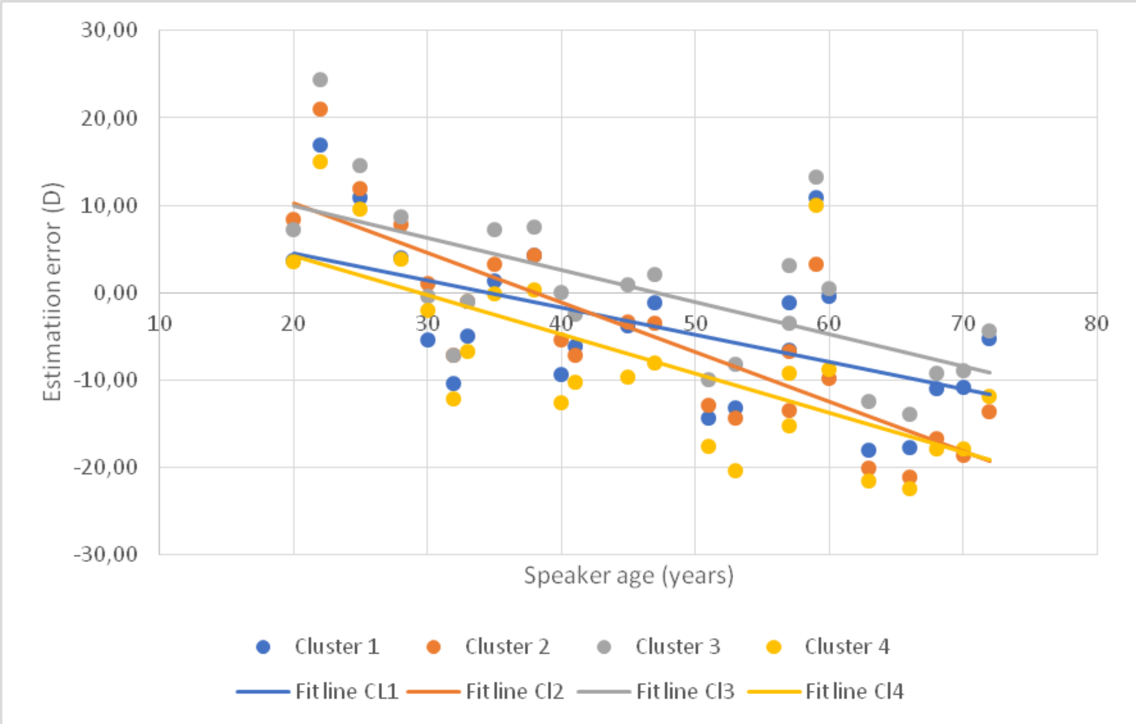
Figure 3. Boxplot chart of age estimation error values (D) (4-cluster solution)



The diagram shows that those in Cluster 4 are “strong underesimators”, while listeners in Cluster 1 and 2 are “slight underestimators”. Although the grand means in these clusters are close to each other, there are several cases when the listeners in one of the two clusters gave considerably better estimates than those in the other, which explains the existence of two clusters. The mean value of errors in Cluster 3 is 0, however, they had the largest degree of overestimation of the four clusters.

To better understand differences between the clusters, we also created scatter plots with trend lines fit on the dots for each cluster (Fig. 4). The scatter plot visually demonstrates that members of Cluster 3 overestimated young speakers’ age to the same degree they underestimated that of the older ones, as stated before. Listeners in Cluster 4 were better at estimating young speakers’ age but they very strongly underestimated the age of most of the speakers already from the age of 30. Listeners belonging to Cluster 1 and 2 are in between: Cluster 1 members are somewhat more balanced in age estimates, being better at estimating the age of older speakers, while listeners in Cluster 2 were worse, both estimating older and younger speakers’ age.

Figure 4. A scatter plot of estimation errors as a function of speaker age (4-cluster solution)



Finally, we calculated the absolute value of the estimation errors for each cluster and calculated their mean values. This calculation yielded 7.91, 9.79, 7 and 11.16 years for Cluster 1, 2, 3, and 4 respectively, which suggests that members of Cluster 3 were the most accurate in age estimation, while the less accurate group was Cluster 4. It should be noted though that these are mean values, which means that large differences may exist across the groups, independently from the main values.

The next step was to analyse the data using a three-cluster structure. Table 6 shows the number of listeners in each group. The similar numbers indicate that there was no subject who was very different from the rest of the participants. Table III (Appendix) shows the Z-standardized values of the differences.

Table 6. Number of listeners in each cluster (3-cluster solution)

Cluster	Listeners
1	26
2	28
3	31
Total	**85**

Again, an analysis of variance was carried out (df=3 and 81) to determine if the individual variables have a significant impact on the formation of the clusters. The *F* and *p* values suggest that all variables have a significant impact on the formation of the clusters.

To examine the differences between the individual clusters, data were displayed on a bar chart (Fig. 5.) This chart also shows clear differences. In Cluster 1 only positive values are found. Cluster 2 and 3 include dominantly negative values but the arrangement of the bars is different: the absolute value of the numbers in Cluster 2 seem to be larger than in Cluster 3, and the age of the same speakers was judged very differently, which is indicated by the fact that in 12 cases the sign of the standardized D values was different in Cluster 2 and 3.

To determine the composition of the clusters, we created further contingency tables to analyse the proportion of male and female, and musician and non-musician participants in the clusters. Table 7 is a crosstabulation table showing the number of male and female participants in each cluster. We performed a Chi-square test of association, yielding $\chi^2(2) = .78$, p = .677. Again, no statistically significant association between cluster number and listener gender was found.

A similar calculation was performed with musicians and non-musicians. The result of the Chi-square test was $\chi^2(2) = 4.503$, p = .105 which means that there is no statistically significant association between cluster number and listeners' musical training.

Figure 5. Bar chart of final cluster centres (3-cluster solution)

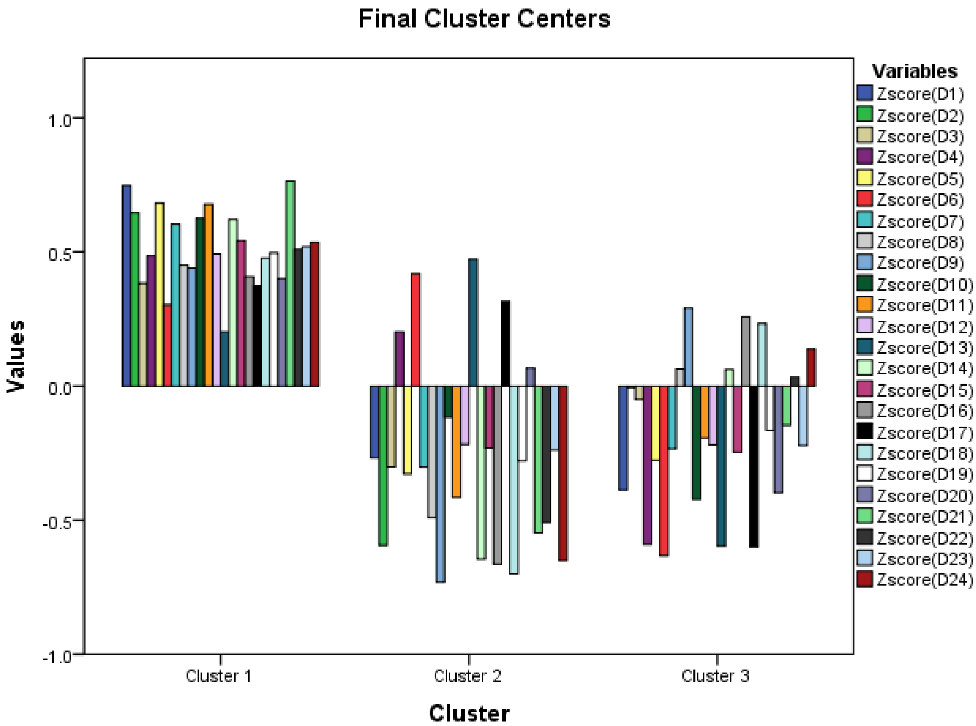


Table 7. The number of male and female listeners in each cluster

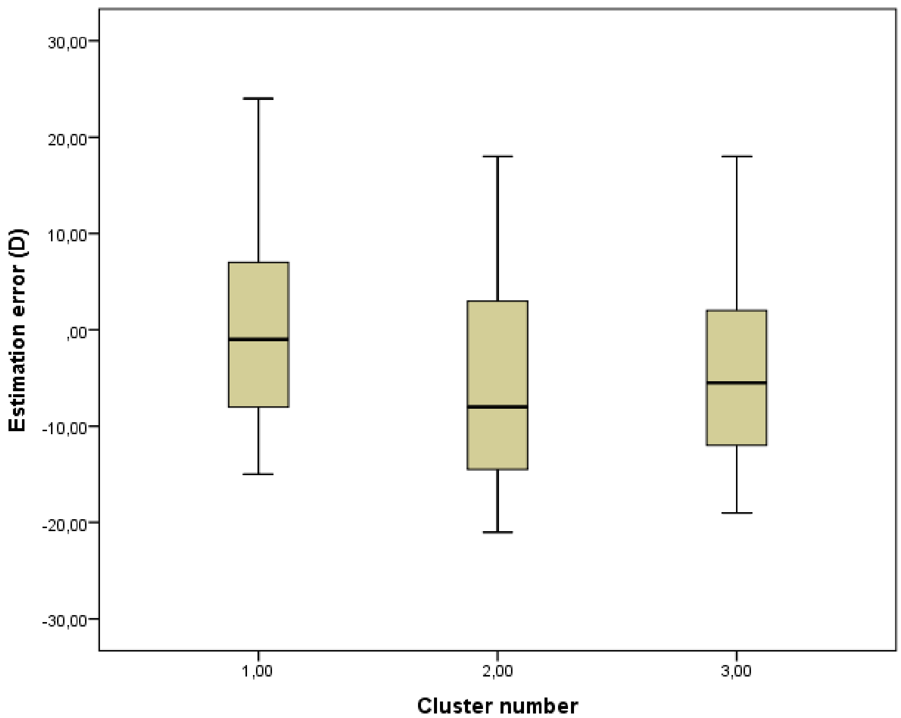
Cluster\Gender	Male	Female	Total
1	8	18	**26**
2	11	17	**28**
3	9	22	**31**
Total	**28**	**57**	**85**

Table 8. The number of musician and non-musician listeners in each cluster

Cluster\Musician	non-musician	musician	Total
1	8	18	**26**
2	11	17	**28**
3	9	22	**31**
Total	**28**	**57**	**85**

Finally, we analysed the raw D values to determine if the positive and negative standardized values mean overestimation or underestimation (Table IV in Appendix) and Figure 6 shows the related boxplots.

Figure 6. Boxplot chart of age estimation error values (D) (3-cluster solution)



Listeners belonging to Cluster 1 are usually “overestimators”, but the grand mean of the estimation error in this cluster is 0. Members of Cluster 2 and 3 are in general “underestimators”, but the grand means in these clusters are close to each other, there are several cases when the listeners in one of the two clusters gave considerably better estimates than those in the other, which explains the existence of two clusters. The scatterplot of the data with trend lines (Fig. 7.) shows tendencies similar to the 4-cluster solution, with the “strong underestimators’” Cluster 4 missing.

We have created a contingency table to find out possible overlaps between the cluster memberships of the 4- and 3-cluster solutions. Table 9 shows that 24 listeners of Cluster 3 in the 4-cluster solution are in Cluster 1 of the 3-cluster version. In addition, members of Cluster 4 are added to Cluster 2 and 3 in the 3-cluster solution.

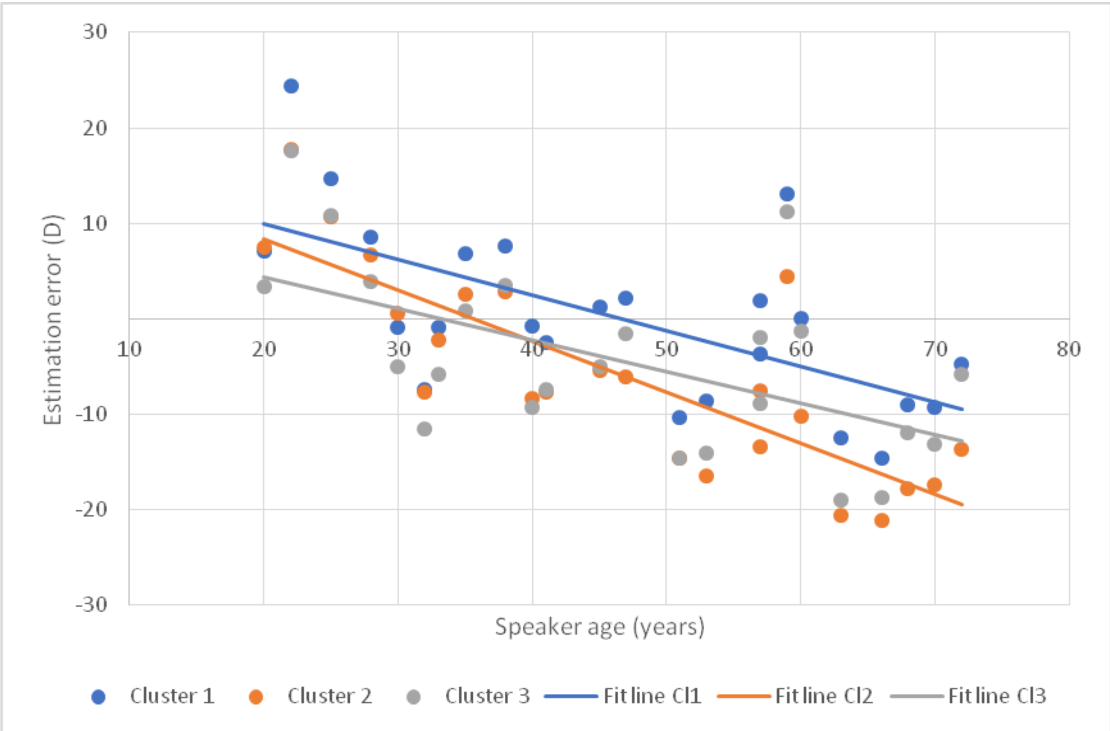


Figure 7. A scatter plot of estimation errors as a function of speaker age (3-cluster solution)

Table 9. Cluster memberships in the 4- and 3-cluster solutions

Clusters	1	2	3	4	Total
1	1	1	24	0	26
2	1	19	0	8	28
3	25	1	0	5	31
Total	27	21	24	13	85

Finally, we calculated the average value of the absolute differences in each cluster. This yielded 7.29, 10.20 and 8.624 years for Cluster 1, 2, and 3 respectively, which may reflect that members of Cluster 1 are the most accurate in general, but as Fig. 6. shows, they are overestimating speaker age more than those in the two other clusters.

Conclusions

This research was a first attempt to find individual differences in speaker age estimation by identifying listener groups that behave differently in a speaker age estimation experiment.

We found slightly different correlation coefficients between mean age estimates and chronological age when the whole listener group was broken down into subgroups, but in all cases correlation coefficient demonstrated a strong association. This means that our results are in agreement with previous findings (Table 1).

The cluster analyses, both the 4 and 3-cluster solutions revealed that several groups of listeners exist that differ in age estimation patterns. This confirms our first hypothesis. One possible explanation is that members of the individual clusters have developed different age estimation mechanisms, i.e. rely on different sets of acoustic parameters, or even other parameters of speech, or use the same set of parameters in different ways. A number of studies, such as Bóna (2015) have demonstrated strong correlation between age and tempo parameters. It is therefore possible that those with more accurate age estimates use tempo parameters as cues to age to a larger degree than those who provided less accurate estimates. Others may rely on other parameters more.

Another possible explanation is that own-age bias in speaker age estimation (Moyse et al., 2014) has developed in different ways in the listeners. Although listeners in this experiment constituted a homogenous group we found great differences in estimation errors, especially with elderly speech (Figures 4 and 7).

We found no gender differences in our experiment, which is in line with previous findings and this also confirms our second hypothesis. The third hypothesis was not confirmed, as musicianship did not result in more accurate age estimates, however, it was found to be a significant property in one of the clusters. It therefore requires further analyses why this exception occurred.

We believe that the results published here may contribute to a better understanding of the mechanisms of speaker age estimation. Further research should address the role of acoustic parameters and analyses age estimates in listeners belonging to other age groups.

References

Allport, G.W. & Cantril, H. (1934) Judging personality from voice. *Journal of Social Psychology; Political, Raical and Differential Psychology* 5. pp. 37-55.

Amilon, K., van de Veijer, J. & Schötz, S. (2007) The Impact of Visual and Auditory Cues in Age Estimation. In: Müller, C. (ed.) *Speaker Classification II. Lecture Notes in Computer Science*. Vol 4441. Berlin, Heidelberg: Springer. pp. 10-21.

Amir, O. – Enger, M. – Shabtai, E. – Amir, N. (2012) Identification of children's gender and age by listeners. *Journal of Voice* 26/3. pp. 313–321.

Bóna J. (2013) *A spontán beszéd sajátosságai az időskorban*. Budapest: ELTE Eötvös Kiadó,

Bóna J. (2015) Különböző beszédtípusok temporális sajátosságai az életkor és a nem függvényében. *Magyar Nyelvőr* 139/2. 201-213.

Braun, A. & Cerrato, L. (1999) Estimating speaker age across languages. *Proceedings of ICPhS* 99. pp. 1369-72.

Eppley, B. D., & Mueller, P. B. (2001). Chronological age judgments of elderly speakers: The effects of listeners' age. *Contemporary Issues in Communication Science and Disorders*, 28. pp. 5-8.

Eriksson, E., Green, J., Sjöstrom, M., Sullivan, K. P. H & Zetterholm, E. (2004) Perceived age: a distracter for voice disguise and speaker identification? *Proc. of FONETIK 2004*. Online: http:// https://www2.ling.su.se/fon/fonetik_2004/sullivan_et_al_fonetik2004.pdf [retrieved: 14.03.2019]

Gocsál Á. (2017) Az artikulációs tempó és az átlagos alaphang szerepe a beszélő életkorának megbecslésében. *Beszédkutatás 25*. 151-167.

Gocsál Á. (2018) Speaker Age Estimation by Musicians and Non-musicians. In Gósy M. & Gráczi T. E. (eds.): *Challenges in analysis and processing of spontaneous speech*. Budapest: MTA Nyelvtudományi Intézet. pp. 185-205.

Gnevsheva, K. & Bürkle, D. M. (2019) Age Estimation in Foreign-accented Speech by Native and Non-native Speakers. *Language and Speech.* DOI: 10.1177/0023830919827621

Gósy M., Gyarmathy D., Horváth V., Gráczi. T. E., Beke A., Neuberger T., & Nikléczy P. (2012) BEA: beszélt nyelvi adatbázis [BEA: Spoken language database]. In M. Gósy (ed.), *Beszéd, adatbázis, kutatások* [Speech, database, research] (pp. 9-24). Budapest: Akadémiai Kiadó. 9-24.

Hartman, D. E. (1979) The perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders* 12. pp. 53-61.

Huckwale M., Webb A. (2015) A Comparison of Human and Machine Estimation of Speaker Age. In: Dediu AH., Martín-Vide C., Vicsi K. (eds) *Statistical Language and Speech Processing. SLSP 2015*. Lecture Notes in Computer Science, vol 9449. Cham: Springer. pp. 111-122.

Hughes, S. M. & Rhodes, B. C. (2010) Making age assessments based on voice: the impact of the reproductive viability of the speaker. *Journal of Social, Evolutionary, and Cultural Psychology* 4/4. pp. 290-304.

Hummert, M. L., Mazloff, D. & Henry, C. (1999) Vocal characteristics of older adults and stereotyping. *Journal of Nonverbal Behavior* 23/2. pp. 111-132.

Hunter, E. J., & Ferguson, S. H. (2017). Listener estimates of talker age in a single-talker, 50-year longitudinal sample. *Journal of Communication Disorders* 68. pp. 103-112.

Huntley, R., Hollien, H. & Shipp, T. (1987) Influences on listener characteristics on perceived age estimations. *Journal of Voice* 1/1. pp. 49-53.

Jiao, D., Watson, W., Gig-Jano Wong, S., Gnevsheva, K. & Nixon, J. S. (2019) Age estimation in foreign-accented speech by non-native speakers of English. *Speech Communication* 106. 118-126.

Krepsz V. & Gósy M. (2016) A hangzásidő és a megakadásjelenségek hatása az életkorbecslésre In: Balázs G. & Veszelszki Á. (eds.) *Generációk nyelve: Tanulmánykötet,* Budapest: Inter; Magyar Szemiotikai Társaság; ELTE Mai Magyar Nyelvi Tanszék. 49-62.

Liang, C., Earl, B., Thompson, I., Whitaker, K., Cahn, S. Xiang, J., Fu, Q.-J., & Zhang, F. (2016). Musicians are better than non-musicians in frequency change detection: Behavioral and Electrophysiological evidence. *Frontiers in Neuroscience* 10.464. doi: 10.3389/fnins.2016.00464

Micheyl, Ch., Delhommeau, K., Perrot, X. & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research* 219. pp. 36-47.

Moyse, E. (2014) Age estimation from faces and voices: A review. *Psychologica Belgica* 54/3. pp. 255-265.

Moyse, E., Beaufort, A. & Brédart, S. (2014) Evidence for own-age bias in age estimation from voices in older persons. *European Journal of Ageing* 11. 241-247.

Pettorino, M. & Giannini, A. (2011) The speaker's age: a perceptual study. *Proceedings of ICPhS XVII*. pp. 1582-85.

Reubold, U., Harrington, J. & Kleber, F. (2010) Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication* 52. pp. 638-651.

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica* 138. pp. 1-10.

Sandman, K., am Zehnhoff-Dinnesen, A., Schmidt, C-M., Rosslau, K., Lang-Roth, R., Burgmer, M., Knief, A., Matulat, P., Vauth, M., & Deuster, D. (2014). Differences between self-assessment and external rating of voice with regard to sex characteristics, age, and attractiveness. *Journal of Voice*, 28(1), 128. e11-128.e18.

Schipp, T., Qi, Y., Huntley, R. & Hollien, H. (1992) Acoustic and temporal correlates of perceived age. *Journal of Voice* 6/3. 211-216.

Schötz, S. (2004) The role of F0 and duration in perception of female and male speaker age. *Proceedings of Speech Prosody 2004*. pp. 379-382.

Skoog Waller S, Eriksson M. & Sörqvist, P. (2015) Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology* 6/978. doi:10.3389/fpsyg.2015.00978

Sullivan, K.P.H., Bayard, D., Weatherhall, A., Gallois, C., Schlichting, F. & Pittam, J. (2000) Does media exposure to an accent impact upon the estimation of the age of speakers? *Proceedings of SST-2000: 8th Australian International Conference on Speech Sciences and Technology*. pp. 336-341.

Stölten, K. & Engstrand, O. (2003) Effects of perceived age on perceived dialect strength: A listening test using manipulations of speaking rate and F0. *PHONUM* 9. pp. 29-32.

Tatár, Z. (2013) Beszélőprofil-alkotás lehetőségei a kriminalisztikai fonetikában. Alkalmazott Nyelvtudomány XIII/1-2. 121-130.

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal NKFIH-FK-128814 számú pályázata támogatta.

The research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. FK-128814.

Appendix

Table I. The standardized D values (4-cluster solution)

	Final Cluster Centers				**ANOVA**	
	Cluster				F	sig.
	1	2	3	4		
Zscore(D1)	-.40645	.11204	.81342	-.83851	14.802	.000
Zscore(D2)	.27382	-.60423	.66983	-.82925	14.538	.000
Zscore(D3)	.05016	-.30572	.44412	-.43025	3.309	.024
Zscore(D4)	-.39548	.46405	.45384	-.76611	9.183	.000
Zscore(D5)	-.02429	-.24711	.69789	-.83878	9.660	.000
Zscore(D6)	-.53616	.65229	.33488	-.55837	10.451	.000
Zscore(D7)	-.06715	-.01821	.56637	-.87672	7.270	.000
Zscore(D8)	.13842	-.40692	.55893	-.66202	6.947	.000
Zscore(D9)	.36663	-.71905	.48688	-.49879	10.432	.000
Zscore(D10)	-.33661	-.00778	.69695	-.57500	7.901	.000
Zscore(D11)	-.07188	-.35039	.68868	-.55610	7.427	.000
Zscore(D12)	-.18315	.06092	.53773	-.71075	5.629	.001
Zscore(D13)	-.67405	.55526	.28346	-.02031	8.816	.000
Zscore(D14)	.11869	-.25994	.61029	-.95330	9.910	.000
Zscore(D15)	-.33569	.15376	.54537	-.55800	5.739	.001
Zscore(D16)	.34589	-.62286	.43928	-.52320	8.199	.000
Zscore(D17)	-.34087	.42083	.41665	-.74102	7.448	.000
Zscore(D18)	.18696	-.86894	.51189	.07035	10.274	.000
Zscore(D19)	-.05434	-.04391	.47741	-.69758	4.463	.006
Zscore(D20)	-.38715	.26099	.43123	-.41365	4.572	.005
Zscore(D21)	.00278	-.20407	.83162	-1.21142	20.799	.000
Zscore(D22)	.30807	-.66366	.54352	-.57120	10.268	.000
Zscore(D23)	-.21837	-.00468	.50618	-.47339	3.795	.013
Zscore(D24)	.26501	-.49934	.49544	-.65844	7.710	.000

Table II. The unstandardized D values (4-cluster solution)

	Cluster Number of Case			
	1	2	3	4
	Mean	Mean	Mean	Mean
D1	-9	-5	0	-13
D2	-7	-13	-4	-15
D3	-18	-21	-14	-22
D4	-5	-1	-1	-7
D5	-6	-7	-2	-10
D6	4	8	7	4
D7	-4	-3	1	-10
D8	-1	-7	3	-9
D9	-5	-14	-4	-12
D10	1	3	7	0
D11	-18	-20	-12	-22
D12	-14	-13	-10	-18
D13	-5	1	0	-2
D14	-1	-4	2	-8
D15	17	21	24	15
D16	0	-10	0	-9
D17	-10	-7	-7	-12
D18	11	3	13	10
D19	4	4	7	0
D20	4	8	9	4
D21	-13	-14	-8	-20
D22	-11	-19	-9	-18
D23	11	12	15	10
D24	-11	-17	-9	-18
mean	**-4**	**-5**	**0**	**-8**

Table III. The standardized D values (3-cluster solution)

	Final Cluster Centers			**ANOVA**	
	Cluster			F	sig.
	1	2	3		
Zscore(D1)	.74898	-.26679	-.38720	13.864	.000
Zscore(D2)	.64558	-.59354	-.00535	13.409	.000
Zscore(D3)	.38396	-.30075	-.05038	3.407	.038
Zscore(D4)	.48682	.20126	-.59008	11.253	.000
Zscore(D5)	.68205	-.32713	-.27657	10.760	.000
Zscore(D6)	.30263	.41872	-.63202	12.539	.000
Zscore(D7)	.60387	-.30160	-.23406	8.008	.001
Zscore(D8)	.45148	-.48983	.06377	6.929	.002
Zscore(D9)	.44026	-.73142	.29138	15.138	.000
Zscore(D10)	.62684	-.11525	-.42165	9.721	.000
Zscore(D11)	.67740	-.41483	-.19345	11.110	.000
Zscore(D12)	.49343	-.21664	-.21817	4.993	.009
Zscore(D13)	.20106	.47303	-.59588	11.439	.000
Zscore(D14)	.62114	-.64467	.06132	14.356	.000
Zscore(D15)	.54197	-.23014	-.24669	6.182	.003
Zscore(D16)	.40778	-.66473	.25839	11.794	.000
Zscore(D17)	.37538	.31493	-.59929	10.847	.000
Zscore(D18)	.47711	-.70053	.23258	13.960	.000
Zscore(D19)	.49690	-.27865	-.16507	5.190	.008
Zscore(D20)	.40141	.06819	-.39826	5.065	.008
Zscore(D21)	.76332	-.54764	-.14557	16.595	.000
Zscore(D22)	.50965	-.50971	.03294	8.243	.001
Zscore(D23)	.51987	-.23815	-.22092	5.621	.005
Zscore(D24)	.53478	-.65028	.13883	12.706	.000

Table IV. The unstandardized D values (3-cluster solution)

	Cluster Number of Case		
	1	2	3
	Mean	Mean	Mean
D1	-1	-8	-9
D2	-4	-13	-9
D3	-15	-21	-19
D4	-1	-2	-6
D5	-3	-8	-7
D6	7	8	3
D7	1	-5	-5
D8	2	-8	-2
D9	-5	-14	-6
D10	7	3	1
D11	-13	-21	-19
D12	-10	-15	-15
D13	-1	1	-5
D14	2	-6	-1
D15	24	18	18
D16	0	-10	-1
D17	-7	-8	-12
D18	13	4	11
D19	8	3	4
D20	9	7	4
D21	-9	-16	-14
D22	-9	-17	-13
D23	15	11	11
D24	-9	-18	-12
mean	**0**	**-6**	**-4**