View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repository of the Academy's Library

SPEAKER AGE ESTIMATION BY MUSICIANS AND NON-MUSICIANS

Ákos GOCSÁL
University of Pécs. Faculty of Music and Visual Arts, Institute of Music &
Hungarian Academy of Sciences, Research Institute for Linguistics
gocsal.akos@pte.hu

Abstract
Speaker age estimation is one of the most commonly researched fields in the domain of social perception based on voice. Previous findings confirm a strong correlation between the estimated and calendar age of speakers, however, younger adult speakers are usually perceived to be older, while older speakers are thought to be younger than their actual age. Effects of listener factors, such as age and gender have also been researched. The purpose of the present study is to examine if a more sophisticated auditory mechanism, which can be attributed to music training, results in more accuracy in speaker age estimation. The present research found correlation coefficients between calendar ages and mean estimated ages comparable to those reported in the literature, and musicianship and listener gender were not proven to have a significant effect on age estimations. Linear mixed models, implemented on three age groups, revealed some marginal differences between musicians and non-musicians, implying musicians' more accurate age estimations in some cases.

Keywords: speaker age estimation, social perception, musicianship

1 Introduction

1.1 Social perception

When we hear another person speaking, it is not only the linguistic message that we decode from the acoustic structures of speech. As Krauss (2002) says, in addition to what is said, human voice conveys considerable information about the speaker, and listeners use this information in human interaction. Based on speech, we infer a variety of speaker traits and make social judgments. These judgments are, however, based on vocal stereotypes and may even lead to prejudices (Drager, 2010). While a wide range of experiments have been conducted to explore humans' ability to infer different objective traits, such as speaker gender (Gelfer & Bennett, 2014), body size parameters (Rendall et al., 2007; van Dommelen & Moxness, 1995), or age (Moyse et al., 2014), a growing body of literature deals with the formation of voice-based impressions and

DOI: http://doi.org/10.18135/CAPSS.185

attitudes, including pleasantness (Hughes & Harrisson, 2013), attractiveness (Feinberg et al., 2005; Abend et al., 2015), perceived dominance (Puts et al., 2006; Fraccaro et al., 2013) or competence (Klofstad et al., 2015). Also, the perception of accented language or a dialect may evoke certain attributions and attitudes towards the speaker (Rubin et al., 1991, Cargile & Giles, 1997, Cross et al., 2001). Such impressions and attitudes may serve as bases for further decisions and actions; for example, voting behavior (Klofstad et al., 2015; Tigue et al., 2012) or mate choice (Collins, 2000; Shoup-Knox & Pipitone, 2015) may be influenced.

Little is known, however, about whether the perceptual behavior of individuals with different characteristics, such as age, gender, or any other factor, differs in their judgments. The results obtained so far seem to be equivocal. For example, Rendall et al. (2007) found no significant differences between male and female listeners' speaker size judgments; however, in an experiment by Charlton et al. (2013), men were better at classifying the apparent size of stimuli than female participants. In a similar experiment, Pisanski et al. (2016) did not find significant differences between sighted and congenitally blind subjects, and those who lost their sight later in life; and they did not find gender differences, either.

One factor that may differentiate listeners in their judgments is the difference in their auditory skills. In the context outlined above, the main purpose of this work is to examine if individuals with musical training are more accurate in one area of voice-based social perception, i.e., age estimation. In this work, those individuals are defined to have "better skills" and be "more accurate" whose age judgments are closer to the calendar age of the speakers. As discussed later in detail, several authors pointed out that classical musical training enhances auditory processing skills in many ways. Does this lead to more accurate age judgments? If results support this, one can infer that voice-based social perception in musicians is based on more reliable foundations than in non-musicians, which may possibly influence decisions in social interactions.

1.2 Speaker age estimation

Humans are in general able to judge the age of the speaker, purely based on the voice, although certain inaccuracies exist. Among the first researchers, psychologists Allport and Cantril (1934) published results of age estimation. They found that estimates were centered around a median of 35-40 years of age, however, they only used three speakers (actual ages: 27, 36, 51 years). More advanced methodology with a larger sample of speakers was used by Shipp and Hollien (1969). Their results suggested that strong correlation exists between the actual and the perceived ages ($r = 0.88$). However, strong correlation does not imply accurate judgments. While listeners tended to underestimate older speakers' age, they overestimated young adults' age (Huntley et al., 1987). Both the strong correlation between actual and estimated age and tendencies of underestimation and overestimation have been confirmed by many other researchers. For example, high degree of correlation between actual and estimated ages was found by Winkler et al. (2003) (r = 0.864 for spontaneous speech and 0.862 for reading), Cerrato et al. (2000), who used telephonic voices (r = 0.77), or Stölten and Engstrand (2002), who found $r = 0.92$ when their younger and the older speakers' groups were collapsed.

In speaker age estimation, two main approaches exist. Cross-sectional experiments use different speakers from a selected period, most commonly, recordings from speakers of the time when the research is carried out. In longitudinal settings, researchers use speech samples of the same persons recorded at different time points.

In a cross-sectional experiment, Hughes and Rhodes (2010) examined more closely how accurately the age of speakers belonging to different age groups can be judged. They found that raters were fairly accurate when determining the age of children and adolescents. A slight overestimation of speaker age was found with young male speakers, while the underestimation of the female speakers' age was more prominent for speakers in the age group of 23-34. The degree of underestimation, however, was higher for the male speakers in age groups 35-45 and 46-55. Finally, the age of speakers over 56 years of age was also underestimated but no significant difference between male and female speakers was found. Sandman et al. (2014) also reported a "tendency for estimates to gravitate to middle age", implying correct estimates for middle-aged speakers, while more prominent differences were found between the estimated and real ages for younger or older speakers.

A longitudinal experiment was carried out by Reubold et al. (2010). Two recordings from Queen Elizabeth II and three from broadcaster Alistair Cooke were used as stimuli. Listeners were accurate in identifying the Queen's age as younger when listening to a 1972 recording, while they were right to identify her as older when hearing a 1983 recording. Cooke was identified younger on a 1947 recording than on the 1970 recording; however, no significant differences between the 1970 and 1990 recordings were found. It was also found that the manipulation of f0, while other parameters remained constant, resulted in an unequivocal effect on perceptual age, while shifting F_1, while keeping f0 constant, provided mixed results. In another longitudinal study, 60 samples were extracted from public addresses by a male speaker over 48 years (aged 48-97). When the speaker was between 49 and 68 years old, he was estimated to be between 58 and 68, overestimating his age by 6 years on average. When the talker reached age 68, the estimates were in line with his calendar age; however, about 5 years of underestimation on average occurred (Hunter & Ferguson, 2017).

Concerning potential effects of listener factors, fewer results are available. Eppley and Mueller (2001) found that both the young and the old listener group underestimated the age of their elderly speakers; however, the older listeners' estimations were, on the average, some four years closer to the actual age of the speakers than those made by the younger group but the difference was not significant. These background factors were summarized by Moyse (2014), who concluded that younger listeners are more accurate than older ones, irrespective of the age of stimuli, and the age of female voices is more accurately estimated than that of male speakers.

Since the present study is carried out with Hungarian subjects, it is important to briefly review available literature. Previous findings with Hungarian speakers and listeners are in harmony with international results. A strong correlation between the real and estimated age (Gocsál, 1998; Bóna, 2013), inaccuracies of judgments, including overestimation or underestimation, have been documented in a variety of experimental settings (Gósy, 2001; Tatár, 2013; Tóth, 2014; Krepsz & Gósy, 2016; Gocsál, 2017).

1.3 Auditory perceptual skills of musicians

In the literature of speaker age estimation, listener factors such as age and gender have been considered so far. Little is known about other factors that may differentiate age judgments. One possible factor of this kind, which may play a role here, is the "quality" of the auditory processing mechanism of the listeners. How do listeners with more sophisticated auditory skills perceive extralinguistic contents of speech? Are they better at age estimation? One specific group of people in which auditory processing skills are expected to be better than those of others is musicians. For assessing potential differences in the auditory mechanisms of musicians and non-musicians, a large part of research uses non-linguistic acoustic stimuli. In most of the cases, musicians do demonstrate more sophisticated skills, i.e., they are more sensitive to smaller differences in the acoustic sign. For example, musicians are better at identifying changes of frequency of pure tones both in silent and noisy conditions (Liang et al., 2016). Further experimental evidence for enhanced performance on frequency discrimination (Micheyl et al., 2006; Eadie et al., 2010; Madikal Vasuki et al., 2016; Meha-Bettison et al., 2018), better auditory temporal-interval discrimination (Banai et al., 2012), or both (Boebinger et at,. 2015). Also, better pitch contour identification was found in musicians; however, training with pitch discrimination exercises results in the improvement of both musicians' and non-musicians' performance (Micheyl et al., 2006; Wayland et al., 2010).

In some cases, no such differences were found. For example, in tonal processing, sense of completion of a melody was rated similarly by musicians and non-musicians, even though neural responses were different. This suggests that there may be specific mechanisms available for non-musicians, compensating for their lack of musical training (Amemiya et al., 2014).

Plasticity of the auditory cortex was also found to be induced by musical training, and Pantev and Herholz (2011) suggested that making music may even contribute to a more effective recovery from impaired auditory or motor skills caused by lesions. Musical training may also mitigate changes in auditory perception commonly occurring in aging adults (Alain et al., 2013).

Discrimination of pure tones does not necessarily explain better performance (Eadie et al., 2008), spectrally more complex stimuli may help determine the underlying differences. It is therefore of particular interest to examine language related perceptual skills. In this context, effects of music training were studied by Flaugnacco et al. (2015). Their findings suggest that music training boosts phonological awareness, rhythmic abilities and reading skills, even when these skills are impaired. Speech segmentation skills are also improved by music training, which may contribute to children's language development (François et al., 2013). Musicians are also more sensitive to subtle pitch changes, both in non-linguistic tones and spoken sentences (Deguchi et al., 2012). Kühnis et al. (2013) used spectrally and temporally manipulated CV syllables, and musicians demonstrated an increased responsiveness to the acoustic stimuli. Alexander et al. (2005) found that English-speaking musicians performed significantly better at identifying and discriminating Chinese lexical tones than English-speaking non-musicians.

Musicianship seems also to be an advantage when one perceives speech in noise. Better perceptual abilities of musicians were demonstrated by Parbery-Clark et al. (2009, 2011). Positive effects of music training on speech in noise perception were also demonstrated in children by a longitudinal research (Stater et al., 2015). A review paper by Coffey et al. (2017) compared research results obtained over a wide range of conditions and confirmed musicians' advantage in speech-in-noise perception. However, some results do not support this conclusion: for example, those found by Boebinger et al. (2015) who concluded that it was nonverbal IQ rather than musical training that predicted speech perception thresholds in noise. A more realistic version of this type of experiment uses the multi-talker masking approach where the masking noise simulates a "cocktail-party" environment (Swaminathan et al., 2015). Musicians outperformed non-musicians when the maskers were spatially separated from the target voice, but no significant differences were found when the masking voices and the target voices were collocated.

In another experiment by Sadakata and Sekiyama (2011), musicians were also better at discriminating and identifying morphed speech sounds, that is, they were more sensitive to subtle temporal and timbre differences of speech sounds when they heard minimal pairs of words, one of them unaltered, while the other chosen from a series of slightly morphed versions of the words. Deme's (2017) results, however, suggest that professional singers do not enjoy a perceptual advantage in the identification of high-pitched, sung vowels over naïve listeners.

Concerning speech and music perception, only few comparative studies have been published. Chartrand and Belin (2006) found that in discriminating instrumental sounds and human voice samples with different timbres, musicians outperformed non-musicians. Musicians used more response time though, which was most likely due to a deeper level processing of the sounds. Another question related to the link between musicianship and speech perception addresses the issue of disorders related to music perception. An experiment by Liu et al. (2015) demonstrated that individuals who suffer from congenital amusia, i.e. are unable to discriminate pitch levels of a melody from birth, also experience difficulties in speech comprehension. Their results also revealed that amusia is not limited only to pitch processing. Amusic subjects achieved lower scores in perceiving flat f0 sentences, which implies that their deficits in speech perception go beyond pitch processing. These results are worth noting because they contradict statements that amusic individuals have a normal understanding of speech (Peretz & Vuvan, 2017). A wide range of other research related to differences in musicians' auditory skills is available in the literature. For example, when evaluating dysphonic voices, musicians demonstrated significantly more agreement in judging the breathiness in dysphonic speakers than non-musicians (Eadie et al., 2008).

The question whether auditory processing mechanisms for voice and music are separate or are at least in part overlapping has been discussed by several researchers. Chartrand and Belin (2006) proposed that voice timbre and instrument timbre discrimination involve similar mechanisms. If music related perceptual mechanisms were completely independent from voice related mechanisms, experiments would only demonstrate musicians' advantage of discriminating instrument timbres, but not vocal timbres. However, their results showed that musicians outperformed non-musicians both in voice and instrument timbre discrimination tasks as well, thus, overlapping perceptual mechanisms were suggested. Hausen et al. (2013) found that music and speech perception is shared by the perception of rhythm and pitch. Strait and Kraus (2011) mentioned several specific common mechanisms such as auditory attention, working memory, neural function in challenging listening environments, sequential sound processing, and sensitivity to temporal and spectral aspects of sound. Perhaps the most comprehensive model that connects the perception of speech and music is proposed by Patel (2014). His expanded OPERA hypothesis (O = overlap between neural networks processing speech and music, P = higher precision of processing is demanded when music is heard, E = emotion, R = repetition of musical activities, A = focused attention demanded by music), proposes that when music and speech share auditory perceptual mechanisms, and music places higher demands on those auditory mechanisms than speech does, speech processing may be enhanced in musicians (Patel, 2014). One possible explanation is that musical training improves the generic constitutional properties of the auditory system (Kühnis et al., 2013).

In the context of the present research, perception of non-linguistic contents of speech is in focus. One direction of this kind of research is related to the perception of emotions. More intense patterns of emotional activation were observed in musicians than in non-musicians when they were listening to classical music (Mikutta et al., 2014), and out of the different types of emotive stimuli, negative emotions expressed in music are more arousing for musicians while responses to happiness in music evoke no specific activations (Park et al., 2014). It is therefore an interesting question to examine how musicians perceive emotions from speech. Results suggest that musicians have better skills in identifying the emotions conveyed by tone sentences mimicking the prosody of spoken sentences. Also, musically trained adults were better at identifying spoken utterances that were emotionally neutral (Thompson et al., 2004). In an experiment by Pinheiro et al. (2015), musicians were more accurate in recognizing angry prosody from sentences, and, their general conclusion was that extensive musical training may impact different stages of vocal emotional processing.

1.4 Research questions

It has been demonstrated that humans are capable of estimating the speaker's age, based on the speaker's voice, although certain inaccuracies occur in estimations. Previous studies have found that speech rate, and in certain cases, speaking fundamental frequency are key properties of speech that listeners use (Reubold et al., 2010; Stölten & Engstrand, 2003; Winkler, 2007; Skoog Waller et al., 2012) in age judgments. If musicians' auditory skills are more sensitive to temporal and frequency differences (Tervianemi et al., 2005; Elmer et al., 2012), it seems reasonable to raise the question if musicians' age judgments are closer to the speakers' calendar age than those of non-musicians. In particular, the research questions are as follows: (1) Do musicians' and non-musicians' age estimations correlate with the speakers' calendar age? (2) Do musicians' age estimations differ from those of non-musicians in different age groups of speakers? (3) Do differences in male and female listeners exist? Since we assume that musicians have more sophisticated auditory mechanisms, therefore it is hypothesized that (1) musician listeners' age estimations correlate stronger with the speakers' calendar ages than non-musicians' estimations and (2) in each of three different age groups of speakers, musicians' age estimations are more accurate than those of non-musicians, and (3) we expect no differences between male and female listeners' age estimations.

2 Methods

2.1 Acoustic stimuli

24 male speakers (age range: 20-72 years) were selected from the BEA spontaneous speech database (Gósy et al., 2012) in a way that their ages were approximately evenly distributed over the age range of the whole sample, resulting a mean difference of 2.16 years (range: 0-4 years) between two adjacent speakers. The speakers were nonsmokers and had either BSc or higher degrees or were students. Of the recordings, samples of approx. 20-30 seconds were chosen from the "interview" or "argument" task of the BEA recording protocol. The chosen speech samples were from longer monologues in which the speakers were talking about general themes, e.g. their job, local transportation, hobbies etc. in an emotionally neutral way. The speech samples included no textual information that the listeners may have used for inferring the speakers' age.

2.2 Listeners

Listeners were students of the University of Pécs (n = 85, age range: 19-37, median: 22), without any prior training in phonetics. There were 42 listeners who study music (instrumental players of classical music, with at least 8 years of music education), and 43 students of other fields (sociology, fine arts, media), who cannot play an instrument and never received any kind of music training, apart from the compulsory singing classes at school. There were some non-musician participants who had received some music training, but they were excluded from the study. Table 1 shows the gender distribution of musician and non-musician participants. No participant reported any kind of hearing impairment, complaints or previous medical treatment that may have influenced their auditory processes.

Table 1. Participants of the study

	musicians	**non-musicians**
males	14	14
females	28	29

2.4 Procedure

The listening tasks took place in a silent seminar room of the Zsolnay campus of the University of Pécs, through good quality multimedia speakers, in groups of 5-10. Listeners were instructed to estimate the speaker's age in years and to write down their estimations on an answer sheet. Prior to the experiment, listeners were familiarized with the task by playing three sound samples to them. At the same time, the experimenter tested if all the listeners can clearly hear the sound samples in the seminar room. Each of the sound samples was introduced by a 1 second 440 Hz beeping sound and a 2 second silent pauses. Each sound sample was played once. A sound sample was only played when all the subjects were ready with writing down their estimations for the previous sample. The speech samples were played in the same, randomized order in each group. Once the data were obtained, Pearson's correlation coefficients were calculated and linear mixed models were fitted using the SPSS 23.0 software.

3 Results

3.1 Correlations between mean estimated age and real age

To establish possible similarities with previous findings, mean values of the age estimations were calculated for each speaker, and Pearson's correlation coefficients between the mean estimated ages and calendar ages were calculated. Figure 1 shows the scatterplot of the estimated mean ages against calendar ages with all listener groups collapsed. Pearson's correlation coefficient indicates strong correlation ($r = 0.806$, $p < 0.001$). A strong correlation, however, does not necessarily suggest accurate judgments. The deviation of the regression line (dashed line) from the y = x line (solid line) suggests a tendency for younger speakers to be perceived older, while older speakers are believed to be younger than their actual age. The intersection of the two lines suggests accurate age estimations for speakers between 35 and 40 years of age.

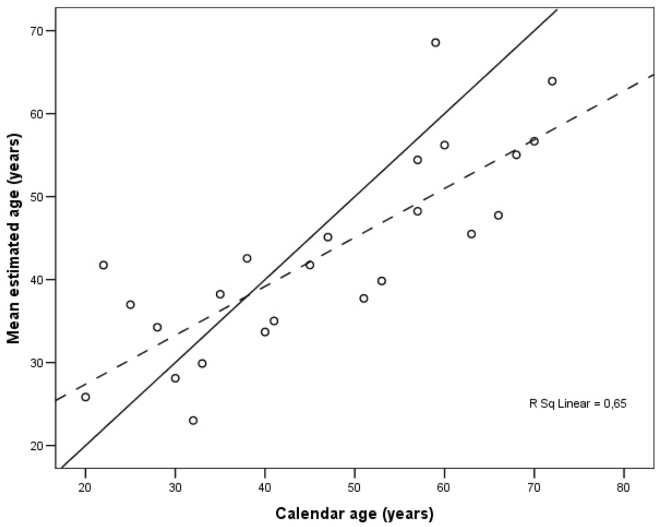

Figure 1.
Scatterplot of mean estimated age and calendar age, all listener groups collapsed

Figure 2 shows musician participants' mean age estimations against the calendar age of the speakers. For both musician males and females, correlation coefficient is comparable or identical with what was found with the overall group ($r = 0.803$ in males and $r = 0.806$ in females, $p < 0.001$ in both cases). The regression lines predict virtually identical age estimations in younger speakers, while female listeners' age estimations are predicted to be slightly closer to the real age of the older speakers.

Figure 3 illustrates non-musician participants' mean age estimations against the calendar age of the speakers.

Again, significant correlation coefficients are found (non-musician males: $r = 0.839$, non-musician females: $r = 0.777$, $p < 0.001$), which is slightly stronger in males and slightly weaker in females than what was found with the musicians' group. The two regression lines are almost identical.

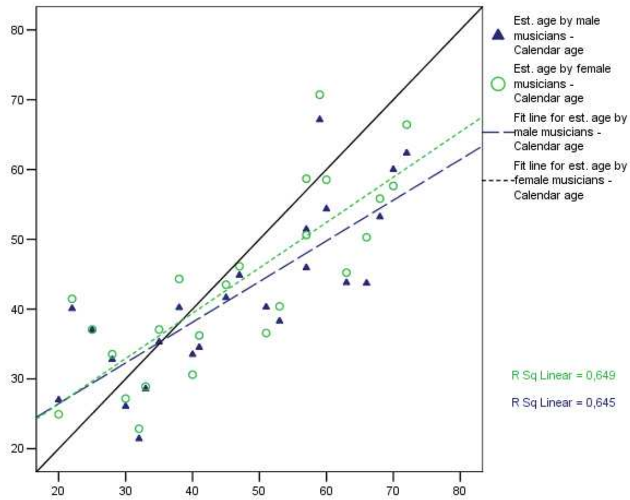

Figure 2.
Scatterplot of mean estimated age and calendar age, musician listeners

Next, to explore whether listener gender and musicianship significantly influence age estimations, a linear mixed model was fitted with estimated age as dependent variable, calendar age as covariate and listener gender and musicianship as fixed factors. For this analysis, all data were used rather than mean values of age estimations. The obtained F-values were as follows: $F(1, 70.591) = 0.988$, $p = 0.324$ for listener gender, $F(1, 70.591) = 0.562$, $p = 0.456$ for musicianship and $F(1, 70.591) = 0.202$, $p = 0.654$. These values suggest that neither the individual fixed factors, nor their interaction is significant. In sum, although minor differences were observed between the listener groups when

Pearson's correlation coefficients were used, a deeper analysis, including all data rather than mean values, revealed no effect of listener gender and musicianship.

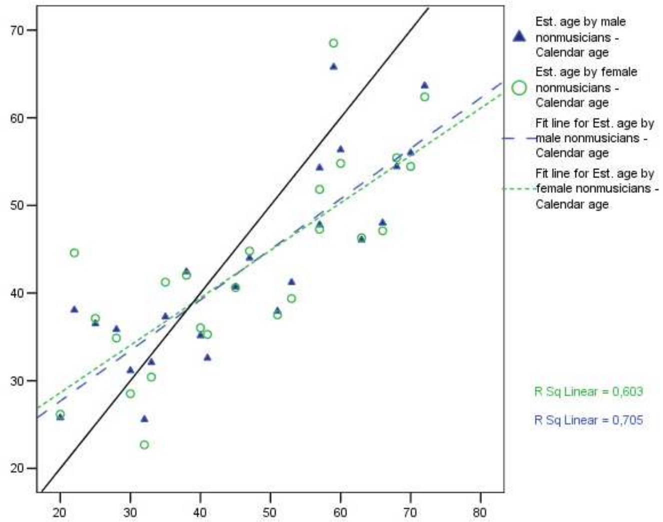

Figure 3.
Scatterplot of mean estimated age and calendar age, non-musician listeners

3.2 Accuracy of age estimations in different speaker age groups

To determine the nature of accuracy of speaker age estimations, linear mixed models were implemented again in a different way. First, for each speaker, differences between the estimated and the calendar ages were calculated, and the difference values were z-standardized. Of the resulting 2040 values (85 listeners x 24 speakers) seven (0.3%) were excluded because of the z-score being over 3 or below –3. Figure 4 demonstrates the related boxplots for all speakers.

Although a visual inspection of the boxplots would suggest a slight difference in the mean values, calculations do not confirm a significant difference. Several covariate structures were tested for repeated effects, but none resulted in significant fixed effects. The lowest AIC value was obtained with the covariate structure "scaled identity". The resulting F-values are as follows: $F(1, 81.058) = 0.089$, $p = 0.766$ for the intercept, $F(1, 81.058) = 1.128$, $p = 0.291$ for the gender of the listener, $F(1, 81.058) = 0.341$, $p = 0.570$ for musicianship, and $F(1, 81.058) = 1.126$, $p = 0.292$ for the interaction of gender and musicianship. These results suggest that there is no significant tendency of underestimation or overestimation of age between the listener groups, i.e., it cannot be stated that when all speakers are collapsed into a single group, age estimations of musicians and non-musicians, males and females significantly differ.

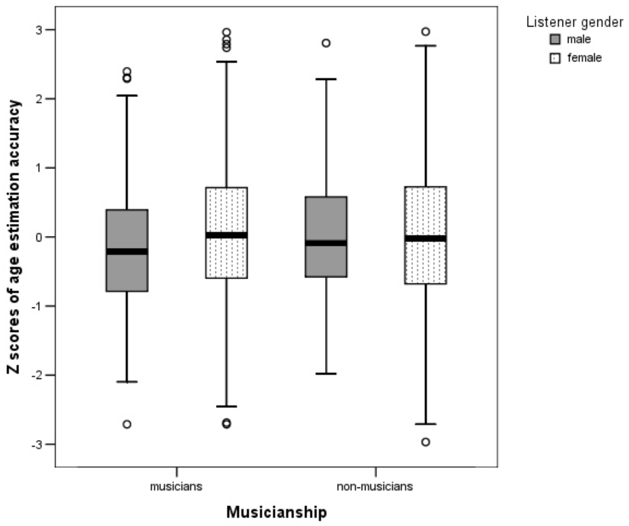

Figure 4.
Boxplot of *z*-scores of age estimations for all speakers

It should, however, be noted that the above calculation does not rule out possible differences in different age groups. For a further analysis, three groups were made of the speakers and the same calculations were repeated separately with the groups. The young speakers' group included speakers where overestimation of age was most likely. The middle-aged speakers' group, accurate estimations were expected, and, in the older speakers' group, underestimation of age was most likely. Again, in each group, *z*-scores over 3 or below –3 were excluded. First, age estimations for 5 speakers, between 22 and 30, were analyzed. Figure 5 shows the results.

Again, a visual inspection of the boxplots shows lower *z*-scores for the male musicians. Since the age of the younger speakers was in general overestimated, lower *z* values imply smaller differences between the estimated age and the calendar age, i.e. better age estimations. In the linear mixed model, different covariate structures resulted in very similar outcomes. Applying the covariate structure "scaled identity", the following *F*-values were obtained: $F(1, 85.339) = 0.119$, $p = 0.731$ for the intercept, $F(1, 85.339) = 0.232$, $p = 0.631$ for the gender of the listener, $F(1, 85.339) = 3.411$, $p = 0.068$ for musicianship, and $F(1, 85.339) = 0.000$, $p = 1.000$ for the interaction of gender and musicianship. These findings suggest a marginal but not significant main effect of musicianship, i.e., the lower z-scores of musician males suggest slightly more accurate estimations than those of the other groups, while non-musician females seem to have been less accurate than the others. But, again, this difference is not significant.

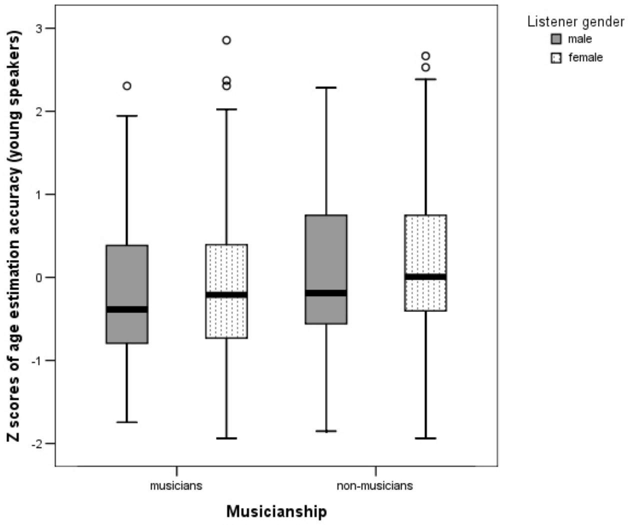

Figure 5.
Boxplot of *z*-scores of age estimations for young speakers

Finally, the calculations were carried out with the older speakers' group (10 speakers, ages between 53-72 years). In this case, one listener group, musician females, seems to be prominent, as Figure 7 shows.

The second group consisted of 9 speakers between 32 and 50 years of age. Figure 6 shows slight differences between the boxplots; however, the differences are not significant. Again, different covariate structures result in substantially the same outcomes. The results, i.e., $F(1, 84.804) = 0.036$, $p = 0.850$ for the intercept, $F(1, 84.804) = 0.510$, $p = 0.477$ for the gender of the listener, $F(1, 84.804) = 1.255$, $p = 0.266$ for musicianship, and $F(1, 84.804) = 0.130$, $p = 0.723$ for the interaction of gender and musicianship, demonstrate not even a marginal effect of any of the factors.

Again, several covariate structures were tested but substantially the same results were obtained. With the "scaled identity" covariate structure, intercept ($F(1, 85.073) = 0.057$, $p = 0.812$) and musicianship ($F(1, 85.073) = 0.709$, $p = 0.402$) were not significant fixed factors, neither was gender ($F(1, 85.073) = 1.349$, $p = 0.249$) or the gender × musicianship interaction ($F(1, 85.073) = 3.047$, $p = 0.085$). Removal of intercept did not improve the effect of the gender × musicianship interaction. This marginal but not significant interaction suggests somewhat better age estimations of female musicians (see Figure 7). In this case, higher *z*-values imply better age estimations since higher estimated ages were closer to the calendar age.

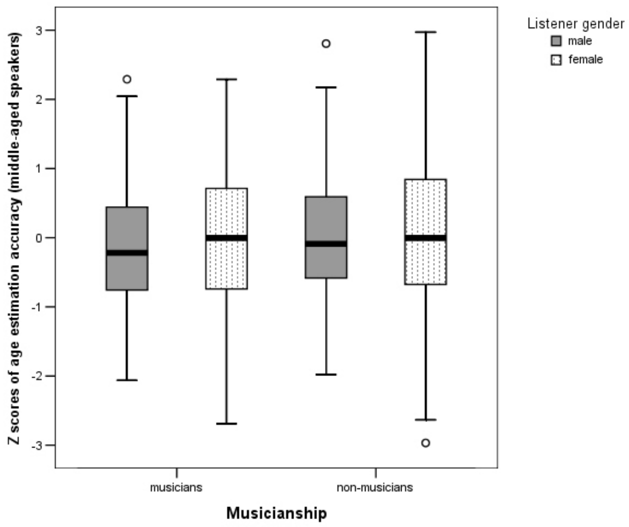

Figure 6.
Boxplot of *z*-scores of age estimations for middle-aged speakers

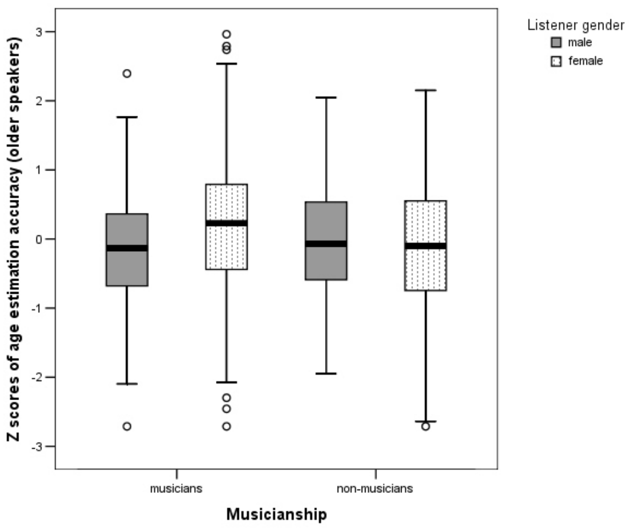

Figure 7.
Boxplot of *z*-scores of age estimations for older speakers

4 General discussion

The main purpose of the present paper was to reveal if musicians are more accurate in speaker age estimation than non-musicians. Our first hypothesis was not confirmed. Pearson correlation coefficients between mean estimated ages and calendar ages (*r*-values between 0.777 and 0.838) are in line with those reported in previous studies (Huntley et al., 1987; Cerrato et al., 2000; Winkler et al., 2003), and differ between the listener groups, but a repeated measures linear mixed model revealed no significant effect of listener gender and musicianship.

Our second hypothesis was not confirmed either. No statistically significant differences were demonstrated by further analyses carried out with different age groups. The third hypothesis was confirmed, no significant differences were found between male and female listeners' estimations. However, there was a non-significant tendency for male musicians to be more accurate in the age estimation of younger speakers, while female musicians were marginally better at estimating the age of the older speakers.

One possible explanation for not finding more prominent differences is that differences between the vocal parameters of different aged speakers are large enough to be perceived even by listeners with less sophisticated perceptual mechanisms. The marginally better performance of male and female musicians in the two cases may reflect that they have "more accurate" vocal prototypes, at least for those age groups. Krepsz and Gósy (2016) outlined the nature of such vocal prototypes, which one builds through experiencing speakers and voices. Vocal prototypes then serve as bases for social perception, including age estimation. Future research should address this marginal difference. Other research has pointed out the role of voice quality in assessing the speaker's dominance or attractiveness (Fraccaro et al., 2013; Puts et al., 2014) so it is possible that for male listeners it is important to develop vocal stereotypes so that they are more accurate in the perception of peers, while, for female listeners, it may be important to develop skills to be more accurate in the perception of more mature males. Musical training may be an advantage in the development of such vocal stereotypes and skills, but again, since non-significant differences were found, further experiments are needed.

One of the main limitations of the present study is that the role of acoustic parameters was not analyzed. Previous studies have demonstrated the role of speech rate (Skoog Waller et al., 2015; Gocsál, 2017), fundamental frequency and formant structure (Reubold et al., 2010), or a combination of several parameters (Winkler, 2007; Harnsberger et al., 2008) in age estimations. Future work should also address the question if musicians and non-musicians use different strategies in using the acoustic parameters of speech in speaker age estimation. Their marginally better estimations in some cases may be attributable to their attention that they pay to subtle details in the acoustic structure of speech. However, the effect of those subtle details is not large enough to make significantly better estimates. Also, further studies will need to examine why female listeners' estimations covered a larger range than those of males (especially with speakers in the middle-aged and older group, see Figures 6 and 7). It is possible that they are more uncertain in giving estimations on opposite-sex speakers, but a similar experiment with female speakers would be necessary for a comparison.

In conclusion, it can be stated that this experiment was the first attempt to demonstrate potential benefits of musicianship in speaker age estimation, and although no prominent differences were found, the marginally better estimations of musicians in some cases raise questions that deserve attention in the future.

References

Abend, P., Pflüger, L. S., Koppensteiner, M., Coquelle, M., & Grammer, K. (2015). The Sound of Female Shape: A Redundant Signal of Vocal and Facial Attractiveness. *Evolution and Human Behavior, 36,* 174-181.

Alain, C., Zendel, B. R., Hutka, S., & Bidelman, G. M. (2013). Turning down the noise: The benefit of musical training on the aging auditory brain. *Hearing Research, 308*, 162-173.

Alexander, J. A., Wang, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *9th European Conference on Speech Communication and Technology, Eurospeech Interspeech* (pp. 397–400). Lisbon, Portugal.

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

Amemiya, K., Karino S., Ishizu, T., Yumoto, M., & Yamasoba, T. (2014). Distinct neural mechanisms of tonal processing between musicians and non-musicians. *Clinical Neurophysiology, 125,* 738-747.

Banai, K., Fisher, S, & Ganot, R. (2012). The effects of context and musical training on auditory temporal-interval discrimination. *Hearing Research, 284,* 59-66.

Boebinger, D., Evans, S., Scott, S. K., Rosen, S., Lima, C. F., & Manly, T. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *Journal of the Acoustic Society of America, 137*(1), 378-387.

Bóna, J. (2013). A spontán beszéd sajátosságai az időskorban [Spontaneous speech in old age]. Budapest: ELTE Eötvös Kiadó.

Cargile, A. C., & Giles, H. (1997). Understanding language attitudes: Exploring listener affect and identity. *Language & Communication, 17*(3). 195-217.

Cerrato, L., Falcone, M., & Paolini, A. (2000). Subjective age estimation of telephonic voices. *Speech Communication, 31*, 107-112.

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters, 405*, 164-167.

Charlton, B. D., Taylor A.M., & Reby, D. (2013). Are men better than women at acoustic size judgements? *Biology Letters, 9,* 1-5.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research, 352*, 49-69.

Collins, S. (2000). Men's voices and women's choices. *Animal Behaviour* 60. 773–780.

Cross, J. B., DeVaney, T., & Jones, G. (2001). Pre-service teacher attitudes toward differing dialects. *Linguistics and Education, 12*(4), 211-227.

Deguchi, Ch., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research, 1455*, 75-89.

Deme, A. (2017). The identification of high-pitched sung vowels in sense and nonsense words by professional singers and untrained listeners. *Journal of Voice, 31*(2), 252.e1-252.e14.

van Dommelen, W. A., & Moxness, Bente H. (1995). Acoustic parameters in speaker height and weight identification: Sex-specific behavior. *Language and Speech, 38*(3), 267-287.

Drager, K. 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass, 4*(7), 473-480.

Eadie, T., Van Boven, L., Stubbs, K., & Giannini, E. (2010). The effect of musical background on judgments of dysphonia. *Journal of Voice, 24*(1). 93-101.

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex, 22,* 650-658.

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,* 5-8.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, M. D., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour, 69,* 561-568.

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE, 10*(9). e0138715. doi:10.1371/journal.pone.0138715

Fraccaro, P. J., O'Connor, J. J. M., Re, D. E., Jones, B. C., DeBruine, L. M., & Feinberg, D. R. (2013). Faking it: deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour, 85,* 127-136.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex, 23,* 2038-2043. doi:10.1093/cercor/bhs180

Gelfer, M. P., & Bennett, Q. E. (2014). Speaking fundamental frequency and vowel formant frequencies: effects on perception of gender. *Journal of Voice, 27*(5), 556-566.

Gocsál, Á. (1998). Életkorbecslés a beszélő hangja alapján [Age estimation based on the speaker's voice]. *Beszédkutatás '98,* 122-134.

Gocsál, Á. (2017). Az artikulációs tempó és az átlagos alaphang szerepe a beszélő életkorának megbecslésében [Age estimation based on the mean f0]. *Beszédkutatás 2017,* 151-168.

Gósy, M. 2001. A testalkat és az életkor becslése a beszéd alapján [Estimation of body configuration and age based on speech]. *Magyar Nyelvőr, 125,* 137-148.

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, darabase, research] (pp. 9-24). Budapest: Akadémiai Kiadó.

Harnsberger, J. D., Shrivastav, R., Brown Jr., W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice, 22*(1), 58-69.

Hausen, M., Torppa, R., Salmela, V. R., Vainop, M., & Särkämö, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology, 4*(566). 1-16. doi: 10.3389/fpsyg.2013.00566

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), 290-304.

Hughes, S. M., & Harrison, M. A. (2013). I like my voice better: Self-enhancement bias in perceptions of voice attractiveness. *Perception, 42,* 941-949.

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

Huntley, R., Hollien, H. & Shipp, T. (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice,* 1(1), 49-52.

Klofstad, C. A., Anderson, R. C., & Nowicki, S. (2015). Perceptions of competence, strength, and age influence voters to select leaders with lower-pitched voices. *PLoS ONE* 10(8): e0133779.doi:10.1371/journal.pone.0133779 (Retrieved 15.03.2018).

Krauss, R., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology, 38*, 618-625.

Krepsz, V., & Gósy, M. (2016). A hangzásidő és a megakadásjelenségek hatása az életkorbecslésre [The effect of the duration of sounding and the disfluencies on the estimation of the speaker's age.]. In G. Balázs, & Á. Veszelszki (Eds.), *Generációk nyelve. Tanulmánykötet.* [Languages of Generations. Book of selected papers] (49–62). ELTE BTK Mai Magyar Nyelvi Tanszék, Inter Nonnprofit Kft. – MSZT, Budapest.

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia, 51,* 1608-1618.

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

Liu F., Jiang, C., Wang, B., Xu, Y., & Patel, A. D. (2015). A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia, 66,* 111-118.

Madikal Vasuki, P. R., Sharma, M., & Demuth, J. A. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research, 342,* 112-123.

Meha-Bettison, K., Sharma, M., Ibrahim, R. K., & Vasuki P. R. M. (2018). Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians. *International Journal of Audiology, 57*(1), 40-52.

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

Mikutta, C. A., Maissen, G., Altorfer, A., Strik, W., & Koenig, T. (2014). Professional musicians listen differently to music. *Neuroscience, 268,* 102-111.

Moyse, E. (2014). Age Estimation from Faces and Voices: A Review. *Psychologica Belgica, 54*(3), 255-265.

Pabery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience, 29*(45), 14100-14107.

Pabery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia, 49,* 3338-3345.

Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews, 35,* 2140-2154.

Park, M., Gutyrchik, E., Bao, Y., Zaytseva, Y., Carl, P., Welker, L., Pöppel, E., Reiser, M., Blautzik, J., & Meindl, T. (2014). Differences between musicians and non-musicians in neuro-affective processing of sadness and fear expressed in music. *Neuroscience Letters, 566,* 120-124.

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308,* 98-108.

Peretz, I., & Vuvan, D. T. (2017). Prevalence of congenital amusia. *European Journal of Human Genetics 2017,* 1-6.

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain & Language, 140,* 24-34.

Pisanski, K., Oleszkiewicz, A., & Sorokowska, A. (2016). Can blind persons accurately assess body size from the voice? *Biology Letters, 12,* 20160063.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior,* 27, 283-296.

Puts, D. A., Doll, L. M., & Hill A. K. (2014). Sexual Selection on Human Voices. In V. Weekes-Shackelford, & T. Shackelford (Eds.), *Evolutionary Perspectives on Human Sexual Psychology and Behavior. Evolutionary Psychology* (pp. 69-86). New York, NY: Springer.

Rendall, D., Vokey, J. R., & Nemeth, C. (2007). Lifting the curtain on the wizard of Oz: Biased voice-based impressions of speaker size. *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1208-1219.

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,* 638-651.

Rubin, D., DeHart, J., & Heintzman, M. (1991). Effects of accented speech and culture-typical compliance-gaining style on subordinates' impressions of managers. *International Journal of Intercultural Relations, 15,* 267-283.

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

Shipp, Th., & Hollien, H. (1969). Perception of the aging male voice. *Journal of Speech, Language, and Hearing Research, 12,* 703-710.

Shoup-Knox, M. L., & Pipitone, R. N. (2015). Physiological changes in response to hearing female voices recorded at high fertility. *Physiology & Behavior, 139,* 386-392.

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). 1-11.

Stölten, K., & Engstrand, O. (2002). Effects of sex and age in the Arjeplog dialect: a listening test and measurements of preaspiration and VOT. *Proceedings of Fonetik, TMH-QPSR, 44*(1). 029-032. Online: http://www.speech.kth.se/prod/publications/files/qpsr/2002/2002_44_1_029-032.pdf (retrieved 07.12.2017)

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,* 29-32.

Strait, D., & Kraus, N. (2011). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music Perception, 29*(2), 133-146.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., & Patel, A. D. (2015). Musical training, individual, differences and the cocktail party problem. *Scientific Reports,* 5. Article number: 11628. doi:10.1038/srep11628

Tatár, Z. (2013). Beszélőprofil-alkotás lehetőségei a kriminalisztikai fonetikában [Possibilities of speaker profiling in forensic phonetics]. *Alkalmazott Nyelvtudomány, XIII*(1-2), 121-130.

Tervianemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*(1), 1-10.

Thompson W. F., Scellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion, 4*(1), 46-64.

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33,* 210-216.

Tóth, A. (2014). Gyermekek nemének és életkorának meghatározása a beszédük alapján [Estimation of children's age and gender based on their speech]. *Beszédkutatás 2014,* 98-111.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics, 38,* 654-662.

Winkler, R., Brückl, M., & Sendlmeier, W. F. (2003). The aging voice: an acoustic, electroglottographic and perceptive analysis of male and female voices In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2869-2872). 3–9 August 2003. Barcelona.

Winkler, R. (2007). Influences of pitch and speech rate on the perception of age from voice. In *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. 1849-1852). Saarbrücken 6-10 August 2007.