Evidence for an own-age bias in age estimation from voices in older persons

Evelyne Moyse, Aline Beaufort and Serge Brédart

(University of Liège, Belgium)

Address for correspondence:

Serge Brédart

Department of Psychology: Cognition and Behaviour

University of Liège

B-4000 Liège

Belgium

Email: serge.bredart@ulg.ac.be

Tel.: +32[0]4 3662015; fax : +32[0]4 3662859

Abstract

Previous studies have investigated the effect of ageing on age estimation from faces as well as the occurrence of an own-age bias in such age estimation from faces. To the best of our knowledge, the occurrence of an own age effect on age estimation from voices has never been examined earlier by using an experimental design in which the age of participants (young *vs* old) and the age of voice stimuli (young *vs* old) were crossed. Results revealed an own-age bias in older adults only. In comparison with younger adults, older participants showed age estimation abilities that are preserved for voices from their own age group and impaired for younger voices. This own age bias was absent in younger participants.

Keywords:

Age estimation, voice, ageing, own-age effect.

Introduction

The ability to estimate the age of a person from her voice is useful in our daily lives but also in the context of police testimony. There are numerous situations in which a witness does not see a target person's face, for instance, when the perpetrator is hooded or the aggression takes place at night. Anonymous telephone calls concerning the planting of bombs, denunciation of unlawful actions or phone harassment are other examples. Therefore the question frequently arises whether we are really accurate when we estimate age from voices.

A number of studies have investigated human capacity for age estimation from voices. Generally results were consistent and indicated that listeners were fairly accurate at storing a voice into an age category (Cerrato, Falcone & Paolini, 2000; Ptacek & Sander, 1966). When the participants' task was to assign a precise age to each voice, an age underestimation was globally observed (Amilon, Van de Weijer & Schötz, 2007; Hartman, 1979) but studies using voices from an age continuum showed that the age of young adults' voices was overestimated, whereas the age of older adults' voices was strongly underestimated (Cerrato et al., 2000; Harnsberger, Brown, Shrivastav & Rothman, 2010; Hughes & Rhodes, 2010; Schötz, 2005; Shipp & Hollien, 1969).

Few studies evaluated whether the accuracy of age estimation from voices was affected by ageing. Huntley, Hollien and Shipp (1987) reported that the difference between perceived age of voices and their chronological age was larger in older participants (aged between 60 and 84 years) than in younger adults (ages between 20 and 30 years). In contrast, Hughes and Rhodes (2010) reported no effect of the age of participants on age estimation from voices but their older participants belonged to a "46 and over" age category that presumably included mainly middle-aged rather than older persons. Given the scarcity of available data, in the present study, the age estimation performance of young and old adults will be compared in order to assess the effect of ageing on age estimation from voices using both measures of error magnitude and error direction (i.e., overestimation *vs* underestimation).

An own-age bias in age estimation from faces has been demonstrated in several studies. Indeed participants are more accurate when estimating faces belonging to their own-age group compared with older or younger faces (George & Hole, 1995; Moyse & Brédart, 2012; Voelkle, Ebner, Lindenberger, & Riediger, 2012). Given the similarities between faces and voices (Belin, Bestelmeyer, Latinus & Watson, 2011), it makes sense to predict that an own-age bias might also affect age estimation from voices. A second aim of the present study was to evaluate the prediction that age estimation errors would be smaller for voices of people belonging to the participants' own age group than for voices from other age groups. In order to do so, we used an experimental design in which age and gender of participants and voice stimuli were crossed.

Before starting to describe the present empirical study, it is important to consider the issue of the measure used to determine the accuracy of age estimation. A commonly used measure in studies of age estimation from voices is the correlation between chronological age and perceived age (e.g. Braun & Cerrato, 1999; Cerrato et al., 2000; Hartman, 1979; Huntley et al., 1987; Krauss, Freyberg & Morsella, 2002; Shipp & Hollien, 1969). However as already pointed out by Braun (1996) and Braun and Cerrato (1999), correlations cannot reveal any systematic errors. For example, a constant error of 10 years for each voice will result in a high correlation and therefore lead to an erroneous conclusion of good performance. In addition, such a measure would not be well-suited to the factorial design adopted in the present study. In other studies of age estimation, a participant's performance is measured by the mean value of the difference between perceived age and chronological age for a given set of stimuli. Using such a dependent measure may attenuate or even fail to reveal the actual degree of estimation error. Indeed, when average error scores are used, the occurrence of a large age overestimation may be cancelled out by the occurrence of an equally large underestimation. Consequently, the average of deviations above and below the actual voice's age may be mistaken with perfect age estimation. One way to deal with this problem is to average the absolute difference between each stimulus' perceived age and chronological age (Amilon et al., 2007; Braun & Cerrato, 1999; Dehon & Brédart, 2001; see also Rhodes, 2009). Nevertheless, the average signed values are usually calculated too, because they are informative with respect to the direction of the error; i.e., whether errors are mainly over- or underestimations (Amilon et al., 2007; Dehon & Brédart, 2001; Rhodes, 2009; Voelkle et al., 2012). Unfortunately, none of the measures discussed above is very effective for comparing error magnitude for voices of different ages. Indeed, making an error of 5 years in estimating the age of a 20 year-old person is a stronger error than making the same 5 years errors in estimating the age of a 70 year-old person. Then, a solution might simply consist in calculating *a percentage of error*; for instance the former error corresponds to 25% (5/20) of the chronological age while the latter corresponds only to 7.14% (5/70) of the chronological age. In the present study, the percentage of error will be used as a measure of the magnitude of the error while the mean value of this participant's estimation errors will be used as a measure of the direction of error.

Method

Participants

Sixty volunteers took part in the study. All participants were Caucasian French native speakers and were randomly recruited in the population. Participants were 30 young adults aged between 20 and 30 years (15 females, mean age = 24.2 years; *SD* = 2.7) and 30 older adults aged between 65 and 75 years (15 females, mean age = 69.1 years; *SD* = 3.3). They were active, healthy and reported no mental health problems. All participants filled out a socio-economic questionnaire. There was no significant difference between the average socioeconomic level of the old participants and the young participants (Mann-Whitney U = 421, p = 0.83). All participants reported having a normal or corrected-to-normal audition.

Stimuli

Thirty-two spoken extracts were used as stimuli. Extracts were made from 8 different young (between 20 and 30 years) female adult voices (mean age: 23.6 years, SD = 2.7), 8 young male voices (mean age: 23.6 years, SD = 2.6), 8 old (between 65 and 75 years) female voices (mean age: 68.8 years, SD = 3.2), and 8 old male voices (mean age: 69.6 years, SD = 2.9). All the speakers were French-speaking Belgians who had never smoked and were not suffering from a cold or a disease affecting their voices. Speakers were recorded individually in a quiet room using a digital recorder (Olympus, WS-450S). Seventy-two spoken extracts were recorded and were transferred to a computer (sampling rate = 44.10 kHz, quantization = 32 bits). Each extract lasted 7 seconds during which time each speaker read the first article of the Declaration of Human Rights. Participants had the opportunity to re-familiarize with the text as long as they wished before being recorded. The number of words varied slightly across the extracts, but each started with the first word of the Declaration. A pilot study, carried out on twenty participants (10 females, mean age = 20.8 years, SD = 2.3), allowed us to select our thirty-two stimuli (each stimulus being a 7-sec long extract) on the basis of both the quality of the recording and the emotional valence of the voices. The selected stimuli were judged emotionally neutral (i.e., on a scale ranging from 1, meaning "very negative", to 7, meaning "very positive"; Young Female: M = 4.10, SD = 0.43; Young Male: M = 4.12, SD = 0.34; Older Female: M = 4.21, SD = 0.31; Older Male: M = 4.02, SD = 0.46) and with a good quality of sounds (i.e., on a scale ranging from 1, meaning "very poor quality", to 7, meaning "very good quality"; Young Female: *M* = 4.84, *SD* = 0.46; Young Male: *M* = 5.19, *SD* = 0.45; Older Female: *M* = 4.62, *SD* = 0.40; Older Male: *M* = 5.06, *SD* = 0.62). Finally, the pilot study's participants were asked to sort the spoken extracts into seven age categories (i.e., "Less than 10 years"; "10-15 years"; "20-30 years"; "35-45 years"; "50-60 years"; "65-75 years" and "80 years and over"). Each extract obtained an averaged categorization computed from the 20 participants. In the four categories of stimuli (Young Female, Young Male, Older Female, Older Male), a Z-score was thus assigned to each record. Stimuli were then selected so that the averaged Z-score of each of the four categories of stimuli was equivalent (Young Female: *M* = -0.21, *SD* = 0.94; Young Male: *M* = -0.20, *SD* = 0.72; Older Female: *M* = -0.20, *SD* = 0.79; Older Male: *M* = -0.19, *SD* = 0.66). Thus, spoken extracts of a category were not more or less typical of their age group in comparison with other categories.

Acoustic analysis

All recordings were imported into Praat speech analysis software (Boersma & Weenink, 2005) so that certain acoustic properties of extracts could be analysed. Using Praat, the fundamental frequency and the intensity of each group of stimuli were measured (see Table 1).

INSERT TABLE 1 ABOUT HERE

A 2 (Age: Young vs. Older Voices) x 2 (Gender: Female vs. Male Voices) ANOVA was carried out on the intensity of stimuli. This analysis revealed no significant main effect of

Age, $F(1, 28) = 1.57$, $p = 0.22$, Gender, $F(1, 28) = 0.04$, $p = 0.84$, or interaction, $F(1, 28) = 0.87$, $p = 0.36$.

A 2 (Age: Young vs. Older Voices) x 2 (Gender: Female vs. Male Voices) ANOVA on mean fundamental frequency showed no significant main effect of Age, $F(1, 28) = 1.71$, $p = 0.20$, and no significant interaction, $F(1, 28) = 1.31$, $p = 0.26$. However, there was a significant main effect of Gender, $F(1, 28) = 141.43$, $p < 0.01$, $\eta^2 p = 0.83$. The fundamental frequency was higher for female voices ($M = 223.89$, $SD = 19.34$) than for male voices ($M = 136.43$, $SD = 22.79$).

Finally, a 2 (Age: Young vs. Older Voices) x 2 (Gender: Female vs.Male Voices) ANOVA was carried out on the standard deviation of fundamental frequency. This analysis revealed no significant main effect of Age, $F(1, 28) = 0.11$, $p = 0.74$), Gender, $F(1, 28) = 0.12$, $p = 0.73$) or interaction, $F(1, 28) = 1.50$, $p = 0.23$.

In addition, the speech rate was measured by counting the number of words in each extract. A 2 (Age: Young vs. Older Voices) x 2 (Gender: Female vs. Male Voices) ANOVA on the number of words revealed no significant main effect of Gender Group, $F(1, 28) = 0.43$, $p = 0.51$, and no significant interaction between the two factors, $F(1, 28) = 0.05$, $p = 0.83$). However, a significant main effect of Age Group was observed, $F(1, 28) = 9.90$, $p < 0.01$, $\eta^2 p = 0.26$. The speech rate was lower for older voices ($M = 20.19$, $SD = 1.94$) than for younger voices ($M = 22.87$, $SD = 2.70$).

Procedure

Each participant was tested individually in a quiet room. The 32 extracts were presented through a headphone with the same intensity to all participants. The E-Prime 1.0 Software was used to present stimuli in a random order. Each trial started with a fixation cross that appeared during 1000 msec on a computer screen; this cross was replaced by a dot for the duration of the extract (7 sec). The experimenter monitored the presentation of stimuli by clicking the spacebar on the computer keyboard and noted each participant's response. At the beginning of the experimental session, the participants were informed that they were going to hear male and female voices of different ages. They were instructed to evaluate the age of each presented voice as accurately as possible by giving an entire number (for instance 22 years, neither 22 and half years nor about twenty), to respond at their own pace, and to give their response orally. The experimenter also pointed out that it was possible that some age ranges would not be present and that, conversely, some voices could be from people of the same age. Each spoken extract could be heard only once. On average the experiment lasted about fifteen minutes.

Results

A mixed design was used, with one between-participants factor (Age group, two levels: Young and Older adults) and one within-participants factor (Age of voices, two levels: Young and Older adults). Two dependent variables were used. First, the percentage of error was calculated, i.e. the deviation from the chronological age was divided by the chronological age. Estimating that a 20 year-old has 15 years or 25 years leads in both cases to a 25% (5/20) error. Then, by participant, the percentage of error was calculated for each stimulus category (Young and Older). The second dependent variable was the signed value of the difference between perceived age and chronological age. In our example, the first errors correspond to a signed value of – 5 (underestimation of 5 years) and the second one to a signed value of +5 (overestimation of 5 years).

Percentage of error (error magnitude)

A 2 (Age group) X 2 (Age of voices) ANOVA with repeated measures on the last factor was carried out on the percentage of error. The analysis revealed a significant main effect of Age group, $F(1, 58) = 21.60$, $p < 0.01$, $\eta^2p = 0.27$. Percentages of error were lower in young ($M = 21.90$, $SD = 6.20$) than in older participants ($M = 28.40$, $SD = 13.55$). This ANOVA also showed a significant main effect of Age of voices, $F(1, 58) = 19.06$, $p < 0.01$, $\eta^2p = 0.25$. Percentages of error were higher for young ($M = 29.18$, $SD = 13.11$) than for older voices ($M = 21.13$, $SD = 6.16$). In addition, this analysis revealed a significant interaction between Age group and Age of voices, $F(1,58) = 17.59$, $p < 0.01$, $\eta^2p = 0.23$.

INSERT TABLE 2 ABOUT HERE

Post-hoc HSD Tukey tests showed that older participants had a lower performance than young participants when estimating the age of young voices ($p < 0.01$). However, there was no significant difference between age groups when the age of older voices was estimated. Planned comparisons indicated that performance of older participants was better when estimating older voices compared with young voices ($p < 0.05$). However, performance of young participants was not significantly different when estimating young and older voices. Descriptive data are presented in Table 2. Speech rate did not correlate with the percentage of errors in any age group for neither younger nor older voices, all *p*s > .20.

Data were also analysed taking the items rather the participants as the random factor. This item analysis revealed the same pattern as the preceding analysis. A 2 (Age of voices) X 2 (Age group) ANOVA with repeated measures on the last factor was carried out on the percentage of error. The analysis revealed a significant main effect of Age group, $F(1, 30) = 16.12$, $p < 0.01$, $\eta^2p = 0.35$. Percentages of error were lower in young ($M = 21.90$, $SD = 7.46$) than in older participants ($M = 28.40$, $SD = 15.15$). This ANOVA also showed a significant main effect of Age of voices, $F(1, 30) = 5.56$, $p < 0.05$, $\eta^2p = 0.16$. Percentages of error were higher for young ($M = 29.18$, $SD = 14.86$) than for older voices ($M = 21.13$, $SD = 7.26$). In addition, this analysis revealed a significant interaction between Age group and Age of voices, $F(1,30) = 22.82$, $p < 0.01$, $\eta^2 p = 0.43$. Post-hoc Tukey tests showed that older participants had a lower performance than young participants when estimating the age of young voices. However, there was no significant difference between age groups when the age of older voices was estimated. Performance of older participants was better when estimating older voices compared with young voices. However, performance of young participants was not significantly different when estimating young and older voices.

In order to facilitate comparison with previous works, a mixed 2 (Age group) X 2 (Age of voices) ANOVA with repeated measures on the second factor was carried out on the absolute errors. The analysis revealed a significant main effect of Age group, $F(1,58) = 4.98$, $p = 0.03$, $\eta^2_p = 0.08$. Error scores were higher in older ($M = 11.37$, $SD = 4.77$) than in young participants ($M = 10.14$, $SD = 6.00$). This ANOVA also showed a main effect of Age of voices, $F(1,58) = 120.96$, $p < 0.01$, $\eta^2_p = 0.68$. Error scores were higher for older adults' voices ($M = 14.74$, $SD = 4.29$) than for young adults' voices ($M = 6.77$, $SD = 2.97$). Finally, this ANOVA revealed a significant interaction between the two factors, $F(1,58) = 8.21$, $p < 0.01$, $\eta^2_p = 0.12$. Post-hoc HSD Tukey tests showed that young participants made smaller ($M = 5.13$, $SD = 1.40$) absolute errors than older participants ($M = 8.42$, $SD = 3.23$) when estimating the age of young voices, $p < 0.01$. However, there was no significant difference between age groups when the age of older voices was estimated. Speech rate did not correlate with the size of absolute errors in any age group for neither younger nor older voices, all *p*s > .30.

Signed errors

A 2 (Age group) X 2 (Age of voices) ANOVA with repeated measures on the last factor was conducted on signed errors. The analysis revealed a significant main effect of Age group, $F(1, 58) = 7.77$, $p < 0.01$, $\eta^2p = 0.12$. This analysis also showed a significant main effect of Age of voices, $F(1, 58) = 395.08$, $p < 0.0001$, $\eta^2p = 0.87$, and a significant interaction between Age group and Age of voices, $F(1,58) = 8.13$, $p < 0.01$, $\eta^2p = 0.12$. Post-hoc HSD Tukey tests showed that there was no significant difference between age groups when the age of older voices was estimated but the two groups significantly differed when estimating the age of young voices ($p < 0.01$). In order to check whether signed errors were over- or underestimation errors, the distribution of signed errors from each cell of the design was separately compared with a theoretical value of zero. The mean of the distribution of signed errors did not differ significantly from zero when young participants estimated young voices, $t(29) = 1.79$, but older participants overestimated the age of young voices, $t(29) = 6.31$, Cohen's $d = 1.15$. Both young and older participants underestimated the age of older voices, respectively $t(29) = -11.66$, Cohen's $d = 2.13$ and $t(29) = -11.08$, Cohen's $d = 2.02$. Descriptive data are presented in Table 3. Speech rate correlated with the size of signed errors when young participants estimated young voices, Pearson $r = 0.55$, $p < .05$, but did not correlate when young participants estimated older voices, nor when older participants estimated either younger or older voices, all *p*s > .20.

An analysis taking items as the random factor showed the same pattern of results. A 2 (Age of voices) X 2 (Age group) ANOVA with repeated measures on the last factor was conducted on signed errors. The analysis revealed a significant main effect of Age group, $F(1, 30) = 24.52$, $p < 0.01$, $\eta^2p = 0.45$. This analysis also showed a significant main effect of Age of voices, $F(1, 30) = 38.48$, $p < 0.0001$, $\eta^2p = 0.46$, and a significant interaction between Age group and Age of voices, $F(1,58) = 14.71$, $p < 0.01$, $\eta^2p = 0.33$. Post-hoc HSD Tukey tests showed that there was no significant difference between age groups when the age of older voices was estimated but the two groups significantly differed when estimating the age of young voices ($p < 0.01$). In order to check whether signed errors were over- or underestimation errors, the distribution of signed errors from each cell of the design was separately compared with a theoretical value of zero. The mean of the distribution of signed errors did not differ significantly from zero when young participants estimated young voices, $t(15) = 0.75$, but older participants overestimated the age of young voices, $t(15) = 4.35$, Cohen's $d = 1.09$. Both young and older participants underestimated the age of older voices, respectively $t(15) = -5.53$, Cohen's $d = 1.38$ and $t(15) = -5.19$, Cohen's $d = 1.30$.

The combination of two kinds of analysis (ANOVAs and comparisons with a theoretical value of 0) allowed establishing the following pattern of results. The age of older voices were underestimated to the same extent both by younger and older participants. The age of younger voices were neither under- or overestimated by younger participants while they were overestimated by older participants.

INSERT TABLE 3 ABOUT HERE

Discussion

The two main objectives of this experiment were to evaluate whether ageing impacted the ability of age estimation from voices, and to assess the occurrence of an own-age bias in an age estimation task from voices. Results showed a main effect of age (younger participants performed better than older participants) that was qualified by an interaction between the age of participants and the age of presented voices revealing an own-age bias. Such an own-age bias was evidenced in older adults only. Their percentage of error was significantly smaller when estimating the age of voices from their own age-group than when estimating the age of young adults' voices. Younger participants did not show this bias: the magnitude of their errors was similar for young and old voices. In comparison with young participants, older people showed preserved age estimation abilities for older voices. It is interesting to note that such a preservation of age estimation for own-age stimuli in older people was previously reported in studies evaluating age estimation from faces (Burt & Perrett, 1995; Moyse & Brédart, 2012).

The absence of own-age effect in young people when estimating the age of voice contrasts with several reports of the occurrence of an own-age effect in participants of the same age as those who were involved in the present study when age is estimated from faces (Anasasi & Rhodes, 2006; George & Hole, 1995, Moyse & Brédart, 2012; Voelke et al., 2012). This difference could be due to the fact that age estimation is more difficult from voices than from faces (see below). This difficulty of estimating the age would not allow younger people to perform better with voices from their age group than with older voices. In addition, this difficulty would create the bias in older people preventing them from performing at the same level as younger people for the less familiar kind of voices. In order to test this hypothetical interpretation, further research is necessary to evaluate whether the own-age bias is different with voices compared with faces in young participants by using both kinds of stimuli in the same experiment.

A second objective was to determine whether age estimation errors were mainly overestimation or underestimation errors. In accordance with previous studies (Hughes & Rhodes, 2010; Schötz, 2005; Shipp & Hollien, 1969), we found a large underestimation of older adults' voices by young participants. In addition, like Braun (1996), we found that older participants overestimated the age of young adults' voices. This pattern of results has previously been explained by a tendency to adjust our age estimates to our own age (Braun, 1996; Hartman, 1979; Huntley et al., 1987; Shipp & Hollien, 1969). However, the fact that, in the present study, older participants underestimated older adults' voices to the same extent as did young participants does not fit this explanation. It is possible that the design used in the present study impacted the over- and underestimation of voices. Because of our main objective was to assess the occurrence of an own age bias, we used an experimental design that was successfully tested in previous studies with faces: the age of stimuli was matched and crossed with the age of participants, the two categories of age being clearly separate. During the experiment, we observed that participants tended to estimate the age of voices according to a continuum, filling the gap between the two age categories. This could explain, in part, the underestimation of older voices and the overestimation of younger voices observed in this study.

Globally, although age estimation is fairly good in the sense that the average percentage error is between 20% and 36% across age groups and conditions (age of voice), it is far from perfect since these values correspond to absolute errors of 14.4 and 15.2 years for older voices estimated by older or young adults respectively, 5.5 years for young voices estimated by young adults and 9.1 years for young voices estimated by older participants, with an overall average error of 10.8 years. This overall average error is quite similar to the 9.7 years value reported by Amilon et al. (2007). These values are clearly higher than errors observed in studies of age estimation from faces using the same experimental designs, i.e. 5.7 years in the Amilon et al. study and 5.3 years in the Moyse and Brédart study.

Harnsberger et al. (2010) showed that speech rate is a clue to estimating the age of voices. The acoustic analysis of stimuli in the present study showed a significant difference between young and older voices in speech rate. Therefore we could also suppose that speech rate is useful in age estimation from voices. However contrary to Hollien and Shipp (1972), we observed no difference in the variability of F0 in the two age groups. In addition, correlational analyses indicated that speech rate did not correlate with age estimation but one single exception. Taken together, these results suggest that speech rate was not crucially involved in the performance of age estimation in the present experiment.

In conclusion, the main result of the present study is that older people are more likely to make larger errors than younger people when estimating the age of young voices, but not when estimating the age of voices from their own age group. In other words, there is no general effect of ageing on the performance of age estimation from voices. Older people's ability to estimate age from voices seems to be preserved for voices from their own age group. This result provides some indications relative to the conditions of reliability of older people's ear-witness testimony.

References

Amilon, K., Van de Weijer, J., & Schötz, S. (2007). The impact of visual and auditory cues in age estimation. In C. Müller (Ed.), *Speaker Classification II. Lectures Notes in Artificial Intelligence.* Springer-Verlag, Berlin, 10-21. doi: 10.1007/978-3-540-74122-0_2

Anastasi, J.S., & Rhodes, M.G. (2006). Evidence for an own-age bias in face recognition. *North American Journal of Psychology, 8,* 237-252.

Belin, P., Bestelmeyer, P.E.G., Latinus, M., & Watson, R. (2011). Understanding voice perception, *British Journal of Psychology, 102*, 711-725.

Boersma, P., & Weenink, D. (2005). *Praat: doing phonetics by computer (Version 5.2.18)* [Computer program]. Retrieved from http://www.praat.org.

Braun, A. (1996). Age estimation by different listener groups. *Forensic Linguistics, 3*, 65-73.

Braun, A., & Cerrato, L. (1999). Estimating speaker age across languages. *Proceedings of ICPhS 99*, San Francisco, CA, 1369-1372.

Burt, D.M., & Perrett, D.I. (1995). Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London: Series B, 259*, 137-143. doi: 10.1098/rspb.1995.0021

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

Dehon, H., & Brédart, S. (2001). An “other-race” effect in age estimation from faces. *Perception, 30*, 1107-1113. doi: 10.1068/p3122

George, P.A., & Hole, G.J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception, 24*(9), 1059-1073. doi:10.1068/p241059

Harnsberger, J.D., Brown, W.S., Shrivastav, R., & Rothman, H. (2010). Noise and tremor in the perception of vocal aging in males. *Journal of Voice, 24*, 523-530. doi: 10.1016/j.jvoice.2009.01.003

Hartman, D.E. (1979). The perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders, 12*, 53-61. doi: 10.1016/0021-9924(79)90021-2

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research, 15,* 155-159.

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*, 290-304.

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

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

Moyse, E., & Brédart, S. (2012). An own-age bias in age estimation of faces. *European Review of Applied Psychology, 62*, 3-7. doi:10.1016/j.erap.2011.12.002

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

Rhodes, M.G. (2009). Age estimation of faces: a review. *Applied Cognitive Psychology, 23*, 1-12.

Schötz, S. (2005) Effects of stimulus duration and type on perception of female and male speaker age. *Proceedings*, FONETIK, 2002. Department of Linguistics, Gothenburg University.

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

Voelkle, M. C., Ebner, N. C., Lindenberger, U., & Riediger, M. (2012). Let me guess how old you are: Effects of age, gender, and facial expression on perceptions of age. *Psychology and Aging, 27,* 265-277.

Acknowledgment

AB is funded by the Belgian National Fund of Scientific Research (F.R.S.-FNRS) as an F.R.S.-FNRS Research Fellow.

The final publication is available at http://link.springer.com

	Mean of F_0 (Hz)	Standard deviation of F_0 (Hz)	Intensity (dB)	Speech rate
Young Female	223.23 (14.16)	33.26 (11.45)	65.42 (4.71)	22.5 (2.51)
Young Male	127.43 (13.48)	43.95 (24.17)	64.33 (4.09)	23.25 (3.01)
Older Female	224.44 (24.50)	43.90 (20.47)	62.19 (4.51)	20 (2.39)
Older Male	145.44 (27.31)	37.89 (18.84)	63.86 (3.20)	20.37 (1.51)

Table 1. Acoustic analysis of stimuli. Standard deviations are given in parentheses.

Age of voices	Young (20-30)	Older (65-75)
Age group		
Young (20-30)	22.06 (6.12)	21.75 (6.29)
Older (65-75)	36.30 (14.40)	20.51 (6.08)

Table 2. Mean percentage of error as a function of the age of participants and the age of voices. Standard deviations are given in parentheses.

Age of voices	Young (20-30)	Older (65-75)
Age group		
Young (20-30)	0.72 (2.21)	-12.18 (5.72)
Older (65-75)	5.67 (4.93)	-11.55 (5.71)

Table 3. Mean directional errors as a function of the age of participants and the age of voices. A negatively signed mean error of -12.18 indicates that the actual age of older voices has been underestimated by 12.18 years on average by young participants. Standard deviations are given in parentheses.