Investigating the effects of talker age and listener age on trait perception from adult voices

Authors: Carolyn McGettigan[1] & Nadine Lavan[2]

Affiliations: 1: Speech, Hearing and Phonetic Sciences, UCL, London UK; 2: Department of Psychology, Queen Mary University of London, London, UK
Email: 1: c.mcgettigan@ucl.ac.uk; n.lavan@ucl.ac.uk

ABSTRACT

We tested whether "first impression" judgements of talkers are affected by the age of the individuals being evaluated, the age of the listeners, and the interaction of these two factors.

We presented listeners in 3 age groups (20-34, 35-49, 50-65 years old) with the vowel [a:] produced by 80 male and female talkers aged from 20 to 65. Talkers were rated on 1-9 Likert scales for perceived trustworthiness, attractiveness, and dominance.

The results showed significant main effects of both talker age (older talkers sounded less trustworthy and less attractive) and listener age (older listeners found talkers more attractive overall), but there were no significant interactions. Within-age-group agreement and cross-age-group correlations in ratings were high. Overall, our study indicates some talker-age effects but minimal listener-age effects on trait perception across the lifespan.

Keywords: Voice, identity, age, traits, first impressions.

1. INTRODUCTION

There is considerable evidence that human listeners can make judgements of person characteristics from voices, including physical attributes (e.g. sex, age, height), traits (e.g. attractiveness, dominance), and social characteristics (e.g. educatedness, professionalism). These "first impression" judgements can be made rapidly and with high agreement [1, 2], on stimuli including vowels, words, phrases, and longer passages [3,4]. When considering trait perception, it has been found that first impressions of voice stimuli can influence not only explicit trait judgements but also decision-making and behaviours toward other individuals – for example, electoral voting choices [5].

For this study, we were interested in the extent to which voices of different ages may yield differing first impressions, as evidence of stereotyping might have important social and economic implications. Several previous studies have focused on how accurately listeners can perceive voice age itself, finding that age perception is to some degree possible [6,7]. However, age estimates are not always precise: for example, listeners tend to overestimate the ages of younger adult voices, while underestimating the ages of older adult voices. There is also some suggestion of effects of *listener age* on judgements – a number of studies have reported an "own-age" bias, where listeners are more accurate at estimating the age of voices from their own age group, versus groups older or younger than themselves (although this effect is not always replicated). See [6] for a detailed review.

Aside from studies directly examining the perception of age, few studies of other aspects of talker first impressions have measured lifespan effects on perception of other talker characteristics. One study found that young listeners associate fewer positive stereotypes with elderly voices [8], while another found that perceived older age in talkers was associated with negative stereotyping, such as being ill-natured, frail, and subdued [9]. However, we are not aware of any previous studies that have comprehensively examined the relationships between listener age and talker age in the formation of trait impressions from voices.

We therefore designed a study to test for effects of talker age, listener age, and their interaction on the formation of trait impressions. Groups of young, middle-aged, and older adults listened to vowels produced by talkers of matched age ranges and provided ratings of their perceived attractiveness, trustworthiness, and dominance (previously identified in a 3-dimensional model underlying first impressions of faces [10]). We predicted that we would find main effects of talker age group (e.g. stereotyped responses where older voices are perceived as less dominant), main effects of listener age group (e.g. positivity bias in older listeners; [11]), and significant interactions of talker age and listener age (e.g. young and older listeners trait ratings showing different trends in relation to talker age; own-age biases [12]).

2. METHOD

2.1. Participants

Listeners were recruited in 3 age groups: 20-34 years old, 35-49 years old, and 50-65 years old. After

exclusions (see section 2.4.3) a final sample of 116 participants included 39 participants aged 20-34 years old (Mean age = 26.7, s.d. = 4.3; 20 female), 39 participants aged 35-49 years old (Mean age = 41.4, s.d. = 3.5; 23 female), and 38 participants aged 50-65 years old (Mean age = 53.8, s.d. = 4.3; 27 female). Participants were recruited from Prolific (prolific.co) and paid at a rate of £7.50 per hour. All participants were native speakers of English, had normal or corrected-to-normal vision, and had no self-reported hearing difficulties. The study was approved by the local Ethics Committee.

2.2. Materials

Audio clips of the sustained vowel [aː] were retrieved from the Saarbrücken Voice Database (http://stimmdb.coli.uni-saarland.de/). We chose to study vowel perception as a first step, and to avoid presenting German connected speech samples to English-speaking listeners. Eighty clips (40 female, 40 male) were selected, covering an age range from 20 to 65. This included 24 voices (12 female) in each of the age ranges 20-34 and 35-49, and 32 voices (16 female) in the age group 50-65.

In Audacity [13], the experimental stimuli were each trimmed to 500ms with a ramp on and ramp off of 25ms. Stimuli were then root-mean-square amplitude normalised in Praat [14], and converted to mp3 format using the MediaHuman Audio Converter.

A further 5 mp3 files featured a synthetic voice saying one of the digits 1-5, for use in vigilance trials. These were root-mean-squared amplitude normed to the experimental stimuli as above.

2.3. Design and Procedure

Participants completed the experiment in the Gorilla Experiment Builder [15]. After providing informed consent, participants completed a headphone volume calibration before proceeding to the main experiment.

The experiment was divided into 3 rating blocks: Attractiveness, Trustworthiness, and Dominance. The order of the blocks was counterbalanced across participants. There were two trial types within each block. The main experimental trials involved a single presentation of a vowel sound followed by onscreen presentation of a visual rating scale from 1-9. Participants used this scale to give their trait rating for that voice, where 1 = "not at all Attractive/Trustworthy/Dominant" and 9 = "very Attractive/Trustworthy/Dominant". Although the task was self-timed, participants were instructed to give their first impression or "gut feeling" when providing a rating. The second trial type was the vigilance trial: When hearing a spoken digit, participants were instructed to respond with the matching number on the rating scale. There were 80 experimental trials (1 per voice) and 5 vigilance trials (1 per digit 1-5) per block; trial order was fully randomised within blocks.

After completing the three rating blocks, participants finally answered a debrief questionnaire, in which they could report technical issues and other observations.

2.4. Statistical Analysis

2.4.3 Data cleaning

Data were prcoessed and analysed using R in RStudio (Version 2022.07.2).

Data were inspected for quality prior to performing statistical analyses. Four participants who failed more than 20% of the vigilance trials were removed, as was 1 participant who used the same rating for >80% of trials for any rating scale.

2.4.2 Inter-rater reliability

The intra-class correlation coefficient (ICC) was used as a measure of inter-rater reliability. After [16], the ICC2k was calculated on raw ratings per trait scale and listener group, using *ICC* function within the *psych* package in RStudio.

2.4.3 Inter-group correlations

The mean trait ratings per talker were calculated by trait scale and listener age group. These values were entered into three Pearson correlations per scale to exhaustively calculate correlations across the 3 listener age groups (i.e. 20-34 vs 35-49; 35-49 vs 50-65; 20-34 vs 50-65).

2.4.4 Effects of talker age and listener age on mean trait impressions

We used generalised linear mixed models (GLMMs) to investigate the effects of listener age, talker age, and their interaction on trait ratings, using the *lme4* package in R. Listener age group (20-34, 35-49, 50-65) and talker age group (binned into categories: 20-34, 35-49, 50-65) were modelled as fixed factors, with stimulus and voice gender as random intercepts. Pairwise comparisons were computed in *emmeans*.

3. RESULTS

3.1. Inter-rater reliability

Table 1 shows the mean intra-class coefficient (ICC) values by trait scale and listener age group. ICCs

above 0.75 indicate good reliability, and values above 0.9 signal excellent reliability [17].

Table 1: Intra-class coefficients for trait ratings, by scale and listener age group.

Trait	Listener Age	ICC	
		Mean	CIs
Attractiveness	20-34	0.857	0.817 - 0.893
	35-49	0.915	0.891 - 0.936
	50-65	0.846	0.801 - 0.884
Trustworthiness	20-34	0.792	0.735 - 0.843
	35-49	0.763	0.698 - 0.821
	50-65	0.712	0.634 - 0.782
Dominance	20-34	0.864	0.826 - 0.896
	35-49	0.861	0.821 - 0.896
	50-65	0.881	0.847 - 0.910

3.2. Inter-group correlations

Table 2 shows Pearson correlation coefficients and significance values across age groups for the 3 evaluated traits. Ratings were significantly correlated across all three age groups, with little evidence that listener groups who more similar in age (20-34 vs 35-49; 35-49 vs 50-65) show more strongly correlated ratings than groups more distant in age (20-34 vs 50-65). Notably, correlations for trustworthiness are overall lower than correlations for the other traits.

Table 2: Inter-group correlations of mean trait ratings. ** indicates p < .001

Trait	Inter-Group Correlations		
	20-34 vs. 35-49	35-49 vs. 50-65	20-34 vs. 50-65
Attractiveness	0.935**	0.922**	0.894**
Trustworthiness	0.769**	0.577**	0.646**
Dominance	0.888**	0.900**	0.842**

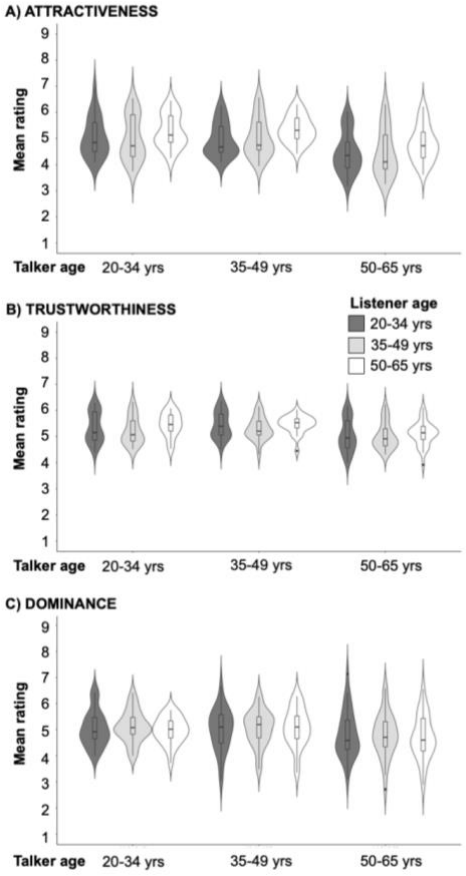
3.3. Effects of talker age and listener age on mean trait impressions

Figure 1 shows plots of mean Attractiveness, Trustworthiness, and Dominance ratings, by talker age and listener age.

3.2.1 Attractiveness

There was a significant main effect of listener age ($\chi2[2] = 71.55$, $p < .001$). The oldest listener group (50-65) gave significantly higher attractiveness ratings than the two younger groups (*p*s < .001). There was also a significant main effect of talker age ($\chi2[2] = 17.64$, $p < .001$). The oldest voices (50-65) were rated as significantly less attractive than the younger voices (*p*s < .001). There was no significant interaction between listener age group and talker age group ($\chi2[4] = 2.94$, $p = .569$).

Figure 1: Mean ratings by trait, talker age and listener age



3.2.2 Trustworthiness

There was a significant effect of listener age group ($\chi2[2] = 10.81$, $p = .005$), but within this none of the pairwise comparisons was significant (ps > .099). There was also a significant main effect of talker age ($\chi2[2]= 71.55$, $p < .001$). The oldest talkers (50-65) were rated as significantly less trustworthy than the other talker groups (ps < .003). There was no significant interaction between listener age group and talker age group ($\chi2[4] = 2.52$, $p = .641$).

3.2.3 Dominance

There were no significant effects of listener group or talker group, nor any significant interaction of the two factors (all $\chi2$s < 3.36, ps > .187).

4. DISCUSSION

We conducted a study investigating the effects of talker age, listener age, and their interaction in the formation of first impressions from voices.

Effects of talker age varied by trait category. For Attractiveness, we replicated findings from the face perception literature, which found that more attractive faces tend to have a youthful appearance [10].

For Trustworthiness, we found a significant effect of talker age, where the oldest talkers were perceived as *less* trustworthy. This trend opposes previous reports in the first impressions literature, where for example more trustworthy faces are associated with older appearance [10]. However, Sutherland's [10] face morphing approach found that more trustworthy faces were both older *and more feminine* in appearance – thus, a positive relationship between age and trust may be dependent on additional characteristics. Alternatively, the cues to ageing in voices may truly yield less positive impressions than apparent ageing in faces. Future work might resolve this issue by comparing explicit trait perceptions across faces and voices using a lifespan sample.

For Dominance, we found no effect of talker age on trait perceptions. The ICC values suggested good inter-rater agreement on dominance judgements, and exploratory inspection of the ratings broken down by talker sex also indicated that listeners reliably rated dominance higher for male than female talkers. Therefore, our interim conclusion is that dominance is perceptible in our talker sample, but impressions of this trait are unaffected by age-related stereotypes.

Attractiveness impressions may be more closely linked to physical characteristics than trustworthiness or dominance. While trustworthiness is conceptually complex and potentially context-dependent [18], attractiveness percepts from voices have been shown to be correlated with actual reproductive success [19]. The relationship between vocal age (as an index of mate fitness) and perceived attractiveness may therefore be more direct than that seen for other traits.

We found a significant effect of listener age group on ratings of talker attractiveness and trustworthiness. Specifically, listeners in the oldest age group rated the voices higher in attractiveness than the listeners in the two other age groups. This finding echoes the "positivity bias" previously reported for older adults [11]. However, this "positivity bias" did not extend to impressions of dominance or trustworthiness.

Our GLMMs revealed no significant interactions of talker and listener age, suggesting that the influence of talker age on trait perception operated similarly regardless of listener age. A set of correlation analyses further supported this conclusion: Listeners across the three listener age groups showed highly correlated talker trait ratings, on all three scales, again suggesting negligible differences in how listeners of different ages ranked the voices. However, there was some evidence of lower correlations between different age groups for trustworthiness. Trustworthiness was also the scale yielding the lowest ICC values, potentially reflecting the conceptual complexity of this trait compared with attractiveness and dominance.

These findings add valuable new evidence in relation to previous claims of listener-talker interactions in the perception of voices. As mentioned in the Introduction, the "own-age bias" reported for age perception is not always replicated. As human social groups are typically multi-generational, it is presumably adaptive for listeners of all ages to be similarly sensitive to socially-relevant information across the lifespan.

There are several possible limitations to the current study: The perception of age from voices is imperfect, thus an alternative way to examine our research questions would be to define talker age based on *perceived* age of the current voice sample. Future work should therefore explore the effects of both chronological and perceived talker age on trait perception, and in relation to listener age.

We used a single sustained vowel as the stimulus material representing each talker. Vowels have the advantage of being short and relatively steady-state samples of the talkers' vocal physiology. However, these stimuli minimise other potentially useful cues, such as speech rate and quality of articulation, which could more reliably communicate cues to physical age. While the current study showed only weak talker-age effects on dominance perception from vowels, a replication using spoken phrases may yield stronger cues to age and thus be more sensitive to detect age-related stereotyping of voices. That said, recent evidence for consistency of trait impressions across stimulus types suggests our results should be generalisable to more realistic speech [e.g. 4].

5. CONCLUSIONS

Our study is one of the first to directly examine how the ages of talkers and listeners might interact in the formation of first impressions of traits from voices.

Our findings have potential societal implications. For example, in call-centre occupations, age-related stereotyping of a voice may lead to less positive evaluations of older workers - if recognised by employers, our findings could help address inequities in the workplace. Future work measuring first impressions of age, and including more naturalistic spoken samples, will help to establish the magnitude and generalisability of these effects.

6. ACKNOWLEDGEMENTS

CM is funded by a Research Leadership Award from The Leverhulme Trust (Grant No. RL-2016-013). NL

is funded by a Sir Henry Wellcome Fellowship from Wellcome (Grant No 220448/Z/20/Z).

7. REFERENCES

[1] McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PloS ONE 9*(3), e90779.

[2] Mileva, M., & Lavan, N. (2022). Trait Impressions from Voices Are Formed Rapidly Within 400ms of Exposure. *PsyArXiv*. doi:10.31234/osf.io/zd4un

[3] Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLoS ONE* 13(10), e0204991.

[4] Groyecka-Bernard, A., *et al.* (2022). Do voice-based judgments of socially relevant speaker traits differ across speech types? *Journal of Speech, Language, and Hearing Research* 65(10), 3674-3694.

[5] 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.

[6] Moyse, E. (2014). Age estimation from faces and voices: A review. *Psychologica Belgica* 54(3), 255-266.

[7] Lavan, N. (2022). The effect of familiarity on within-person age judgements from voices. *British Journal of Psychology* 113(1), 287-299.

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

[9] Mulac, A., & Giles, H. (1996). 'You're Only As Old As You Sound': Perceived Vocal Age and Social Meanings. *Health Communication* 8(3), 199-215.

[10] Sutherland, C. A., Oldmeadow, J. A., Santos, I. M., Towler, J., Burt, D. M., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition* 127(1), 105-118.

[11] Zebrowitz, L. A., Franklin Jr, R. G., Hillman, S., & Boc, H. (2013). Older and younger adults' first impressions from faces: similar in agreement but different in positivity. *Psychology and Aging* 28(1), 202-212.

[12] Zebrowitz, L. A., & Franklin Jr, R. G. (2014). The attractiveness halo effect and the babyface stereotype in older and younger adults: Similarities, own-age accentuation, and older adult positivity effects. *Experimental Aging Research* 40(3), 375-393.

[13] Audacity Team (2022): Audacity [Computer program]. Version 2.4.2, retrieved Dec 19, 2022, from http://audacityteam.org/

[14] Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved Dec 19, 2022, from http://www.praat.org/

[15] Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52(1), 388-407.

[16] Shrout, P. E., & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin* 86(2), 420-428.

[17] Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15(2), 155-163.

[18] Knight, S., Lavan, N., Kanber, E., & McGettigan, C. (2018). The social code of speech prosody must be specific and generalizable. *Proceedings of the National Academy of Sciences* 115(27), E6103-E6103.

[19] Pisanski, K., & Feinberg, D.R. (2018) Vocal Attractiveness. In S. Frühholz, and P. Belin (eds), *The Oxford Handbook of Voice Perception* (pp.607-626). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.27, accessed 20 Dec. 2022.