Don't judge a voice by its cover: Visual interference in vocal age judgments

Jessica Cathcart
University of Victoria
jcathcar@uvic.ca

Previous studies show that individuals perceive the same face as older when preceded by young relative to old adaptor faces, and the same voice as older when preceded by young relative to old adaptor voices. However, research had not yet addressed whether these adaptation effects can occur cross-modally. Therefore, this study sought to determine whether adaptation to young or old *faces* influences the perceived age of *voices*. To do this, 20 participants ages 20-23 years were tested individually over 40 experimental trials. In each trial, participants saw either a young or old face; they then heard a voice and were asked to judge the age of the speaker. It was predicted that voices would be perceived as *older* when preceded by young adaptor faces. Results in fact showed the opposite trend: voices were consistently judged to be *younger* when preceded by young relative adaptor faces. Thus, it appears that adaptation evokes the opposite effect on age judgments when the adaptor and test stimulus differ in modality (i.e. one stimulus is visual while the other is auditory). To explain these results, it is proposed that individuals rely more heavily on visual cues than auditory ones when assessing age in their conversational partners, and that sensory cues from different modalities are unconsciously integrated even when a known incongruence exists.
Keywords: Perception; adaptation; age; face; voice

1 Introduction

Faces are arguably the most conspicuous indicator of a person's chronological age (Burt & Perrett, 1995). Although other factors are certainly prevalent, faces alone generally elicit robust and accurate age judgments (George & Hole, 2000). However, Schweinberger et al. (2010) were able to demonstrate a facial age aftereffect (FAAE) in which exposure to old adaptor faces (approximately 70 years) caused subsequently presented test faces to be perceived as younger than when the same test faces were presented after young adaptor faces (approximately 20 years). These results suggested that age judgments, though usually reliable, could be significantly influenced by adaptation, the habituation to a given stimulus after prolonged or repetitive exposure (Zaske & Schweinberger, 2011, p. 283).

Furthermore, the voice is often described as the auditory equivalent of the face (Belin et al., 2004). Even when a voice is used to produce non-speech sounds, such the cough or cry of a baby, we are able to glean valuable paralinguistic information about the vocalizer's physical and emotional state (Belin et al., 2004). Studies have also demonstrated that listeners can classify speaker age using exclusively vocal cues (Ptacek & Sander, 1966). That is, age judgments are largely accurate even in the absence of visual information. Therefore, to complement Schweinberger et al.'s 2010 study that revealed the FAAE, Zaske and Schweinberger (2011) investigated whether adaptation to young or old voices could influence listeners' perception of age in test voices. Similarly, they found that the same test voice was judged to be older when preceded by a young adaptor voice (20 years) than when preceded by an old adaptor voice (70 years). This phenomenon, deemed the vocal age aftereffect (VAAE), strongly supported their previous findings about the influence of adaptation on age estimates.

Given the assertion that faces and voices are perceptual counterparts, the adaptation effects observed in these studies should also occur cross-modally. That is, exposure to young or old faces should influence the perceived age of test voices, and exposure to young or old voices should influence the perceived age of test faces. Since research had not yet investigated the cross-modal effects of adaptation on age judgments, pairing visual and auditory stimuli was the focus of the present study. Specifically, this study sought to determine whether adaptation to young or old faces could cause listeners to over- or under-estimate age judgments of subsequently presented test voices.

Can adaptation to young or old faces influence the perceived age of test voices? It was expected that test voices would be perceived as *older* when preceded by young relative to old adaptor faces, and *younger* when preceded by old relative to young adaptor faces. These findings were predicted by analogy to Schweinberger et al.'s FAAE (2010) and Zaske and Schweinberger's VAAE (2011).

2 Methodology

The methodology for this study largely followed that of Zaske and Schweinberger (2011), but included a number of adaptations intended to simplify the research design.

2.1 Participants

Participants were 20 native English speakers: 4 male and 16 female, between the ages of 20-23. Since Zaske and Schweinberger only found a significant interaction between speaker and listener gender in their post-adaptation condition, and this study did not include a post-adaptation phase, an equal number of male and female participants was not required. All participants had

© 2015 Jessica Cathcart

normal hearing ability and normal (or corrected to normal) vision, and were recruited mainly by word of mouth.

2.2 Stimuli

Adaptor stimuli were images of 10 young (approximately 20 years) and 10 old (approximately 50 years) faces, half male and half female. 50-year-old faces were chosen over 70-year-old faces for the 'old' condition because this study used a smaller subset of voices and vocal ages than did Zaske and Schweinberger. Since the adaptor faces used by Schweinberger et al. (2010) were not publicly available, adaptor faces for this study were strategically selected from the Internet in hopes that they would be novel to all participants.

Test stimuli were voice recordings of 20 native English speakers. These speakers were approximately 30 or 40 years old, with 10 people (5 male and 5 female) falling into each age category. Speakers were recorded uttering the vowel-consonant-vowel (VCV) syllables /aba/ and /igi/. All utterances were standardized for length to eliminate speaking rate, a common cue for vocal age, as a confounding variable (Harnsberger, 2010). Other properties of speech (such as pitch and vocal quality) were preserved; this ensured that listeners still heard sufficient acoustic information to make accurate vocal age judgments.

2.3 Procedure

Participants were each tested individually on a computer in the University of Victoria phonetics lab. Instructions appeared on a computer screen, not verbally, to ensure that adaptation to the experimenter's voice would not influence the results. Each trial began with a red fixation cross for 500 ms, followed by presentation of an adaptor face for 3500 ms, and then a green fixation cross for 500 ms. The green cross signalled that a voice was about to be played. Participants then heard a voice and indicated by pressing 2, 3, 4, 5, or 6 on the computer keyboard whether they thought the speaker was 20, 30, 40, 50, or 60 years old.

For all trials, the adaptor faces and test voices were gender congruent so as to eliminate gender differences in vocal quality as a confounding variable. With the exception of gender congruency, though, face and voice pairings were randomized. Participants were explicitly told that a voice would never belong to the person in the preceding photograph, but that they should pay close attention to the faces presented nonetheless. They were also told that a question would be asked after the activity to ensure that they had carefully attended to the faces. As a final precaution, participants were instructed to raise their hand if they recognized a face or voice (from an encounter prior to this study) on any given trial.

Each face was presented twice, once with an /aba/ utterance and once with an /igi/ utterance, such that the experiment was comprised of 40 trials in total (20 speakers x two utterances). Since Zaske and Schweinberger (2011) allowed breaks after every 48 trials and this study contained only 40, there was no need for a rest period.

© 2015 Jessica Cathcart

Unlike Zaske and Schweinberger's research (2011), this study did not include a post-adaptation phase (trials in which there was a delay between presentation of the adaptor stimulus and presentation of the test stimulus). This is because the *occurrence* of cross-modal adaptation effects, not the duration for which they persisted, was the focus here. As stated above, the effect of speaker and listener gender congruence on perception was not analyzed in this study, as Zaske and Schweinberger (2011) only found significance for this in post-adaptation trials.

2.4 Data analysis

Since all participants demonstrated strong facial recall when answering the final question and none reported familiarity with experimental stimuli, no data were excluded from analysis.

Data were sorted by the factors *adaptor stimulus* (young vs. old) and *test stimulus* (30 vs. 40). Each combination of factors (young, 30; young, 40; old, 30; old, 40) was then examined to determine the average perceived age (and standard deviation) of test voices under each experimental condition.

3 Results

It was predicted that voices would be perceived as older when preceded by young adaptor faces than when preceded by old adaptor faces. However, results of this study were as follows.

30-year-old voices heard after exposure to a young adaptor face were, on average, judged to be younger (mean = 31.77 years, s.d. = 11.37 years) than 30-year-old voices heard after exposure to an old adaptor face (mean = 34.32 years, s.d. = 13.13 years). Furthermore, 40-year-old voices presented after a young adaptor face were judged to be younger (mean = 37.03 years, s.d. = 12.20 years) than 40-year-old voices played after an old adaptor face (mean = 38.74 years, s.d. = 12.35 years). These values are summarized in Table 1.

Experimental Condition	Mean	Std. Deviation
30-year-old voice, young adaptor face	31.77 years	11.37 years
30-year-old voice, old adaptor face	34.32 years	13.13 years
40-year-old voice, young adaptor face	37.03 years	12.20 years
40-year-old-voice, old adaptor face	38.74 years	12.35 years

Table 1: Means and standard deviations for perceived age of 30- and 40-year-old test voices after adaptation to young or old faces

It was therefore found that voices were consistently judged to be *younger* when preceded by young relative to old adaptor faces and *older* when preceded by old relative to young adaptor faces. Thus, it appears that adaptation has the opposite effect on age judgments when the adaptor and test stimulus differ in modality

© 2015 Jessica Cathcart

(i.e. one stimulus is visual while the other is auditory). Figure 1 depicts the trend in perceived ages of 30- and 40-year-old voices after adaptation to young (dashed line) or old (solid line) faces. It is important to note that perceived vocal age increased with actual age for both adaptation conditions, but that age judgments were consistently higher in the old adaptor face condition. It is also interesting to note that average age estimates all fell between 30 and 40 years, even for 30-year-old voices played after young adaptor faces and 40-year-old voices played after old adaptor faces.

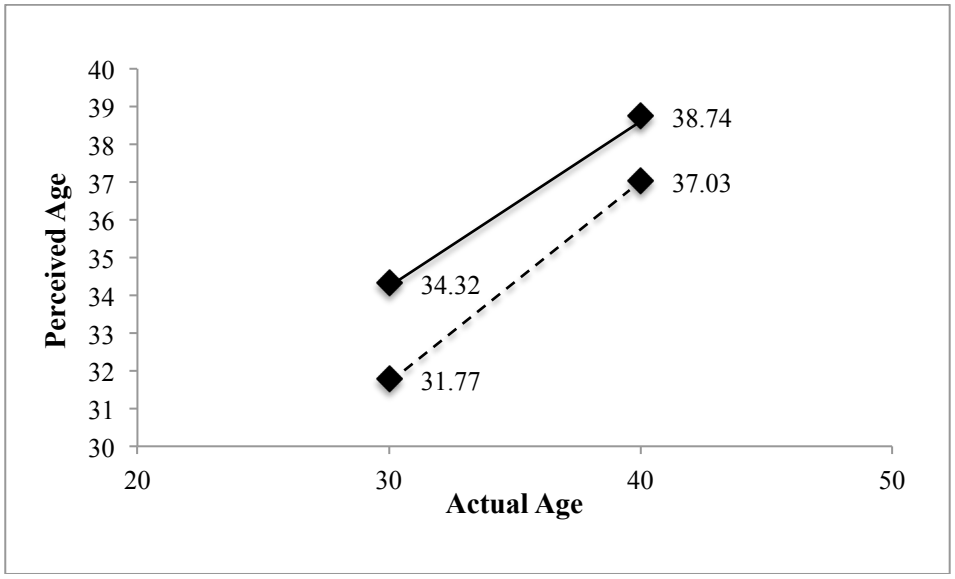

Figure 1: Average perceived ages of 30- and 40-year-old voices after adaptation to old (solid line) or young (dashed line) adaptor voices.

4 Discussion

As illustrated in Figure 1, voices were consistently judged to be *younger* when preceded by young adaptor faces and *older* when preceded by old adaptor faces. While these findings appear to contradict those of Schweinberger et al. (2010) regarding the FAAE, as well as those of Zaske and Schweinberger (2011) concerning the VAAE, there are a number of possible explanations for these results.

First, individuals generally speak to only a fraction of the people that they encounter in a day. That is, they *see* many more people than they *hear*. It is therefore likely that individuals are more practiced at assessing age based on visual cues than auditory ones, so exposure to faces should (and does) readily interfere with vocal age judgments. This is supported by the work of Schweinberger et al. (2011) who found that participants were able to recognize faces more easily than they could recognize voices. Although recognition is a distinct cognitive process from age perception, participants' preference for faces

© 2015 Jessica Cathcart

over voices suggests that faces would also be more heavily relied upon when assessing age in a conversational partner.

Second, it is likely that different strategies and mental processes are at work when judging stimuli from the same modality than when judging stimuli from different modalities. That is, while stimuli from the same modality are presumably evaluated in *contrast* to each other, stimuli from different modalities are likely evaluated in c*onjunction* with each other. In fact, research has shown that the brain quickly and efficiently integrates cues from faces and voices to increase the reliability of our sensory estimates (Campanella & Belin, 2007).

While this is certainly advantageous in most situations, it appears that this integration of visual and auditory information continues even when individuals are instructed to disregard cues from one modality. As a result, information gained from the adaptor faces in this study may have been unintentionally integrated with information from the test voices, thereby causing individuals to over- or under-estimate age in the voices that they heard. Further evidence for this theory comes from a study by Kamachi et al. (2003) which concluded that individuals link whatever visual and auditory information is available to them when constructing a unified perception of "what is being said [and] who is saying it" (p. 1709).

Finally, there were a number of limitations in the execution of this research. Not only were there fewer participants than Zaske and Schweinberger's (2011) study, but there were also far fewer stimuli. In particular, there was a much smaller range of auditory (test) stimuli. While Zaske and Schweinberger used voices that were morphed to sound 30-, 40-, 50-, and 60-years-old, this study used only 30- and 40-year-old voices. This also meant that 50-year-old faces had to be used as the "old" adaptor stimuli. It is possible that these notably younger "old" adaptor faces did not elicit as strong an adaptation effect as did the 70-year-old adaptor faces in Zaske and Schweinberger's study. What most likely accounts for the narrow range of average perceived vocal ages, though, is that the voices used as test stimuli in this study were actual recordings of approximately 30- and 40-year-old speakers rather than electronically morphed voices. Since it was not feasible to produce morphed voices for the purposes of this study, and therefore recordings of real speakers had to be used, a number of the speakers recorded were several years older than 30 or several years younger than 40. In short, some or all of the above confounds may have influenced the experimental results.

5 Conclusion

This study sought to determine whether adaptation to young or old faces influences the perceived age of subsequently presented test voices. Results in fact showed that voices were judged to be younger when preceded by young adaptor faces and older when preceded by old adaptor faces. Although contradictory to previous studies that used stimuli from a single modality, these findings were attributed to routine cognitive operations by which sensory cues from different modalities are rapidly and unconsciously integrated. It was also proposed that

© 2015 Jessica Cathcart

individuals rely more heavily on faces than voices when assessing age in the people they interact with. Furthermore, the relatively small sample size, limited vocal stimuli, and other deviations from Zaske and Schweinberger's (2011) study may have affected these results.

Future studies should seek to further our understanding of cross-modal adaptation effects by adopting a similar procedure in which voices are used as the adaptor stimuli and faces as the test stimuli. Research could also investigate the extent to which accuracy improves when participants are provided with a face, a voice, or both in making age judgments.

Although it is clear that there is much work to be done before a more complete understanding of human perception can be reached, this study was a successful contribution to the existing literature on perception of age in conversational partners.

References

Belin, P., Fecteau, S. & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129-135.

Campanella, S. & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences, 11*(12), 535-543.

Burt, D. M. & Perrett, D. I. (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and color information. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *259*, 137-143.

George, P. A. & Hole, G. J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Visual Cognition, 7*, 485–509.

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*(5), 523-530.

Kamachi, M., Hill, H., Lander, K. & Vatikiotis-Bateson, E. (2003). Putting the face to the voice: matching identity across modality. *Current Biology, 13*, 1709-1714.

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

Schweinberger, S. R., Kloth, N. & Robertson, D. M. C. (2011). Hearing facial identities: Brain correlates of face-voice integration in person identification. *Cortex, 47*(9), 1026-1037.

Schweinberger, S. R., Zaske, R., Walther, C., Golle, J., Kovacs, G. & Wiese, H. (2010). Young without plastic surgery: Perceptual adaptation to the age of female and male voices. *Vision Research, 50,* 2570-2576.

Zäske, R. & Schweinberger, S. (2011). You are only as old as you sound: Auditory aftereffects in vocal age perception. *Hearing Research, 282*(1-2), 283-288.

© 2015 Jessica Cathcart