The effect of speech rate on age estimation in conversational speech

Ka Yan Tsang, Kristen Wing Yan Wong, and Yoonjung Kang
University of Toronto Scarborough

Previous studies found that older speakers generally speak slower than younger speakers and further showed that listeners make use of this age-rate correlation in estimating speakers' age. This study examines how speakers' speech rates affect the perception of speaker age in conversational speech as well as how listeners' own age affects age perception. After hearing a short dialogue in which two speakers' speech rates were varied orthogonally, listeners estimated the age of each speaker. The results showed that listeners judged slower voices as older than faster voices and this effect was more pronounced for older speakers. We found no effect of interlocutors' speech rate, indicating that the listeners were able to reliably separate the speech rate information of the two speakers in the dialogue. We also found a significant effect of listeners' own age; other things being equal, younger listeners judged the speakers to be younger than older listeners.

1   Background

Speech contains important information that listeners use to make inferences about speakers, such as age, height, and weight (Krauss et al. 2002). Listeners rely on different vocal features such as speech rate, pitch, and articulation, as well as the linguistic content of the spoken words to estimate speakers' age in conversational speech (Hartman 1979; Moyse 2014). Among these vocal characteristics, speech rate is a feature that is consistently found to be related to age. Results from cross sectional and longitudinal studies of English speakers suggest speech rate decreases in both men and women as age increases (Bona 2014; Jacewicz et al. 2009). This speech rate to age correlation holds true across dialects of English and across languages.

Speaking rate is also found to be perceptually relevant to age estimation (Harnsberger et al. 2008; Ryan and Burk 1974). Studies that examined the effect of speech rate on age perception by manipulating speech rate while controlling for other vocal features of the speakers found that a significant shift in perceived age could result from the manipulation of speech rate alone (Harnsberger et al. 2008; Skoog Waller et al. 2015). Specifically, an increased speech rate was associated with speakers sounding younger, while a decreased speech rate was associated with speakers sounding older. This speech rate effect was found to be strongest in the age estimates of older speakers.

Some studies suggest the presence of a listener effect in vocal age estimation. Skoog Waller et al. (2015) noted that listeners in their experiment were mostly young adults and the effects of speech rate manipulation were less prominent in the young talker group. They went on to suggest that this may be due to the fact that listeners are more familiar with voices and other cues from their own age group (Moyse et al. 2014; Nagao and Kewley-Port 2005). As a result, those cues might be sufficient to provide a reliable age estimate, which in turn undermined the speech rate's significance as a cue for age estimation. It is notable that there are conflicting results on own-language bias on age estimation — i.e., whether speakers are better at estimating the age of speakers of their native language (Braun and Cerrato 1999; Nagao and Kewley-Port 2005). Moreover, listeners' gender does not appear to yield significant differences in the performance of age estimation where reported (Braun 2013; Braun and Cerrato 1999; Hughes and Rhodes 2010). In short, while the literature suggests a possibility that listeners' own characteristics (age or native

language) can affect how they use speech rate in age estimation, the evidence is inconclusive. The current study probes how listeners' own age affects their estimation of speaker age.

Another underexplored question is how the presence of interlocutors in speech affects listeners' perception of the target speakers' speech rate and estimation of their age. Existing studies that have examined the speech rate effect on age estimation all used single speaker stimuli (Harnsberger et al. 2008; Ptacek and Sander 1966; Skoog Waller et al. 2015). Since daily-oral communication usually takes the form of dialogues involving multiple speakers, the use of conversational speech as stimulus is natural-sounding and thus should better reflect how speech rate influences age estimations in real life. Furthermore, the presence of an interlocutor may introduce more factors that would affect the speech rate effect. It is of interest how the interlocutors with different speech rates would affect each other's perceived speaking rate and perceived age.

Studies on vocal age showed that when listeners were adapted to an older voice, the following voice was perceived as younger than the age estimates made by listeners adapted to a younger voice (Zäske and Schweinberger 2011). The aftereffects were reduced but still significant when the adaptation and test condition were mismatched in the gender of voice (Zäske et al. 2013). Studies on speech rate variation also found that listeners calibrated their durational perception relative to the speech rate of the contextual speech (Summerfield 1981). These results suggest that speech rate perception itself may be relative and be subject to contrastive effects, i.e., a speaker may sound faster and, by inference, younger when the interlocutor speaks slow than fast. On the other hand, Newman and Sawusch (1996) examined how exposure to multiple speakers affected the listeners' rate normalization in speech categorization. It was found that the information from multiple voices was grouped together — i.e., one speakers' speech conditioned the categorization of subsequent speech produced by another speaker. However, when the experimental condition sufficiently motivated separating different speakers' voices, listeners' categorization was affected by the source speaker's speech rate only.

The current study will examine the role of the listener and the interlocutor in age estimation. With regard to the listener effect, we aim to directly test the hypothesis that the apparent asymmetrical effect of speech rate on the perceived age of younger vs. older speakers is due to the listeners' own age bias. Specifically, we vary the age of the speakers and also recruited listeners of diverse age groups in order to provide additional insight on the interaction of listener and speaker characteristics on the use of speech rate in age estimation. With regard to the effect of the interlocutor, we explore how the interlocutor in conversational speech affects the perceived speech rate and age estimation by using conversational speech involving two speakers, where the speech rate of the two speakers are orthogonally varied.

2   Methods

2.1 Speakers

Eight professional voice actors (4 male and 4 females) were recruited from an online freelancer marketplace (www.fiverr.com) to record a scripted dialogue. The speakers were chosen to represent unambiguously younger-sounding and older-sounding voices as impressionistically judged by the authors, based on the sample recordings available for the actors.

This recruitment method was chosen as a relatively easy way to recruit older(-sounding) speakers, as well as younger(-sounding) speakers. While the perceived age generally aligned with the chronological age of the speakers, there are exceptions, such as the 42-year old female speaker (FY2) who sounded distinctively young compared to their chronological age.

While the dialogue was scripted, the voice actors enacted it to make it sound like a natural conversation, ensuring fluency and naturalness of the dialogue while maintaining complete control over the speech material across different speaker conditions. The speakers were all self-identified native speakers of North American English and were selected to represent four younger and four older-sounding speakers. They read both roles of a two-person dialogue (see 2.2) at their normal comfortable rate. Their relative speech rates were measured as a ratio of each speakers' duration of the dialogue, excluding pauses, to the average across

all speakers. As expected, the speech rates generally varied according to their (perceived) age and older speakers spoke slower than young speakers within each gender. The speaker information is summarized in Table 1.

Table 1. Speaker background information

Gender	Age group	Speaker code	Age	Relative speech rate
Female	Older	FO1	55	1.05
		FO2	58	0.97
	Younger	FY1	30	0.87
		FY2	42	0.86
Male	Older	MO1	71	1.15
		MO2	63	1.10
	Younger	MY1	29	1.07
		MY2	25	0.98

2.2 Stimuli

A 324-word dialogue between two speakers (A & B) was scripted. The script is provided in the appendix. The word count was kept to be approximately the same for the two speakers. The recordings were segmented at utterance boundaries and manipulated for speech rate in two directions, 15% faster or 15% slower relative to the overall mean of all eight speakers' natural speech rate. The conversations used in the experiments were constructed by splicing together parts from two different speakers with 250 ms of silence between speaker turns. The amount of rate change and inter-utterance gap were chosen to ensure that the resulting dialogue was both natural and distinct enough to be recognized as fast and slow. After durational manipulation, the resulting dialogue varied between 100 and 135 seconds depending on the speech rate conditions.

One male and one female speaker of the same age group were matched with each other to form a pair, creating a total of four pairs (FO1-MO1, FO2-MO2, FY1-MY1 and FY2-MY2). For each pair of speakers, eight versions of the dialogue were created such that each speaker was heard in each role (A and B) and the speech rate of the two roles were varied orthogonally to create four speech rate conditions (A&B=slow, A&B=fast, A=slow & B=fast, A=fast & B=fast). This produced a total of 32 experimental conditions (4 speaker pairs * 2 speaker roles * 4 speech rate conditions).

2.3 Procedure

Prior to starting the main experiment, participants completed a short questionnaire about their age, gender, and language background. Next, participants reported the model of the earphones or headphones they used, and were asked to adjust their device to a comfortable volume. The main experiment consisted of participants listening to a dialogue and answering questions about the speakers' age. The dialogue was accompanied with the display of a speaker icon and the phrase: "Please listen carefully". Then, participants estimated the age of the two speakers. An audio prompt of "How old am I?" in the voice and speech rate that corresponded to the experimental condition of the speaker of interest was presented. The participants responded by choosing one of the 10-year age bins (10–20, 21–30, 31–40, 41–50, and 51–60).[1] The order of the age question for each of the two speakers was always matched with the order of the speaker role in the dialogue, A and then B. Finally, a multiple-choice, content-focused screening question (e.g., "Which country is Ethan visiting?") was posed to ensure that participants had paid attention to the audio stimulus. The experiment took approximately 3–5 minutes to complete.

---

[1] The choice of 60 and over was omitted due to an experimenter error. This may have affected the responses, especially for the older speakers, but we believe that this should not affect the overall findings and conclusions of the study.

2.4 Participants

A total of 689 participants were recruited on Amazon Mechanical Turk (https://www.mturk.com/) and were paid for their participation. The data from 34 participants were excluded because the participants either did not answer the screening question correctly or they did not indicate the use of earphones in the experiment. The final sample consisted of 655 participants across the 32 experimental conditions with 19 to 25 participants per condition. Each participant heard only one version of the dialogue. 626 participants reported being native English speakers. Analyses including or excluding non-native speakers produced comparable results, except where noted, and we report the results including non-native speaker participants. Table 2 summarizes the age and gender distribution of the participants.

Table 2. Participant age and gender distribution

	<20	**<30**	**<40**	**<50**	**<60**	**60+**	**Total**
Female	8	133	139	56	47	17	400
Male	11	105	79	28	19	13	255

2.5 Statistical analysis

Statistical analyses were conducted in R (R Core Team 2016). A linear mixed-effects regression model was built using the *lmer*() function of the *lme4* package (Bates et al. 2017).The response variable was the estimated age of the speakers converted to a scale of 0 (10–20) to 4 (51–60). The fixed effects predictors in the initial models included the age of the speaker, which is also the interlocutor's age (SPEAKER_AGE: Young, Old), the speech rate of the speaker (SPEAKER_RATE: Fast, Slow), the speech rate of the interlocutor (INTERLOC_RATE: Fast, Slow), and the order of the evaluated speaker in the dialogue (ORDER: first, second). All factors were simple-coded and the reference levels are underlined above. In addition, the reported age of the participant (PART_AGE: converted to a scale of 0 (10–20) to 5 (>60) and centred) was also included as a covariate. Speaker gender and participant gender were also included in the initial model but neither of them contributed significantly to the model and hence they are not discussed further. The model included full interactions of the five predictors. The model also included random intercepts for PARTICIPANT and SPEAKER. The initial model also included a by-SPEAKER random slope for SPEAKER_RATE, which was dropped based on a likelihood ratio test. This model was pared down by backward step-wise regression using the *step()* function. As post-hoc tests, follow-up Wald Chi-square tests were done using *testInteractions*() function of the *phia* package (De Rosario-Martinez 2015).

3   Results

The statistical model output is summarized in Table 3. First, we found a significant main effect of SPEAKER_AGE in the expected direction; older speakers are judged as older (by 12.07 years) compared to younger speakers. Table 3 summarizes the average estimated age for each speaker. While the older speakers are generally judged to be older than younger speakers, on average, the age estimates are generally lower than the actual age, especially for the older speakers. Note that our speech rate manipulation eliminated speech-rate cues to the individual speakers' age. The fact that the speaker age effect was nevertheless significant indicates that besides speech rate, there were additional cues to speaker age in their speech.

Table 3. Mixed-effects model output

	Estimate	t value	Pr (> |t|)	
(Intercept)	1.361	11.718	<0.001	***
SPEAKER_AGE	1.207	7.380	<0.001	***
SPEAKER_RATE	0.100	2.181	0.029	*
ORDER	-0.044	-1.200	0.231	
PARTICIPANT_AGE	0.052	1.758	0.079	
SPEAKER_AGE * SPEAKER_RATE	0.337	5.173	<0.001	***
SPEAKER_AGE: ORDER	0.160	3.081	0.002	**
SPEAKER_AGE: PARTICIPANT_AGE	0.107	2.550	0.011	*

Table 4. Speaker background information

	Actual Age	Mean estimated age (Standard Deviation)
FO1	55	40.8 (0.9)
FO2	58	37.4 (0.9)
FY1	30	26.5 (0.6)
FY2	42	23.2 (0.6)
MO1	71	41.4 (0.8)
MO2	63	35.9 (0.8)
MY1	29	24.5 (0.6)
MY2	25	24.6 (0.6)
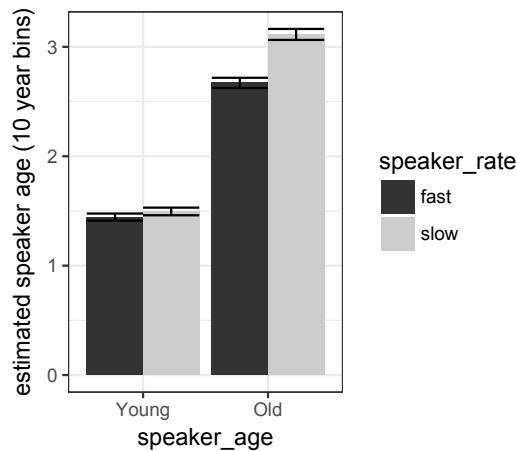
Figure 1. Estimated age by speaker age and speech rate

As for the effect of speech rate on age estimation, we found a significant main effect of SPEAKER_RATE: the slow speech rate condition was judged as older than the fast speech rate condition by a very modest amount (1.00 year). We also found a significant interaction of SPEAKER_AGE and SPEAKER_RATE and the effect of speech rate is modulated by the age of the speaker. Figure 1 illustrates this interaction. The black and grey bars that represent fast and slow speech conditions are significantly different only for the older speakers but not for the younger speakers. A post-hoc test showed that the speech rate effect was much

stronger for the older speaker group than the younger group.[2] This asymmetrical effect of speech rate on older vs. younger speakers replicates previous findings (Skoog Waller et al. 2015). However, the hypothesis that this asymmetry in the speech rate effect is due to the asymmetry in the listeners was not confirmed. In our study, this asymmetrical effect of speech rate was found even though our participants came from a wide age range. Crucially, we found no three-way interaction of PARTICIPANT_AGE (under vs. over 40) * SPEAKER_AGE * SPEAKER_RATE. Figure 2 shows the interaction of speaker age and speech rate separately for younger participants (left panel: 40 years old or younger) and older participants (right panel: above 40 years old). In other words, the differential effect of speech rate on younger and older speakers was found regardless of the age of the listeners and it was not the case that listeners will rely on speech rate less for estimating the age of speakers from their own age group.
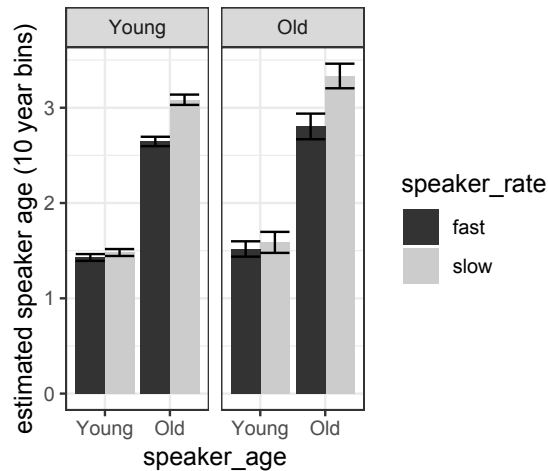
Figure 2. Estimated age by speaker age (x-axis), participant age (panel) and speech rate

As for the effect of interlocutor's speech rate, there are two possible ways that the interlocutor's speech rate could affect the speech rate perception, and hence the age estimation, of the target talker. One possibility is that the interlocutor's speech rate would provide a contrastive effect, making the target talker's speech sound faster or slower when the interlocutor's speech rate is slow or fast, respectively. The other possibility is that the speech rate of the two speakers in a conversation is perceived additively, such that fast and slow speech rates of the interlocutor will make the target talker sound faster and slower, respectively. The effect of the interlocutor speech rate is summarized in Figure 3 and we found no consistent pattern. Descriptively, for older talkers' fast speech condition and younger talkers' slow speech condition, the interlocutor's speech rate had a contrastive effect: when the interlocutor speaks fast (black bar), the target talker sounded older than when the interlocutor speakers slow (grey bar). On the other hand, for older talkers' slow speech condition and younger talkers' fast speech condition, the pattern was the opposite: the interlocutor's fast speech made the target talker sound younger than the interlocutor's slow speech. The effect of the interlocutor speech rate did not survive the stepwise regression and none of the effects involving interlocutor speech rate turned out to be statistically significant. In other words, participants' estimation of the target speaker's age was not systematically influenced by whether the interlocutor was speaking fast or slow in the dialogue.

---

[2] When only native English speakers are included, the speech rate effect for the younger speakers does not reach significance. It is not clear whether this is due to a meaningful difference between the native and non-native speakers' use of rate information in age judgment or merely an issue of statistical power.
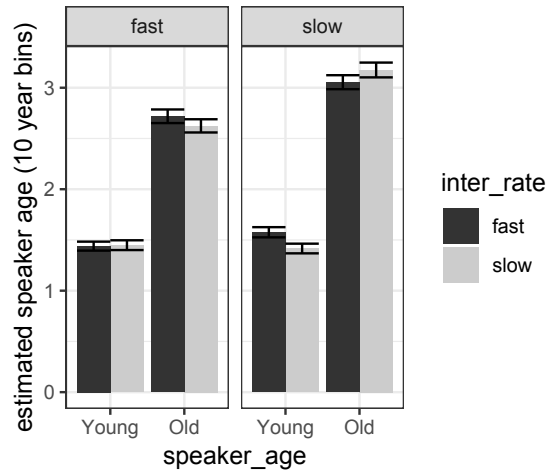
Figure 3. Estimated age by speaker age (x-axis), target talker's speech rate (panel) and interlocutor's speech rate

Now we turn to the effect of speaker order. The main effect of ORDER was not significant, but this factor interacted significantly with SPEAKER_AGE, as shown in Figure 4. A post-hoc test shows a statistically significant effect for older speakers; the speaker was judged as older when they were the second person to talk in the dialogue (and also the second person to be asked about age) than when they were the first speaker. For the younger speakers, the effect was in the opposite direction without reaching statistical significance. Note that the two speakers in the dialogue were always matched in their general age group in our design. Thus, it seems that after the participants made an age estimation of the first speaker as older or younger sounding, they more readily made the same judgment for the second speaker.
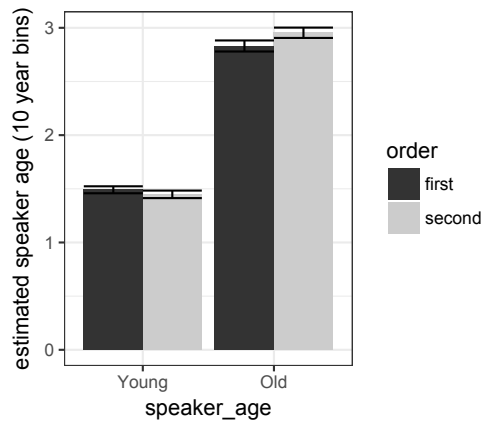
Figure 4. Estimated age by speaker age and dialogue order

Finally, we found a marginally significant main effect of participants' own age; the older the participant, the older they judged the speakers, as shown in Figure 5. In other words, it seems that with other things being equal, participants tend to judge speakers as more similar to their own age. This tendency was more pronounced for older speakers as shown by the significant interaction of SPEAKER_age and PARTICIPANT_AGE.
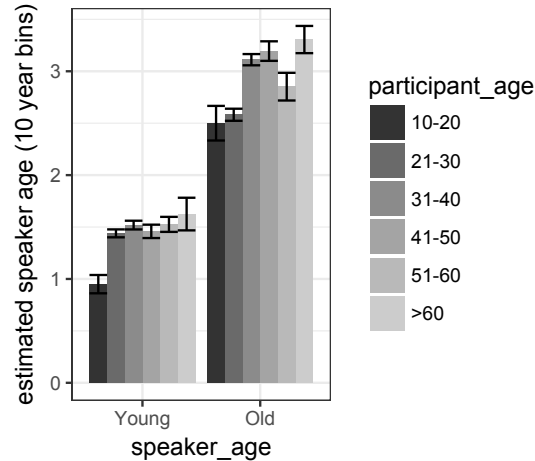
Figure 5. Estimated age by participant age

4   Discussion

Our study investigated the effectiveness of speech rate as a cue in age estimation. We aimed to fill the gap of previous research by including a wide age range of participants to probe the role of the listener effect and its interaction with speaker age estimation. We also used conversational speech involving multiple speakers to investigate the role of interlocutor in age estimation based on speech rate variation. We largely replicated findings of previous studies: the speech rate of the speaker played an important role, though the effect was more reduced for younger speakers and more substantial for older speakers. We did not find evidence to support that this asymmetrical effect of speech rate is due to the participant bias. Therefore, this asymmetry is still in need of an explanation. It seems that for younger talkers, non-speech rate cues to age override speech rate cues while non-speech rate cues are more ambiguous for older talkers. We instead uncovered another type of listener bias effect: other things being equal, younger listeners judged the speakers as younger than older listeners.

Finally, we did not find any systematic effect of interlocutors' speech rate on the listeners' judgement of speaker age. This may be because the two speakers always differed in gender in the experiment, allowing the listeners to reliably segregate the information from the two speakers and make subsequent inferences about the speakers without the interference of interlocutor information. However, we did find a different kind of interlocutor effect, i.e., the second talker of the younger pair was judged to be slightly younger than the first talker and likewise, the second talker of the older pair was judged to be slightly older than the first talker. This shows that the age perception of a talker is influenced by the perception of another talker in the same conversation after all.

References

Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, and Gabor Grothendieck. 2017. lme4 package, version 1.1-13 [Comupter software]. Available from https://cran.r-project.org/web/packages/lme4/.

Bona, Judit. 2014. Temporal characteristics of speech: The effect of age and speech style. *The Journal of the Acoustical Society of America* 136(2): EL116-EL121.

Braun, Angelika. 2013. Age estimation by different listener groups. *International Journal of Speech Language and the Law* 3(1): 65–73.

Braun, Angelika, and Loredana Cerrato. 1999. Estimating speaker age across languages. Paper presented at the *Proceedings of ICPhS*, 1999.

De Rosario-Martinez, Helios. 2015. phia package, version 0.2-1.

Harnsberger, James D, Rahul Shrivastav, WS Brown Jr, Howard Rothman, and Harry Hollien. 2008. Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice* 22(1): 58–69.

Hartman, David E. 1979. The perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders* 12(1): 53–61.

Hughes, Susan M, and Bradley C Rhodes. 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.

Jacewicz, Ewa, Robert A Fox, Caitlin O'Neill, and Joseph Salmons. 2009. Articulation rate across dialect, age, and gender. *Language Variation Change* 21(2): 233–256.

Krauss, Robert M., Robin Freyberg, and Ezequiel Morsella. 2002. Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology* 38(6): 618–625.

Moyse, Evelyne. 2014. Age estimation from faces and voices: a review. *Psychologica Belgica* 54(3): 255–265.

Moyse, Evelyne, Aline Beaufort, and Serge Brédart. 2014. Evidence for an own-age bias in age estimation from voices in older persons. *European Journal of Ageing* 11(3): 241–247.

Nagao, Kyoko, and Diane Kewley-Port. 2005. The effect of language familiarity on age perception. Paper presented at the *International Research Conference on Aging and Speech Communication. Bloomington, Indiana*, 2005.

Newman, Rochelle S, and James R Sawusch. 1996. Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics* 58(4): 540–560.

Ptacek, Paul H, and Eric K Sander. 1966. Age recognition from voice. *Journal of Speech and Hearing Research* 9(2): 273–277.

Ryan, William J, and Kenneth W Burk. 1974. Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders* 7(2): 181–192.

Skoog Waller, Sara, Mårten Eriksson, and Patrik Sörqvist. 2015. Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology* 6: 978.

Summerfield, Quentin. 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception* 7(5): 1074.

Team, R Development Core. 2016. R: A Language and Environment for Statistical Computing. Version 3.3.2. Vienna, Austria: R Foundation for Statistical Computing.

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

Zäske, Romi, Verena G Skuk, Jürgen M Kaufmann, and Stefan R Schweinberger. 2013. Perceiving vocal age and gender: An adaptation approach. *Acta Psychologica* 144(3): 583–593.

Appendix: Conversation script

A: Hey Ethan/Sara! How are you?

B: I'm well, Ethan/Sara. How are you?

A: I'm fine. This morning I'm going to the zoo with Michelle! She really loves animals.

B: Oh! How nice! I was there last May. The weather then was amazing! Are you also seeing the marine life show there?

A: Yes! We love their live show! When we're through with the zoo, we're rushing home for a movie marathon! We'll have some wine and cheese, sweets, and more!

B: Oh how fun! I'm seeing a movie with Alex, this evening as well! You know, the new Mulan movie everyone's fussing over.

A: I hear they got some strong reviews!

B: Yeah, I heard the same thing! Alex is a real fan of Mulan.

A: Nice! I saw the original film with Michelle the other evening. Though she really favours Mushu over Mulan. She thinks he's hilarious!

B: I am with her! He's so funny! Especially in that scene with the horses!

A: Mmhmm! I also really love the songs they use in the film! They're absolutely amazing! Overall, it's a really excellent film!

B: Oh for sure! It's one of my favourite scores! I must say John Williams has no rival in the realm of film music.

A: Right, he's well-known for a reason; his work for the Star Wars films was phenomenal! How else are you enjoying this marvellous weather?

B: I'm leaving for New Zealand soon!

A: That sounds amazing! When will you leave?

B: I'm leaving near the end of June and I will get to see John!

A: That's so sweet! He's lived there for a while now hasn't he?

B: This is his fifth year there! He loves New Zealand. Are you flying anywhere this summer?

A: No, this summer I'm staying home. I had my share of flying last month when I was in Asia. I think this summer I'll just relax.

B: That sounds nice! Well, I should run, I'll see you soon though! We're still on for next week?

A: Yes, absolutely! I'll see you then!