APRIL 06 2016

Effects of age on speech and voice quality ratings[a)]

Huiwen Goy; M. Kathleen Pichora-Fuller; Pascal van Lieshout

J Acoust Soc Am 139, 1648–1659 (2016)
https://doi.org/10.1121/1.4945094

Effects of age on speech and voice quality ratings[a)]

Huiwen Goy[b)] and M. Kathleen Pichora-Fuller
Department of Psychology, University of Toronto, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada

Pascal van Lieshout
Department of Speech-Language Pathology, University of Toronto, 500 University Avenue, Toronto, Ontario M5G 1V7, Canada

(Received 13 February 2015; revised 31 October 2015; accepted 18 March 2016; published online 6 April 2016)

The quality of communication may be affected by listeners' perception of talkers' characteristics. This study examined if there were effects of talker and listener age on the perception of speech and voice qualities. Younger and older listeners judged younger and older talkers' gender and age, then rated speech samples on pleasantness, naturalness, clarity, ease of understanding, loudness, and the talker's suitability to be an audiobook reader. For the same talkers, listeners also rated voice samples on pleasantness, roughness, and power. Younger and older talkers were perceived to be similar on most qualities except age. Younger and older listeners rated talkers similarly, except that younger listeners perceived younger voices to be more pleasant and less rough than older voices. For vowel samples, younger listeners were more accurate than older listeners at age estimation, while older listeners were more accurate than younger listeners at gender identification, suggesting that younger and older listeners differ in their evaluation of specific talker characteristics. Thus, the perception of quality was generally more affected by the age of the listener than the age of the talker, and age-related differences between listeners depended on whether voice or speech samples were used and the rating being made. © *2016 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4945094]

[MSS] Pages: 1648–1659

I. INTRODUCTION

The quality and intelligibility of speech are not necessarily highly correlated, but both can contribute to listeners' appraisal of talkers. Studies of speech processed through hearing aids have demonstrated that speech samples may not be perceived to have high quality even if speech is highly intelligible (Gabrielsson and Sjögren, 1979; Kates and Kozma-Spytek, 1994). Other research suggests that speech intelligibility may be one of many components of perceived speech quality (Preminger and Van Tasell, 1995), and that speech intelligibility can be dissociated from other talker qualities such as accent (Munro and Derwing, 1999). Studies of speech quality rating have often been conducted to evaluate speech perception by listeners who wear hearing aids, or to evaluate the communication ability of talkers who have various pathologies. However, perceived speech quality is also important in everyday interactions because listeners use the acoustic properties of speech to form judgments about a talker's characteristics, such as their physical traits (Bruckert *et al.*, 2006), personality (Klofstad *et al.*, 2012), emotional state (Scherer, 1995), cognitive ability (Dilley *et al.*, 2013), gender (Singh and Murry, 1978), and age (Hartman and Danhauer, 1976). In turn, these perceptions may then influence interactions with a conversational partner. For example, a person speaking to an older adult who is perceived to be frail and cognitively impaired may use exaggerated intonation and more repetition of information than when that person speaks to an older adult who is perceived to be active and healthy (Kemper *et al.*, 1998). These changes in communication may be helpful in the transmission of important information (Cohen and Faulkner, 1986), or may make speech sound more unpleasant and convey a lack of respect for a conversational partner (Ryan *et al.*, 1991). Importantly, the quality of speech has effects on the listener beyond those of intelligibility; a listener's judgments of speech quality may be associated with biological characteristics such as pathological conditions and social characteristics such as gender or age.

With respect to judgments related to age, many studies that compared the speech of healthy younger and older talkers have investigated topics such as the accuracy of age perception by listeners (Linville and Korabic, 1986) or the perceived competence of younger and older talkers (Ryan and Laurie, 1990). However, studies have rarely investigated age-related differences in both talkers and listeners on measures of perceived speech and voice quality. Understanding the true extent of differences between the productions of younger and older talkers and the influence of age on the perception of talkers may lead to a better understanding of the factors that affect communication with older adults, especially the potential role of negative stereotypes based on the perceived age of a talker (Ryan *et al.*, 1986).

[a)]Portions of this work were presented in "What makes a good talker? Perceptual ratings by younger and older listeners of speech and voice samples produced by younger and older talkers" at a poster session at the Aging and Speech Communication Conference in Bloomington, Indiana, October 2011.
[b)]Electronic mail: huiwen.goy@utoronto.ca

 0001-4966/2016/139(4)/1648/12/$30.00

When it comes to the acoustical parameters of speech, there are some well-known age-related differences in speech production, such as the lowering of F_0 in female talkers (Mueller, 1997) and the tendency for older talkers to speak more slowly than younger talkers (Duchin and Mysak, 1987). However, some findings of age-related differences in voice characteristics were based on studies that examined small groups of younger and older adults. In fact, the acoustic qualities of the voices of many younger and older talkers may not differ greatly, but the variation in acoustic characteristics tends to be larger in older adults such that voice measurements in older adults contain more extreme values (on the poorer end of the scale) compared to those of younger adults (Goy *et al.*, 2013). When a sample of 159 healthy younger adults were compared to 133 healthy older adults on a range of voice and speech characteristics under well-controlled recording conditions, the only significant age-related differences were that older males had more shimmer in their voices and spoke more softly compared to younger males, while older females had lower F_0, less variation in speaking F_0, and a longer maximum phonation time compared to younger females (Goy *et al.*, 2013). Thus, the acoustic evidence from voice measurements suggests that it could be difficult to distinguish healthy older talkers from younger talkers in terms of their voice quality. Nevertheless, despite the limited acoustic differences between younger and older talkers, listeners tend to judge both the acoustic and the social qualities of older talkers to be worse than those of younger talkers. Even if talkers are healthy older adults, listeners describe older talkers as being hoarser and more imprecise in their articulation than younger talkers (Hartman and Danhauer, 1976) and generally less "normal" than younger adults (Amerman and Parnell, 1990). Older talkers are also perceived to have negative social qualities such as being less powerful and less engaged (Montepare *et al.*, 2014), and listeners tend to associate older talkers with negative personality stereotypes (Hummert *et al.*, 1999). Thus, communication could be adversely affected if older talkers produce speech that is perceived to be less clear or more unpleasant by listeners or if negative ageist stereotypes are triggered.

One factor that may have contributed to some of the perceived differences between younger and older talkers reported in previous studies is that listeners were informed of the ages of talkers before rating their speech qualities (Hartman and Danhauer, 1976), which makes it difficult to isolate listeners' true perception of the quality of younger and older talkers from listeners' expectations of how talkers should sound based on age-related stereotypes. It has been demonstrated that older talkers are not given credit for being competent even when they supply a verbal description that is just as effective as that provided by younger talkers (Ryan and Laurie, 1990). The perceived differences between younger and older talkers may also reflect biases due to the characteristics of listeners rather than true differences between talkers. Listeners who participate in rating studies are frequently younger adults who differ from older adults in terms of hearing abilities and social experience, and some studies have reported that younger listeners behave differently than older listeners when judging the speech samples of younger and older talkers. These differences in rating behavior may be partly due to age-related changes in the auditory system that affect listeners' perception of speech, such an age-related increase in F_0 difference limens which can adversely affect concurrent vowel perception (Vongpaisal and Pichora-Fuller, 2007). These age-related auditory changes may then lead listeners of different age groups to use different types of acoustic information for judging speech. For instance, one study found that older listeners were less accurate than younger listeners at judging the age of talkers, possibly because younger listeners used F_0 as the main cue for perceiving the age of a talker, whereas older listeners did not find F_0 information to be as useful (Jacques and Rastatter, 1990). Another study found that older listeners used more cues than younger listeners to determine a talker's age insofar as older listeners incorporated formant information when judging a talker's age whereas younger listeners did not (Linville and Korabic, 1986). In contrast, other studies have concluded that a listener's age does not influence rating behavior; for example, older listeners did not appear to differ from younger listeners when asked about their preference for particular talkers (Hollien *et al.*, 1991). It is possible that these apparently contradictory findings result from younger and older listeners responding differently on some measures (e.g., judging age), but similarly on other measures (e.g., talker preference). In sum, it appears that listeners' characteristics such as age may affect their perception of talker speech or voice quality, but these characteristics have rarely been taken into account in past studies of the perception of quality.

In addition to listener characteristics, the extent of age-related differences in listeners' perceptions may also depend on the measure of quality used in a study. Many measures of speech quality have been used in studies on speech perception, but it is unclear how these different measures are related to one another. It is likely that some measures are better than others at distinguishing groups of talkers and would therefore be more useful in describing and classifying talkers in terms of the quality of their speech. Some subjective measures may also be more strongly related to acoustic measures than others. Simple vowel production is driven more directly by biological factors, compared to more complex speech production that involves both voicing and articulation. Since perceived speech quality and perceived voice quality have usually been measured in separate studies without comparisons within listeners, it has not been possible to determine how voice and speech quality are inter-related or what their relative weightings are when listeners appraise talkers. Additionally, it is not clear if younger and older listeners make use of the different kinds of information available in speech and voice in the same way. The voice is an easy signal to capture and measure in the laboratory or clinic, but speech is what is used for daily communication in the real world. It would be important to know whether listeners' perceptions of speech and voice are driven by similar aspects of each type of signal.

In the present study, the speech qualities that were used were similar to those used in a range of studies with healthy

talkers: pleasantness (Gelfer, 1988; Eadie and Doyle, 2002), naturalness (Gatehouse and Akeroyd, 2006), clarity of articulation (Bele, 2005), and loudness (Gelfer, 1988). Since speech clarity and speech intelligibility are not equivalent, in addition to rating clarity of articulation, listeners were also asked how easy it was to understand the talker (Preminger and Van Tasell, 1995; Eisenberg *et al.*, 1998). To place the rating task in an everyday context, listeners were also asked to judge the talker's suitability as an audiobook reader. For voice samples, listeners were asked to rate talkers on pleasantness (a parallel measure for the speech pleasantness rating scale), roughness [one of the best-known perceptual qualities, which is frequently correlated with acoustic measures of jitter, shimmer, and harmonics-to-noise ratio (HNR)], and power (a parallel measure for the speech loudness rating scale).

Importantly, younger and older listeners rated speech and voice samples from healthy younger and older talkers on a range of qualities without any prior knowledge of talkers' age or gender. Speech samples consisted of a single sentence extracted from a longer passage that had been read aloud; it was thought that these samples would be of sufficient length for the purposes of the current study as it has been shown that listeners are able to form an impression of a talker even from a brief utterance (McAleer *et al.*, 2014). Age groups were compared for both talkers and listeners and the relationships between their perceptual ratings and acoustic measures were evaluated. It was hypothesized that (a) speech and voice samples from younger talkers would not be rated more positively than samples from older talkers because listeners who were unaware of talkers' chronological ages would not be influenced by the limited acoustic differences between healthy older and younger adult talkers (Goy *et al.*, 2013), nor by negative age-related stereotypes, (b) older listeners would rate samples differently than younger listeners, based on previous findings that older listeners use different acoustic cues to evaluate talkers, (c) talkers who were perceived to articulate clearly would not necessarily be perceived to have pleasant-sounding speech, since speech intelligibility and speech quality can be rated differently, and (d) perceptual ratings would relate to some acoustic measurements; specifically, higher perturbation (shimmer or jitter) in talkers' voices would be perceived as unpleasant, similar to past findings for artificial voices (Hillenbrand, 1988). By testing talkers and listeners of both age groups and using a variety of rating scales, this study aimed to further our understanding of how the perception of quality is affected by the interaction between talker and listener characteristics and to extend existing knowledge about age-related differences in speech and voice production.

II. METHOD

A. Talker selection and stimuli preparation

Speech and voice samples of talkers were selected from a normative database (Goy *et al.*, 2013) that contained recordings of talkers reading the first paragraph of the Rainbow Passage (Fairbanks, 1960) and phonating the vowel [a] for 8 s. The final sample of talkers consisted of 30 younger adults with a mean age of 19.0 years (range 18–23) and 30 older adults with a mean age of 71.3 years (range 63–81), with equal numbers of male and female talkers in each age group. To be selected from the database, a panel of three experimenters had to agreed that a talker did not speak with a distinctive regional dialect and had at least one speech sample and one voice sample that did not contain artifacts such as clipping. Up to three speech samples and three vowel samples were available for each talker; if all three samples met the selection criteria then the second sample was chosen. For each talker, the sentence "People look, but no one ever finds it" was extracted from the chosen speech recording, and a 4-s sample was extracted from the chosen vowel recording. The extracted portion of the vowel was the portion with the most stable intensity and the least amount of jitter and shimmer across the entire recording as measured by the speech analysis program Sonneta (MintLeaf Software, Inc.).

B. Listeners

Younger listeners were 40 undergraduate students at the University of Toronto Mississauga who participated in the study for course credit; they had a mean age of 19.6 years (range 18–28) and pure-tone audiometric thresholds of $\leq$25 dB hearing level (HL) from 250 to 8000 Hz in both ears. Older listeners were 40 community-living volunteers who had responded to newspaper advertisements inviting participation in studies on healthy aging at the University of Toronto Mississauga. They received an honorarium of $10/h. Their mean age was 74.1 years (range 66–84). They had pure-tone audiometric thresholds of $\leq$25 dB HL from 250 to at least 3000 Hz in at least one ear, and no more than two frequencies at which inter-aural differences were greater than 15 dB. All listeners had learned English before the age of 5 yr in an English-speaking country. There were 12 male and 28 female listeners in each age group.

C. Ratings

For both the speech and voice samples, after hearing each utterance, listeners first decided on the talker's gender and then selected which of 15 five-year age categories (from 16–20 to 86–90) they thought corresponded to the age of the talker. For speech samples, listeners rated each talker on pleasantness, naturalness, clarity, ease of understanding, loudness, and the talker's suitability as an audiobook reader, in this order. For voice samples, listeners rated each talker on pleasantness, roughness, and power, in this order. A talker's perceived gender is likely to be an important characteristic that will affect a listener's estimation of the talker's age given that some voice qualities have a different aging trajectory for males and females. For instance, F_0 provides a useful cue for differentiating talkers by gender, and F_0 decreases significantly with age for females, whereas there is a less dramatic age-related change in F_0 for males (Linville, 1996). There is some evidence that listeners who have been informed that a talker is older will rate the content of the talker's speech more negatively, even when it is identical to that of a younger talker (Ryan and Laurie, 1990). Thus, in combination, a talker's perceived gender and perceived age

may influence a listener's perception of other speech and voice qualities. In this study, qualities such as pleasantness and naturalness were considered to describe a talker more generally, whereas qualities such as loudness or roughness were more specific. Thus, speech and voice qualities were tested in the above order to encourage listeners to judge talkers on overall quality before judging them on more specific aspects because the judging of specific qualities is more likely to affect the judging of broader qualities than the other way around. Similar to other previous studies on quality rating of healthy talkers by non-expert listeners (Fagel *et al.*, 1983), listeners evaluated speech and voice samples using semantic differential rating scales with seven categories, where the first category denoted a very poor quality and the last denoted a very good quality, with the middle category as a neutral midpoint. The labelled endpoints of voice quality scales were: Very pleasant/very unpleasant, very smooth/very rough, and very powerful/very weak. Likewise, the labelled endpoints of speech quality rating scales were: Very pleasant/very unpleasant, very natural/very unnatural, very clear/very unclear, very easy to understand/very difficult to understand, and very loud/very soft. For the audiobook reader quality scale, the scale had five categories and a neutral midpoint, with endpoints labelled "yes" and "no." Since the endpoints on this rating scale were decisions of yes and no rather than a quality, it seemed that one additional label ("maybe") between each endpoint and the neutral midpoint was sufficient for listeners to express some degree of uncertainty, but that adding two categories between the endpoints and neutral midpoint would not necessarily have been meaningful to listeners.

D. Procedures

Listeners were seated in a double-walled sound-attenuating booth and stimuli were presented binaurally through Sennheiser Linear HD265 headphones at 70 dB sound pressure level using Tucker Davis Technologies System III hardware. The sampling rate was down-sampled from 48000 to 24414 Hz and all sound files were equated on RMS before presentation. Computer touchscreens were used by listeners to control the timing of the presentation of sound files, and to record ratings for each stimulus. On each trial, the listener rated one talker's speech or voice sample. After selecting the ratings on various scales for a particular talker, listeners then pressed a "next" button to go on to the next talker.

The speech and voice samples were presented in separate sessions. Half the listeners were tested on speech trials first, while the rest were tested on voice trials first. Listeners heard both speech and voice samples from the same talkers, although listeners were not told that the samples in the two sessions belonged to the same talkers. Prior to each test session, listeners rated the speech or voice samples of two talkers for practice; samples from these two talkers were not presented during the test session.

One of 20 lists, each with 18 trials, was selected to test each participant (see experimental design below). In each session, the trials in the selected test list were presented in a fixed order. The stimuli in each list had been previously randomized, and the same order of stimuli was used for each listener who heard a particular list. On each trial, all rating scales were listed in vertical order on the touchscreen and listeners had to supply a rating for each scale in the order stated in Sec. II C (e.g., a listener could not supply a loudness rating before supplying a pleasantness rating). Listeners were allowed to change any of the ratings that they had already supplied before proceeding to the next talker sample, and they were also allowed to replay the sample up to three times during the trial. The data showed that listeners proceeded to the next trial without replaying the sample on 63.6% of speech trials and 69.1% of voice trials; the third and final allowed replay for a sample was used only on 2.4% of speech trials and 1.2% of voice trials.

E. Experimental design

Samples from 60 talkers were distributed across 20 lists, which were presented to 40 younger listeners and 40 older listeners, using the same experimental design for both listener groups. Of the 60 talkers, four talkers (one from each of the four talker groups: younger/older; male/female) were rated by all participants in each listener group (younger/older) to test inter-rater reliability. The remaining 56 talkers were each rated by eight different participants from each listener group. In each list of samples, at least two of the samples were repeated so that intra-rater reliability could be evaluated. Some listeners were presented with more than two repeated talkers in order for each listener to complete a total of 18 trials and for each talker to be rated by eight different listeners. Thus, although all listeners completed 18 trials, 70% of listeners rated 16 unique talkers, 10% rated 15 unique talkers, 10% rated 13 unique talkers, and 10% rated 12 unique talkers. For each of the seven speech qualities and four voice qualities for each talker, ratings were averaged across listeners in each age group (after discarding the second of any repeated trials) to yield one average rating by each listener age group of each talker's speech sample and one average rating by each listener age group of each talker's voice sample. The rationale for having listeners rate subsets of talkers rather than the entire set of talkers was partly based on practical considerations with regards to testing time and preventing listener fatigue, especially for older listeners. Furthermore, by varying the voices heard by different listeners, consistent effects of other voices on the ratings of a given voice were minimized. In addition, since listeners in the real world are not all exposed to the same set of talkers, having listeners rate certain talkers embedded within slightly different sets of talkers would presumably reflect the exposure to a mix of talkers in real-world experience.

F. Data analysis

Rating scales were recoded from descriptors to values from 1 to 7, with 1 denoting the poorest quality on the scale and 7 denoting the best quality; the audiobook reader scale was similarly recoded to values from 1 to 5. Each age range category was recoded to the midpoint value of that range (e.g., the age range of 21–25 years was recoded to 23). When calculating intra-rater test-retest agreement and inter-

tester reliability, we included all trials, including those on which the talker's gender had been incorrectly identified. For all other analyses, we discarded trials in which the listener incorrectly identified the talker's gender from either the talker's speech or voice sample (10.2% of all trials) because a talker's perceived gender was expected to influence other ratings such as perceived age.

For each of the 11 speech and voice qualities, an average rating from at least 8 younger listeners and an average rating from at least 8 older listeners was obtained for each of the 60 talkers. The majority of ratings collected from younger and older listeners were normally distributed; thus, two multivariate analyses of variance (MANOVA) were conducted, one for speech ratings and another for voice ratings, with the age group of talkers as the between-subjects factor and the age group of listeners as the within-subject factor. *Post hoc* *t*-tests were conducted for individual dependent variables. To investigate the correlation between younger listeners' and older listeners' ratings for the 60 talkers, Pearson's correlations were calculated between the 60 average ratings of younger listeners and the 60 average ratings of older listeners for each scale.

To evaluate the effects of acoustic predictors on rating outcomes, a linear model with standardized coefficients was constructed for each speech and voice quality rating scale. The acoustics of the speech and voice stimuli were measured using PRAAT (Boersma and Weenink, 2014). In these models, speech rate, mean speaking F_0, and speaking F_0 standard deviation were tested as predictors of speech qualities (Harnsberger *et al.*, 2008; Medrado *et al.*, 2005; Rothman *et al.*, 2001). Mean F_0, jitter (relative average perturbation), shimmer (11-point amplitude perturbation quotient), and HNR were tested as predictors of voice qualities (Hillenbrand, 1988). Separate models were constructed for male and female talkers and for younger and older listeners. As some scales were correlated with others, partial correlations were calculated among speech quality scales and among voice quality scales to examine the relationships between different scales. The Holm-Bonferroni correction was applied to all *t*-tests, linear models, and correlations to control for type I error, with $\alpha = 0.05$.

III. RESULTS

A. Intra-rater agreement

All listeners rated at least two talkers twice. Following previous studies (Bunton *et al.*, 2007), the first and second ratings of a talker were considered to agree if the second rating fell within one unit of the first rating. Table I shows the agreement between first and second ratings across all listeners, talkers and scales, which was higher for speech qualities than for voice qualities. Younger and older listeners showed similar levels of agreement between first and second ratings; 82.4% and 83.4% of all repeated ratings were within ±1 unit of the original rating for younger and older listeners, respectively.

B. Inter-rater reliability

Inter-rater reliability was assessed for each rating scale for each listener age group using intra-class correlation coefficients (ICC), with raters as random factors and a composite of 40 ratings [case (2, k) in Shrout and Fleiss (1979)]. Most ICC values were >0.8 (Table II), higher than or comparable to those of previous studies in which listeners rated a fixed set of talkers rather than different subsets of talkers as in the current study (Maryn and Debo, 2014; Mulac and Giles, 1996).

TABLE I. Intra-rater agreement between original and repeated ratings on speech and voice quality rating scales, showing the percentages of repeated ratings that were the same as the original ratings, and the percentages of repeated ratings that were within ±1 unit of the original rating.

	Younger listeners		Older listeners	
Speech quality	Same (%)	±1 unit (%)	Same (%)	±1 unit (%)
Perceived age	42.9	87.5	38.4	76.8
Pleasantness	56.3	89.3	54.5	91.1
Naturalness	42.0	79.5	48.2	83.0
Clarity	55.4	87.5	58.9	85.7
Ease of understanding	56.3	92.0	53.6	90.2
Loud	56.3	92.9	71.4	99.1
Reader	48.2	89.3	50.9	84.8
Voice quality	Same (%)	±1 unit (%)	Same (%)	±1 unit (%)
Perceived age	34.8	74.1	24.1	54.5
Pleasantness	50.9	78.6	50.0	88.4
Roughness	49.1	78.6	45.5	85.7
Power	42.9	75.0	50.9	89.3

C. Accuracy of gender and age judgments

As shown in Table III, gender identification by both younger and older listeners was generally highly accurate (>90%) for both younger and older talkers. The exception was that younger listeners frequently mistook the voice samples of older females for those of younger males. In contrast, older listeners made this error much less frequently than younger listeners (about 20 percentage points).

On trials in which older females were wrongly perceived to be male, talkers were perceived to be young males with an

TABLE II. Intra-class correlation coefficients with lower and upper bounds for different rating scales and two listener age groups.

	Younger listeners			Older listeners		
Speech	ICC	Lower bound	Upper bound	ICC	Lower bound	Upper bound
Perceived age	0.98	0.95	1	0.99	0.95	1
Pleasant	0.96	0.88	1	0.85	0.574	0.99
Natural	0.93	0.8	1	0.72	0.24	0.98
Clarity	0.83	0.52	0.99	0.69	0.18	0.98
Ease	0.79	0.42	0.98	0.66	0.14	0.97
Loudness	0.88	0.646	0.99	0.91	0.725	0.99
Reader	0.86	0.574	0.99	0.77	0.2826	0.98
Voice						
Perceived age	0.96	0.87	1	0.93	0.772	0.99
Pleasant	0.98	0.93	1	0.96	0.87	1
Rough	0.98	0.93	1	0.95	0.83	1
Power	0.88	0.628	0.99	0.91	0.726	0.99

TABLE III. Percentage of trials on which the listener correctly identified the talker's gender.

	Younger listeners	Older listeners
Speech		
Younger males	98.0	97.4
Older males	98.7	96.7
Younger females	100.0	100.0
Older females	99.3	100.0
Voice		
Younger males	96.1	91.4
Older males	99.3	96.1
Younger females	96.1	97.4
Older females	64.5	82.8

average age of 24 years and to have less pleasant voices than those of older females who had been correctly identified as female ($p = 0.006$). When all gender-incorrect trials were included in the analysis, older talkers were perceived to be poorer readers than younger talkers ($p = 0.025$), whereas this talker age-related difference was absent when gender-incorrect trials were excluded ($p = 0.09$), suggesting that incorrect identification of a talker's gender led to downstream effects in the perception of the talker's other qualities.

Including only gender-correct trials, regardless of whether speech or voice samples were presented, both listener age groups over-estimated younger talkers' true ages and underestimated older talkers' true ages. As shown in Table IV, the mean perceived ages were at least one age category above the true ages for younger talkers, and several age categories below the true ages for older talkers. The most accurate age estimation occurred when younger listeners evaluated the speech samples of younger females (Table IV).

D. Relationships between rating scales

Partial correlation coefficients were calculated to evaluate the strength of relationships between different rating scales (Table V). The relationship between perceived age and other qualities was different between younger and older listeners. For younger listeners, when talkers were perceived to be older, their speech was also perceived to be less pleasant ($p = 0.01$) and louder ($p = 0.02$). For older listeners, talkers who were perceived to be older also had voices that were perceived to be rougher ($p = 0.03$).

The relationship between pleasantness and other qualities was similar between younger and older listeners. For both younger and older listeners, speech that was perceived to be more pleasant was also rated as being more natural ($p < 0.01$) and louder ($p < 0.05$), and talkers with a more pleasant speech quality were rated as being more suited to be audiobook readers ($p < 0.001$). Both younger and older listeners perceived more pleasant voices to be less rough ($p < 0.001$).

The clarity of speech was highly correlated with ease of understanding for both younger and older listeners ($p < 0.001$). Younger listeners perceived speech to be easier to understand if it was also perceived to be softer ($p = 0.002$). For younger listeners, more pleasant voices were also perceived to be more powerful ($p = 0.02$), whereas for older listeners, voices that were less rough were perceived to be more powerful ($p = 0.04$).

Comparing speech samples with voice samples, speech pleasantness and voice pleasantness were correlated when younger listeners evaluated younger talkers ($r = 0.62$, $p < 0.001$) and when older listeners evaluated older talkers ($r = 0.42$, $p = 0.02$); however, there was no significant relationship between speech pleasantness and voice pleasantness when listeners evaluated talkers who belonged to a different age group than themselves ($p > 0.1$).

E. Effects of talker and listener age group on quality ratings

Younger and older talkers were rated similarly on all speech and voice qualities except for perceived age and perceived voice pleasantness and roughness, such that older

TABLE IV. Comparison of actual ages of talkers to their mean perceived ages by listeners, with *t*-values from independent samples *t*-tests and the associated *p*-values with Holm-Bonferroni correction.

		Perceived age (years) from speech samples					
		Younger listeners			Older listeners		
Talker group	Actual age	Perceived age	*t*	*p*	Perceived age	*t*	*p*
Younger males	19.0	24.7	−5.02	<0.001	27.4	−6.3	<0.001
Older males	71.8	49.6	10.37	<0.001	50.3	11.26	<0.001
Younger females	18.9	20.6	−2.78	0.011	23.6	−5.72	<0.001
Older females	71.0	47.9	8.78	<0.001	48.3	9.7	<0.001
		Perceived age (years) from voice samples					
		Younger listeners			Older listeners		
Talker group	Actual age	Perceived age	*t*	*p*	Perceived age	*t*	*p*
Younger males	19.0	26.7	−6.56	<0.001	34.5	−8.87	<0.001
Older males	71.8	39.2	12.87	<0.001	43.3	12.37	<0.001
Younger females	18.9	22.5	−5.79	<0.001	28.9	−8.17	<0.001
Older females	71.0	38.8	10.35	<0.001	42.6	11.09	<0.001

TABLE V. Partial correlation coefficients for relationships between different speech quality scales and between different voice quality scales. Boldface values indicate significant partial correlations between qualities ($p < 0.05$ with Holm-Bonferroni correction).

Younger listeners						
Speech quality	Age	Pleasant	Natural	Clarity	Ease	Loud
Perceived age						
Pleasant	**−0.34**					
Natural	0.21	**0.48**				
Clarity	0.07	−0.12	0.29			
Ease	0.00	0.07	−0.23	**0.83**		
Loud	**0.32**	**0.28**	−0.21	0.12	**−0.41**	
Reader	0.20	**0.65**	0.02	0.01	0.23	−0.01
Voice quality	Age	Pleasant	Rough	Power		
Perceived age						
Pleasant	−0.07					
Rough	−0.26	**0.81**				
Power	−0.04	**0.31**	−0.23			
Older listeners						
Speech quality	Age	Pleasant	Natural	Clarity	Ease	Loud
Perceived age						
Pleasant	−0.21					
Natural	0.13	**0.41**				
Clarity	−0.18	0.01	0.03			
Ease	0.18	0.09	0.15	**0.75**		
Loud	0.01	**0.41**	0.02	−0.11	−0.14	
Reader	0.20	**0.52**	0.23	0.26	−0.06	−0.12
Voice quality	Age	Pleasant	Rough	Power		
Perceived age						
Pleasant	0.13					
Rough	**−0.30**	**0.77**				
Power	−0.05	0.06	**0.28**			

TABLE VI. Comparison of ratings by younger and older talkers for each age group of listeners. Standard deviations are shown in parentheses. Boldface rows indicate significant differences between means ($p < 0.05$ with Holm-Bonferroni correction).

Younger listeners				
Speech quality	Younger talkers	Older talkers	t	p
Perceived age	**22.67 (3.92)**	**48.78 (7.77)**	**−16.43**	**< 0.001**
Pleasantness	4.49 (0.77)	4.30 (0.96)	0.84	1.00
Naturalness	4.89 (0.76)	4.98 (0.82)	−0.45	0.66
Clarity	5.55 (0.84)	5.82 (0.49)	−1.51	0.83
Ease of understanding	5.82 (0.66)	5.97 (0.44)	−0.99	1.00
Loudness	3.32 (0.47)	3.47 (0.52)	−1.11	1.00
Reader	2.95 (0.63)	3.09 (0.71)	−0.83	0.82
Voice quality	Younger talkers	Older talkers	t	p
Perceived age	**24.59 (4.03)**	**39.01 (9.64)**	**−7.56**	**< 0.001**
Pleasantness	**4.18 (0.98)**	**3.53 (0.80)**	**2.83**	**0.013**
Roughness	**4.37 (1.00)**	**3.65 (0.93)**	**2.88**	**0.017**
Power	4.43 (0.42)	4.23 (0.67)	1.37	0.18
Older listeners				
Speech quality	Younger talkers	Older talkers	t	p
Perceived age	**25.52 (4.52)**	**49.29 (6.61)**	**−16.26**	**< 0.001**
Pleasantness	4.42 (0.64)	4.51 (0.71)	−0.55	1.00
Naturalness	4.82 (0.54)	5.03 (0.77)	−1.23	0.90
Clarity	5.53 (0.51)	5.67 (0.43)	−1.14	0.78
Ease of understanding	5.61 (0.47)	5.80 (0.38)	−1.71	0.47
Loudness	3.71 (0.34)	3.67 (0.34)	0.44	0.66
Reader	2.80 (0.55)	3.09 (0.71)	−1.75	0.51
Voice quality	Younger talkers	Older talkers	t	p
Perceived age	**31.68 (6.31)**	**42.94 (7.91)**	**−6.10**	**< 0.001**
Pleasantness	4.09 (0.58)	4.14 (0.70)	−0.30	0.76
Roughness	4.15 (0.71)	4.05 (0.66)	0.58	1.00
Power	4.38 (0.40)	4.25 (0.55)	1.01	0.95

talkers received higher scores for perceived age and lower scores for pleasantness and roughness than younger talkers. Younger and older listeners rated speech and voice qualities similarly except for perceived age and loudness, on which older listeners gave higher scores than younger listeners. More specifically, for speech sample ratings, a MANOVA showed a main effect of talker age, $F(7, 52) = 50.21$, $p < 0.001$ and a main effect of listener age, $F(7, 52) = 4.97$, $p < 0.001$, but no interaction of talker age with listener age, $p = 0.09$. For voice sample ratings, there was a main effect of talker age, $F(4, 55) = 15.46$, $p < 0.001$, a main effect of listener age, $F(4, 55) = 13.68$, $p < 0.001$, and an interaction of talker age with listener age, $F(4, 55) = 3.99$, $p = 0.007$. As shown in Table VI, *post hoc* independent samples t-tests showed that older talkers were generally perceived to be older than younger talkers on both speech and voice samples ($p < 0.001$), but otherwise younger and older talkers were not rated differently on speech qualities. Compared to younger listeners, older listeners rated talkers as louder on speech samples ($p < 0.001$) and older from their voice samples ($p < 0.001$).

To investigate interactions between talker and listener age groups for voice samples, the dataset was split by listener age group and younger and older talkers were compared on voice qualities using t-tests with Holm-Bonferroni corrections. As illustrated in Fig. 1, when listening to voice samples, younger listeners perceived older voices to be rougher and less pleasant than younger voices, but older listeners perceived no overall difference between younger and older voices on either of those qualities. The ratings of younger and older listeners were highly correlated for the perceived ages of talkers and moderately correlated on all other scales (Table VII).

F. Acoustic measures of speech and voice samples

Comparisons of acoustic measures were made between younger and older talkers (Table VIII). Both male and female younger adults spoke faster than their older counterparts. Older females had a lower F_0 than younger females in their running speech; the average F_0 of older females in voice samples did not differ significantly from the average F_0 of younger females, but the range of F_0 in voice samples

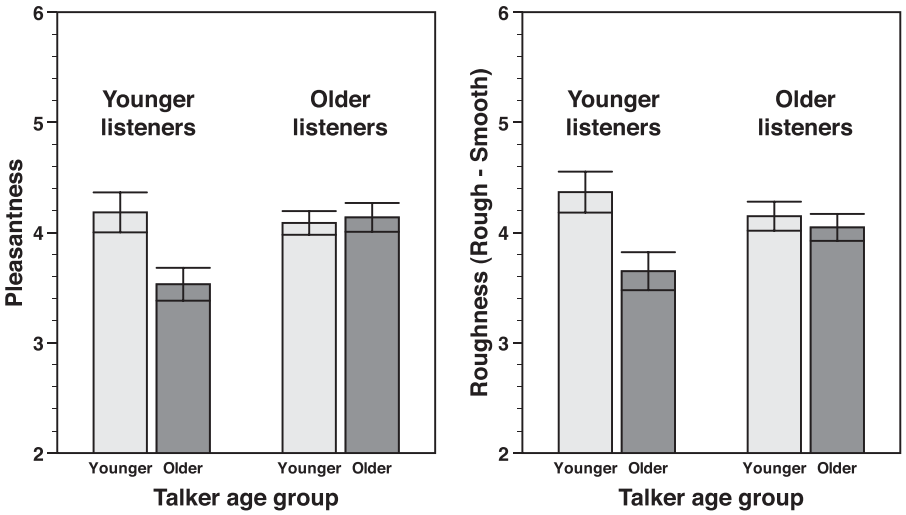

FIG. 1. Mean levels of pleasantness and roughness of young and older talkers' voice samples, as perceived by younger and older listeners. Standard error bars are shown.

of older females was considerably larger than that of younger females. There were no significant differences between talkers on other acoustic measures.

G. Relationship between acoustics and perceptual ratings

Results from linear models showed that speech rate and F_0 predicted listeners' ratings on a limited number of scales that were mainly related to perceived age. Older listeners perceived male talkers to be younger when speech rate was faster ($\beta = -0.5628$, $p = 0.001$) and when male voices had a higher mean F_0 ($\beta = -0.3445$, $p = 0.03$), while younger listeners perceived female talkers to be younger when speech rate was faster ($\beta = -0.3692$, $p = 0.03$) and when female talkers' speaking F_0 was higher ($\beta = -0.4348$, $p = 0.01$). Younger listeners perceived female talkers to have rougher voices when voice samples had a lower mean F_0 ($\beta = 0.7084$, $p < 0.001$). There were no other significant relationships between acoustic measures and listener ratings.

IV. DISCUSSION

In the present study, speech and voice samples produced by 30 younger and 30 older talkers were rated by 40 younger and 40 older listeners on a variety of perceptual scales. Younger and older talkers were compared on their speech and voice qualities as perceived by listeners, and younger and older listeners were compared on their ratings of talkers. The associations between speech and voice quality scales were examined, and ratings were correlated with acoustic measurements.

TABLE VII. Correlations between composite ratings of 60 talkers by younger listeners and composite ratings of 60 talkers by older listeners.

Speech quality	r	p
Perceived age	0.95	<0.001
Pleasantness	0.56	<0.001
Naturalness	0.41	0.0011
Clarity	0.44	<0.001
Ease of understanding	0.40	0.0017
Loudness	0.56	<0.001
Reader	0.46	<0.001
Voice quality	r	p
Perceived age	0.78	<0.001
Pleasantness	0.52	<0.001
Roughness	0.58	<0.001
Power	0.53	<0.001

A. Comparison of ratings for younger and older talkers

The first hypothesis was that younger and older talkers would receive similar ratings on quality from listeners due to minimal acoustic differences between talkers in different age groups and the lack of *a priori* activation of negative age-related stereotypes on the part of listeners. The results

TABLE VIII. Means (standard deviations) of acoustic characteristics of younger and older talkers. Boldface rows denote significantly different means between younger and older talkers.

	Male talkers			
Speech samples	Younger	Older	t	p
Rate (syllables/sec)	**5.5 (0.8)**	**4.5 (0.7)**	**3.77**	**0.002**
Mean speaking F_0 (Hz)	126 (21)	117 (22)	1.17	0.50
Speaking F_0 SD (Hz)	39 (33)	27 (23)	1.15	0.26
Voice samples	Younger	Older	t	p
Mean F_0 (Hz)	131 (25)	123 (29)	0.81	1.00
Jitter RAP (%)	0.23 (0.12)	0.23 (0.15)	−0.003	1.00
Shimmer APQ11 (%)	**1.99 (0.56)**	**3.43 (1.83)**	**−2.92**	**0.039**
HNR (dB)	22.5 (3.1)	22.9 (5.4)	−0.28	1.00
	Female talkers			
Speech samples	Younger	Older	t	p
Rate (syllables/sec)	**5.3 (0.9)**	**4.6 (0.6)**	**2.67**	**0.04**
Mean speaking F_0 (Hz)	**232 (32)**	**203 (30)**	**2.52**	**0.04**
Speaking F_0 SD (Hz)	57 (16)	57 (26)	0.008	0.99
Voice samples	Younger	Older	t	p
Mean F_0 (Hz)	267 (38)	227 (60)	2.21	0.15
Jitter RAP (%)	0.21 (0.12)	0.18 (0.10)	0.69	1.00
Shimmer APQ11 (%)	1.55 (0.66)	1.68 (1.23)	−0.37	1.00
HNR (dB)	26.4 (3.2)	26.5 (3.7)	−0.14	0.89

supported this hypothesis; younger and older talkers received similar ratings on nearly all speech and voice qualities except perceived age. Even for perceived age, younger and older talkers were judged to be closer in age than they actually were. The lack of differences between the ratings of younger and older talkers suggests that there are either no reliable acoustical bases for the two age groups to be perceived differently, or that age-related differences in past studies may have resulted from listeners' prior knowledge of the ages of talkers and the activation of negative ageist stereotypes. It is possible that the lack of age-related differences between talkers in this study was due to listeners' perception of the talkers as being younger and thus also better, but this explanation seems unlikely to be correct as there is no significant correlation between perceived age and speech qualities. Another possible explanation for the relative lack of age-related talker differences is the use of samples of read speech rather than spontaneous speech. In general, read speech tends to have a lower articulation rate (Laan, 1997), fewer pauses and different stresses compared to spontaneously produced speech (Howell and Kadi-Hanifi, 1991). It is not known if younger and older talkers show similar differences between read and spontaneous speech; however, it is possible that age-related differences would be less pronounced during reading than in spontaneous speech because reading is more constrained, leading both age groups to produce more similar utterances than they would have if they had been speaking spontaneously.

Past studies have shown that listeners are able to classify talkers into broad age categories at above-chance accuracy levels, but that classification errors are frequently made, especially if listeners only had information from voice samples. For instance, when vowels produced by female talkers between the ages of 20 and 80 years were presented to young adult listeners, only 51% of talkers were correctly classified as being young, middle-aged or old (Linville and Fisher, 1985). Another study showed that older listeners performed at near-chance accuracy levels when they were asked to classify younger and older talkers into two age categories based on vowel samples (57% accuracy), whereas younger listeners performed better (71% accuracy; Jacques and Rastatter, 1990). Listeners in the present study were asked to make direct age estimations of talkers rather than classifying talkers into broad age categories. For speech samples, younger and older listeners made similar judgments of age, but for voice samples, older listeners judged talkers (especially younger talkers) to be older than the age judged by younger listeners. From the results of the present study and the study by Jacques and Rastatter (1990), it seems that younger listeners are more accurate on the task of age estimation when presented with vowels, but younger and older listeners do not differ on age estimation when listening to speech. Similar to past studies in which listeners were asked to make direct age estimations of talkers (Harnsberger *et al.*, 2010; Huntley *et al.*, 1987; Ryan and Capadano, 1978), the present results showed that both younger and older listeners overestimated the ages of younger talkers and underestimated the ages of older talkers from both speech and voice samples. In one previous study, listeners were found to increase their age estimations of talkers after exposure to voices of younger talkers and vice versa (Zäske and Schweinberger, 2011). Since studies usually present stimuli from talkers of different age groups in randomized order (including the present study), the general tendency of listeners to overestimate younger talkers' ages and underestimate older talkers' ages may be partly due to the effects from interspersing samples from younger and older talkers. Alternatively, listeners may simply find it difficult to estimate talkers' ages from voice samples because vocal aging does not follow the same time course as chronological aging (Huntley *et al.*, 1987).

B. Comparison of ratings by younger and older listeners

Our second hypothesis was that the age of listeners would have an effect on their ratings of speech and voice samples. The results showed that the age of listeners affected quality ratings, but more so for voice samples than for speech samples. For speech samples, relative to younger listeners, older listeners perceived talkers to be louder (though the group difference for perceived loudness was approximately 1/3 of a unit on a scale of 7 units and is not likely to be meaningful), but the two age groups of listeners did not differ on their average ratings for other speech qualities. Notably, the two age groups of listeners were similar in their average perception of talkers, and their ratings were moderately correlated for other speech qualities besides perceived age, suggesting that younger and older listeners differed subtly on their perception of individual talkers.

For voice samples, younger listeners perceived the voices of older talkers to be rougher and less pleasant than the voices of younger talkers, but older listeners did not show such an effect, suggesting a negative bias towards older voices by younger listeners. This bias in younger listeners is unexpected in light of listeners' general underestimation of older talkers' ages. One possible explanation for this finding is that while the acoustic quality of older voices did not fit younger listeners' stereotypes of what older talkers should sound like (Ryan and Capadano, 1978), younger listeners may have perceived enough of the subtle acoustic differences in the voices of older talkers compared to the voices of younger talkers that these older talkers sounded less pleasant (or more unhealthy), whereas older listeners did not perceive these acoustic cues in the same way. Alternatively, it is possible that younger listeners attributed negative qualities to talkers whom they perceived to be middle-aged, whereas older listeners did not have such an association.

Despite the high correlation of perceived age between younger and older listeners, older listeners in the present study were less accurate at judging the ages of younger talkers from voice samples. Specifically, younger listeners judged younger talkers to be in the next age category up (21–25 years) from their true age (19 years), while older listeners judged younger talkers to be three age categories beyond their true age (31–35 years). This pattern of findings is similar to that of previous studies (Huntley *et al.*, 1987; Linville and Korabic, 1986), which have attributed these age-related differences in listeners' behavior to listeners' familiarity with talkers who are in

the same age group. However, familiarity effects do not explain why older listeners in this study selected a younger age category (41–45 years) than younger listeners did (46–50 years) when listening to the voice samples of older talkers whose true age was 71 years on average.

In addition to estimating a talker's age, gender is another talker characteristic that may be judged based on the individual's voice and speech. For judgments based on speech, both listener age groups performed well when identifying talkers' age and gender. However, in contrast to our finding that age was estimated less accurately by older listeners than by younger listeners based on voice samples, older listeners were better than younger listeners at identifying the gender of older female talkers from voice samples. In other words, younger listeners seemed to rely more on speech than voice information for gender identification, whereas older listeners performed fairly well in judging gender based on either speech or voice. Our data are in line with other studies of older listeners in showing that both formant and F_0 information may help with gender identification (e.g., Schwartz and Rine, 1968). Conversely, older listeners seemed to rely more on speech than voice information for age identification, whereas younger listeners performed fairly well in judging age based on either speech or voice. Consistent with our finding that older listeners rely more on speech than voice information for age identification, a previous study found that older listeners used formant information in addition to F_0 information to judge the age of talkers from voice samples, whereas younger listeners only used F_0 information (Linville and Korabic, 1986), but older listeners were still found to be less accurate than younger listeners when performing an age estimation task. Thus, the combined evidence from past and present studies suggests that younger and older listeners may differ in terms of which auditory cues they rely on for judging age and gender, perhaps due to a combination of differences in social experience and age-related changes in hearing.

C. Correlations among speech and voice rating scales

The third hypothesis was that the clarity of articulation would not be significantly correlated with the pleasantness of speech, as past studies have found that the quality and the intelligibility of talkers can vary independently. In line with this hypothesis, we found that the pleasantness of speech samples was more closely related to naturalness and to the talker's suitability as an audiobook reader than to clarity of articulation or ease of understanding. However, it should be noted that ease of understanding did not vary much across talkers, likely due to the quiet listening conditions in which the speech samples were presented, the talkers' homogeneity with regards to dialect and health status, and the recording of read rather than spontaneous speech. It is possible that the listeners' perceptions of which talkers speak clearly or are easy to understand would be different in a listening environment that more closely approximates naturalistic listening conditions (for instance, with background noise, accented talkers or talkers with pathologies) insofar as talkers' perceived intelligibility in quiet may not correspond well to their actual intelligibility in a more challenging listening environment.

The varying relationships between different rating scales in the present study provides insights into whether the results of past studies may be compared depending on which scales were used (for instance, it may be more reasonable to compare clarity of articulation as a measure of talker quality with ease of understanding than with ratings of pleasantness).

It is noteworthy that the relationships between scales differed depending on whether listeners were younger or older. Pleasantness was associated with naturalness, loudness and suitability as an audiobook reader in the same way for both younger and older listeners. For younger listeners, the perceived age of the talker was associated with speech qualities (pleasantness and loudness), whereas for older listeners, perceived age was associated with voice quality (roughness). While younger and older listeners had comparable intra-rater agreement for speech scales, younger listeners had slightly lower intra-rater agreement than older listeners for voice scales, which may have contributed to the non-significant correlations for voice rating scales for younger listeners (e.g., between perceived age and roughness). Nevertheless, as also discussed in Sec. III B for age and gender ratings, it is possible that younger and older listeners used different information from speech and voice samples to evaluate some aspects of a talker.

Another point about the present pattern of findings is that speech and voice ratings were significantly associated with each other under specific circumstances. Specifically, speech and voice pleasantness ratings for the same talkers were significantly correlated only when listeners rated talkers in their own age group. In other words, despite perceiving younger talkers to be somewhat older and older talkers to be somewhat younger than their true ages, listeners showed greater consistency in rating the same talkers when evaluating their peers than when evaluating talkers of a different age group, which suggests that some effect of familiarity or same-age bias was present. This finding is in line with previous findings that listeners are sometimes more accurate in gauging a particular characteristic (e.g., age) of talkers within their peer group compared to talkers outside the group (Huntley *et al.*, 1987).

In general, the findings show that when comparing listeners' intra-rater reliability for speech samples to their reliability for voice samples, listeners in the present study were more reliable for speech samples than for voice samples. Furthermore, when listening to the same talker's speech and voice, listeners' age estimations and gender identification were more accurate with speech samples than with voice samples, demonstrating a benefit from the greater amount of information in speech samples relative to that of voice samples, similar to past findings (Ramig, 1986).

D. Correlations between speech and voice ratings and acoustic measures

The fourth hypothesis was that some of the rating scales would be related to acoustic measures; specifically, that the perceived roughness of voice samples would be significantly correlated with perturbation measures. However, linear models showed that listeners' perception of vocal roughness was

not predicted by either jitter or shimmer. In fact, speech and voice quality ratings were generally not well-predicted by any acoustic measure except for speech rate (all listeners perceived faster talkers to be younger) and F_0 (a higher F_0 was associated with younger perceived age, and also with smoother voices in female talkers). The slowing of speech rate with age has been objectively measured across a variety of speaking tasks (Duchin and Mysak, 1987), and when the speech of young adults was manipulated in one study to have a slower rate, listeners' age estimations of talkers increased (Harnsberger *et al.*, 2008), showing that listeners (including those in the present study) use speech rate as a valid cue for age. Although conclusions about age-related changes in F_0 in males are inconsistent across studies (Goy *et al.*, 2013), there is certainly a large amount of evidence showing an age-related decline in F_0 in females, and it seems that listeners in the present study used this F_0 cue appropriately. Although both younger and older listeners in the present study used speech rate to estimate the age of talkers, only older listeners used shimmer as a cue for estimating the age of male talkers from voice samples and F_0 as a cue for estimating the age of female talkers in running speech. The age-related increase in vocal shimmer in males and the lowering of F_0 in females have been shown in a previous study (Goy *et al.*, 2013), demonstrating again that older listeners used a larger variety of valid age-relevant information to evaluate talkers compared to younger listeners.

V. CONCLUSIONS

In the present study, younger and older listeners rated the speech and voice qualities of healthy younger and older talkers. The age of listeners had a somewhat larger effect on the perception of quality than on the perception of the age of talkers, and the extent of age-related differences between listener groups was larger when listeners heard voice samples than when they heard speech samples. Younger and older listeners seemed to differ on the kind of information they derived from speech and voice samples for the purposes of making judgments about talkers, resulting in better performance by older adults than younger adults on gender identification, but relatively poorer performance by older adults on age estimation. Younger listeners also showed a negative bias towards older voices, whereas older listeners did not. Such interactions between effects due to the age of listeners and effects due to the age of talkers may explain some of the inconsistencies in the literature on the perception of speech and voice qualities in younger and older adults. These findings also further our understanding of how characteristics of a listener may subtly influence their perception of a talker and, in turn, possibly alter their social interactions.

ACKNOWLEDGMENTS

The authors wish to thank James Qi for technical assistance and Sumaiya Farooq and Ivian Tchakarova for their assistance in stimuli preparation and data collection. This project was funded by grants from the Natural Sciences and Engineering Research Council of Canada awarded to M. K. P.-F. (No. 138472) and P. v. L. (No. 312308), a Canada Research Chair grant awarded to P. v. L. (No. 950-213162), and the University of Toronto Mississauga Work-Study Program.

Amerman, J. D., and Parnell, M. M. (**1990**). "Auditory impressions of the speech of normal elderly adults," Br. J. Disord. Commun. **25**, 35–43.
Bele, I. V. (**2005**). "Reliability in perceptual analysis of voice quality," J. Voice **19**, 555–573.
Boersma, P., and Weenink, D. (**2014**). "Praat: Doing phonetics by computer," http://www.praat.org (Last viewed December 28, 2014).
Bruckert, L., Liénard, J.-S., Lacroix, A., Kreutzer, M., and Leboucher, G. (**2006**). "Women use voice parameters to assess men's characteristics," Proc. R. Soc. London, Ser. B: Biol. Sci. **273**, 83–89.
Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., and Kent, J. F. (**2007**). "Listener agreement for auditory-perceptual ratings of dysarthria," J. Speech Lang. Hear. Res. **50**, 1481–1495.
Cohen, G., and Faulkner, D. (**1986**). "Does 'elderspeak' work? the effect of intonation and stress on comprehension and recall of spoken discourse in old age," Lang. Commun. **6**, 91–98.
Dilley, L. C., Wieland, E. A., Gamache, J. L., McAuley, J. D., and Redford, M. A. (**2013**). "Age-related changes to spectral voice characteristics affect judgments of prosodic, segmental, and talker attributes for child and adult speech," J. Speech Lang. Hear. Res. **56**, 159–177.
Duchin, S. W., and Mysak, E. D. (**1987**). "Disfluency and rate characteristics of young adult, middle-aged, and older males," J. Commun. Dis. **20**, 245–257.
Eadie, T. L., and Doyle, P. C. (**2002**). "Direct magnitude estimation and interval scaling of pleasantness and severity in dysphonic and normal speakers," J. Acoust. Soc. Am. **112**, 3014–3021.
Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Martinez, A. S. (**1998**). "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," J. Speech Lang. Hear. Res. **41**, 327–339.
Fagel, W. P. F., Van Herpt, L. W. A., and Boves, L. (**1983**). "Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation," Speech Commun. **2**, 315–326.
Fairbanks, G. (**1960**). *Voice and Articulation Drillbook*, 2nd ed. (Harper, New York).
Gabrielsson, A., and Sjögren, H. (**1979**). "Perceived sound quality of hearing aids," Scand. Audiol. **8**, 159–169.
Gatehouse, S., and Akeroyd, M. (**2006**). "Two-eared listening in dynamic situations," Int. J. Audiol. **45**, 120–124.
Gelfer, M. P. (**1988**). "Perceptual attributes of voice: Development and use of rating scales," J. Voice **2**, 320–326.
Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., and Van Lieshout, P. (**2013**). "Normative voice data for younger and older adults," J. Voice **27**, 545–555.
Harnsberger, J. D., Brown, W. S., Jr., Shrivastav, R., and Rothman, H. (**2010**). "Noise and tremor in the perception of vocal aging in males," J. Voice **24**, 523–530.
Harnsberger, J. D., Shrivastav, R., Brown, W. S., Jr., Rothman, H., and Hollien, H. (**2008**). "Speaking rate and fundamental frequency as speech cues to perceived age," J. Voice **22**, 58–69.
Hartman, D. E., and Danhauer, J. L. (**1976**). "Perceptual features of speech for males in four perceived age decades," J. Acoust. Soc. Am. **59**, 713–715.
Hillenbrand, J. (**1988**). "Perception of aperiodicities in synthetically generated voices," J. Acoust. Soc. Am. **83**, 2361–2371.
Hollien, H., Pausewang Gelfer, M., and Carlson, T. (**1991**). "Listening preferences for voice types as a function of age," J. Commun. Disord. **24**, 157–171.
Howell, P., and Kadi-Hanifi, K. (**1991**). "Comparison of prosodic properties between read and spontaneous speech material," Speech Commun. **10**, 163–169.
Hummert, M. L., Mazloff, D., and Henry, C. (**1999**). "Vocal characteristics of older adults and stereotyping," J. Nonverbal Behav. **23**, 111–132.
Huntley, R., Hollien, H., and Shi, T. (**1987**). "Influences of listener characteristics on perceived age estimations," J. Voice **1**, 49–52.
Jacques, R. D., and Rastatter, M. P. (**1990**). "Recognition of speaker age from selected acoustic features as perceived by normal young and older listeners," Folia Phoniatr. **42**, 118–124.
Kates, J. M., and Kozma-Spytek, L. (**1994**). "Quality ratings for frequency-shaped peak-clipped speech," J. Acoust. Soc. Am. **95**, 3586–3594.

Kemper, S., Ferrell, P., Harden, T., Finter-Urczyk, A., and Billington, C. (**1998**). "Use of elderspeak by young and older adults to impaired and unimpaired listeners," Aging Neuropsychol. Cog. **5**, 43–55.
Klofstad, C. A., Anderson, R. C., and Peters, S. (**2012**). "Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women," Proc. R. Soc. London, Ser. B: Biol. Sci. **279**, 2698–2704.
Laan, G. P. M. (**1997**). "The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style," Speech Commun. **22**, 43–65.
Linville, S. E. (**1996**). "The sound of senescence," J. Voice **10**, 190–200.
Linville, S. E., and Fisher, H. B. (**1985**). "Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females," J. Acoust. Soc. Am. **78**, 40–48.
Linville, S. E., and Korabic, E. W. (**1986**). "Elderly listeners' estimates of vocal age in adult females," J. Acoust. Soc. Am. **80**, 692–694.
Maryn, Y., and Debo, K. (**2014**). "Is perceived dysphonia related to perceived healthiness?," Logoped. Phoniatr. Vocol. **40**, 122–128.
McAleer, P., Todorov, A., and Belin, P. (**2014**). "How do you say 'Hello'? Personality impressions from brief novel voices," PLoS One **9**, e90779.
Medrado, R., Ferreira, L. P., and Behlau, M. (**2005**). "Voice-over: Perceptual and acoustic analysis of vocal features," J. Voice **19**, 340–349.
Montepare, J. M., Kempler, D., and McLaughlin-Volpe, T. (**2014**). "The voice of wisdom: New insights on social impressions of aging voices," J. Lang. Soc. Psychol. **33**, 241–259.
Mueller, P. B. (**1997**). "The aging voice," Sem. Speech Lang. **18**, 159–168.
Mulac, A., and Giles, H. (**1996**). "'You're only as old as you sound': Perceived vocal age and social meanings," Health Commun. **8**, 199–215.
Munro, M. J., and Derwing, T. M. (**1999**). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," Lang. Learn. **49**, 285–310.
Preminger, J. E., and Van Tasell, D. J. (**1995**). "Quantifying the relation between speech quality and speech intelligibility," J. Speech Hear. Res. **38**, 714–725.
Ramig, L. A. (**1986**). "Aging speech: Physiological and sociological aspects," Lang. Commun. **6**, 25–34.
Rothman, H. B., Brown, W. S., Jr., Sapienza, C. M., and Morris, R. J. (**2001**). "Acoustic analyses of trained singers perceptually identified from speaking samples," J. Voice **15**, 25–35.
Ryan, E. B., Bourhis, R. Y., and Knops, U. (**1991**). "Evaluative perceptions of patronizing speech addressed to elders," Psychol. Aging **6**, 442–450.
Ryan, E. B., and Capadano, H. L. III (**1978**). "Age perceptions and evaluative reactions toward adult speakers," J. Gerontol. **33**, 98–102.
Ryan, E. B., Giles, H., Bartolucci, G., and Henwood, K. (**1986**). "Psycholinguistic and social psychological components of communication by and with the elderly," Lang. Commun. **6**, 1–24.
Ryan, E. B., and Laurie, S. (**1990**). "Evaluations of older and younger adult speakers: Influence of communication effectiveness and noise," Psychol. Aging. **5**, 514–519.
Schwartz, M. F., and Rine, H. E. (**1968**). "Identification of speaker sex from isolated, whispered vowels," J. Acoust. Soc. Am. **44**, 1736–1737.
Scherer, K. R. (**1995**). "Expression of emotion in voice and music," J. Voice **9**, 235–248.
Shrout, P. E., and Fleiss, J. L. (**1979**). "Intraclass correlations: Uses in assessing rater reliability," Psychol. Bull. **86**, 420–428.
Singh, S., and Murry, T. (**1978**). "Multidimensional classification of normal voice qualities," J. Acoust. Soc. Am. **64**, 81–87.
Vongpaisal, T., and Pichora-Fuller, M. K. (**2007**). "Effect of age on F_0 difference limen and concurrent vowel identification," J. Speech Lang. Hear. Res. **50**, 1139–1156.
Zäske, R., and Schweinberger, S. R. (**2011**). "You are only as old as you sound: Auditory aftereffects in vocal age perception," Hear. Res. **282**, 283–288.