Published in final edited form as:

J Commun Disord. 2017 July ; 68: 103–112. doi:10.1016/j.jcomdis.2017.06.001.

Listener Estimates of Talker Age in a Single-talker, 50-year Longitudinal Sample

Eric J. Hunter¹ and Sarah Hargus Ferguson²

¹Department of Communicative Sciences and Disorders Michigan State University, East Lansing MI

²Department of Communication Sciences and Disorders University of Utah, Salt Lake City, UT

Abstract

Purpose: While many studies have shown that listeners make relatively accurate age judgments from hearing talkers' voices, most have used just one age per talker. The present study evaluated listeners' age estimation abilities for a longitudinal sample: a single talker recorded over nearly five decades.

Method: We extracted 60 samples from addresses given by a male talker over 48 years. For each sample, listeners gave a direct estimate of his age. We also asked listeners whether they were able to identify the talker, who was known locally, from the recordings.

Results: While correlations between the talker's chronological age and direct estimates of age were generally strong, the relationship was particularly strong when the talker was older than 68 years, although listeners underestimated chronological age by approximately 5 years. When the talker was between 49 and 68 years, listeners' direct age estimates were less accurate and less strongly correlated with his chronological age. Additionally, direct age estimates were more accurate when listeners recognized the talker.

Conclusions: Corroborating previous cross-sectional studies, our listeners overestimated the single talker's age when he was younger and underestimated it when he was older. However, the crossover between overestimation and underestimation, or the point at which estimated and chronological ages were equal, occurred at a later talker age in the current longitudinal study than in previous cross-sectional studies. Additionally, listeners who recognized the talker made significantly more accurate age estimates in both experiments. We propose additional studies to explore the relationship between talker familiarity and estimated age.

Keywords

Age-related speech changes; Age estimation; Longitudinal speech

1 Introduction

As the Baby Boom generation has gotten older, more research has focused on age-related health issues. Age-related changes to the laryngeal system, while not typically life-threatening, have become increasingly important. Beginning in the fifth decade, specific degenerative processes affect the speech production system. These physiological changes (e.g., atrophy of muscles, dystrophy of nerves and cells, edema) cause speech production to deteriorate, affecting quality of life (Golub, Chen, Otto, Hapner, & Johns, III, 2006; Schneider, Plank, Eysholdt, Schutzenberger, & Rosanowski, 2011). Such deterioration includes dysphonia (Golub et al., 2006); slower speech and elongated words (Ramig, 1983a; Harnsberger, Shrivastav, Brown, Jr., Rothman, & Hollien, 2008); tremor, instability, and aspiration (Ramig, 1983b; Linville, 1988; Harnsberger, Brown, Jr., Shrivastav, & Rothman, 2010); reduced loudness (Ramig & Ringel, 1983; Baker, Ramig, Sapir, Luschei, & Smith, 2001); and voice fundamental frequency changes (Baken, 2005; Stathopoulos, Huber, & Sussman, 2011). Further, decreased pulmonary function may cause a weaker voice and shortened phrase lengths (Huber & Spruill, 2008).

Previous studies have examined whether listeners can perceive these age-related speech changes in physiology and production, as well as whether they are capable of making accurate judgments of age from hearing a talker's voice. These studies have varied in terms of both talker characteristics (e.g., gender, age range) and listener characteristics (e.g., gender, age, number) as well as in the speech materials the talkers produced (e.g., sustained vowels, spontaneous speech) and the perceptual tasks the listeners performed (e.g., category judgments, direct age estimates). Among studies in which listeners made category judgments, designs have ranged from two age categories (under 35 and over 65; Ptacek and Sander, 1966) to six age categories each spanning 5 years (Neiman and Applegate, 1990). Despite this, category judgments of age consistently have been more accurate for connected speech materials than for vowels. Furthermore, most category judgment studies (e.g., Linville & Fisher, 1985; Goy, Fernandes, Pichora-Fuller, and van Lieshout, 2016) have shown that listeners tend to underestimate the age of older talkers, with few exceptions (i.e., Ptacek & Sander, 1966). Caruso, Mueller, and Xue (1994) also found that their two older talkers were much more likely to be judged as young or middle-aged than to be judged as old.

Age underestimation for older talkers has also been evident in direct age estimation studies, most of which have covered several decades of talker ages (Shipp & Hollien, 1969; Ryan & Capadano, III, 1978; Hartman, 1979; Huntley, Hollien, & Shipp, 1987; Harnsberger et al., 2010). Hunter, Newman, and Ferguson (2015) conducted a retrospective data analysis of direct age estimate studies of English-speaking talkers and listeners. First, they identified seven studies that provided data in tables or scatter plots rather than just reporting means. They then used MATLAB to capture, verify, and recreate data provided in scatter plots from digital scans of the articles and pooled them with data provided in tables. In this meta-analysis, which yielded 530 data points for talkers aged 10–90 years, listeners seemed to overestimate age by about 5–7 years until talkers reached age 40. After talkers reached age 50, listeners underestimated age by about 5 years. This underestimation gradually increased until about age 80, at which point listeners underestimated talker age by 12 years.

Nevertheless, results varied among individual studies. For example, for talkers aged roughly 70–80 years, listeners underestimated talkers' age by about 8 years in Huntley et al. (1987) and by about 11 years in Moran, McCloskey, and Cady (1995), but by only about 5 years in Ryan and Burk (1974)¹. Eppley and Mueller (2001) explored possible causes of this underestimation by including both younger and older listeners but only older talkers. Since both listener groups underestimated the age of these older talkers by nearly 17 years, they concluded that the underestimation seen in other studies did not result from including only young listeners, nor was it due to regression to the mean. This suggested that the age underestimation of older talkers is a real phenomenon and not an artifact of experimental protocols or analysis techniques.

For all of the methodological variability among the previous studies, in nearly all of them the listeners were asked to identify each talker's age based on a single age sample (i.e., they used a cross-sectional design). While such designs enable general trends to be observed, longitudinal studies allow us to track age-dependent characteristics without the effects of talker variability (Linville, 1988). After a detailed search, only two age estimation studies which included longitudinal elements were found. In Brückl (2007), listeners heard recordings of 9 talkers who had been recorded on two occasions separated by 5 years; they were able to discriminate between samples produced by the same talker 5 years apart about two-thirds of the time. In another study, Reubold, Harrington, and Kleber (2010) used phrases extracted from broadcast recordings of two prominent individuals as well as manipulated, resynthesized versions of these phrases to assess the effects of age-related speech fundamental frequency and first formant changes on age estimation. For both individuals (Queen Elizabeth, recorded at ages 46 and 57 years, and Alistair Cooke, recorded at ages 39, 62, and 82 years), listeners underestimated the talkers' ages when they were older. These studies went beyond the traditional cross-sectional report of changes with age and added valuable within subject age changes, though the time points were limited to just a few ages.

The dearth of longitudinal age estimation studies is not surprising given that there have actually been very few longitudinal speech-related studies of any sort. A rare exception to this is an early speech aging study which tracked the vowel formants of seven talkers over as many as 29 years (Endres, Bambach, & Flosser, 1971). One barrier to such studies is the limited availability of longitudinal speech databases. Harrington, Palethorpe, and Watson (2000) addressed this dilemma by using 50 years of recorded Christmas broadcasts given by Queen Elizabeth II to examine age-related changes in the voice. Reubold et al. (2010) revisited these data, adding recorded broadcasts from four additional well-known speakers. However, the brief length of the broadcasts (1 m 16 s to 13 m 24 s) limits their usefulness: it is unknown whether such brief segments would be representative of longer speech samples or whether they might reflect warm-up effects. Further, with such well-known figures as Queen Elizabeth or Margaret Thatcher, listeners' familiarity may affect their age estimates.

¹We note that the original articles did not segment the results by age range. These estimates were approximated from tables and figures provided in the original articles.

Building on these previous attempts to use prerecorded public addresses as a listening database, the present study used recorded addresses given by a single talker over a period spanning nearly 50 years (ages 48–97). The goal was to examine listeners' ability to perceive within-talker age-related speech production changes. We then compared the accuracy with which listeners were able to estimate the talker's age to that observed in the cross-sectional age perception literature. Additionally, because the talker was a known individual to a significant proportion of the population local to the location of the experiment (as discussed in more detail in the Methods section), we were also able to discuss the potential effects of talker recognition on age estimation. The public addresses used here have been described in Hunter, Kapsner-Smith, Pead, Engar, and Brown (2012) and in Hunter and Titze (2016), both of which examined changes in vocal characteristics with age. Additionally, the public addresses were 25 to 50 min in length, allowing us to extract multiple samples from each recording.

The research question for the current study is, "How do chronological age estimates in a longitudinal sample compare to those observed in cross-sectional studies?" We hypothesized that our listeners' estimates of the talker's chronological age would be similar to those seen in a cross-sectional design. Additionally, we hypothesized that familiarity with the talker would improve age estimations. We expected that the listeners who recognized the talker would remember him being a very old man, and would accordingly give estimates reflecting this memory (i.e., judge him as being in his 80s and 90s for most of the samples). To facilitate comparison with the previous cross-sectional literature, we carried out a direct age estimate experiment similar to those found in the literature (Shipp and Hollien, 1969; Ryan and Capadano, 1978; Hartman, 1979; Huntley et al., 1987; Harnsberger et al., 2010; Moran et al., 1995; Eppley and Mueller, 2001).

2 Methods

Adult listeners listened to short speech samples. They then judged the talker's age using a direct age estimation experiment. The University of Utah Institutional Review Board approved all methodology.

2.1 Participants

We recruited 25 listeners from the University of Utah Department of Psychology Participant Pool. All subjects were college students with self-reported normal hearing. After testing, data were excluded from participants who were younger than 18 ($n = 1$) or older than 32 ($n = 3$), or who had not reported their age ($n = 1$). The perceived age data reported here thus came from 20 young adults (aged 18–32 years, mean age 22.8 years, 13 female).

2.2 Materials

Research assistants extracted sixty samples (described later) from recordings of addresses given over a period of almost 50 years. The talker was a lay leader of an international church who delivered frequent public addresses (both religious and academic in nature) in an enclosed university arena. We obtained audio recordings in MP3 format from a publicly available archive of these addresses. These recordings spanned the years 1958 (when the

talker was 48 years old) through early 2007 (when he was 97 years old). Additional details regarding the talker, recording procedures, and environment may be found in Hunter et al. (2012).

In addition to the longitudinal breadth of the recorded archive, several other characteristics make the talker and his body of addresses unique. First, the speaking style he used is comparable to that of the general population; unlike many individuals who give frequent public addresses, the subject apparently received no training as a public speaker. Further, although he was a religious leader, the addresses do not possess the traditional rhetorical characteristics of many sermons. Instead, he gave his addresses as lectures at a podium using a microphone with written notes and, in later years, a teleprompter. Second, the subject maintained a lifestyle that included abstinence from smoking, coffee, and alcohol. While this lifestyle is not representative of much of the aging population, the absence of these common vocal irritants removed variables that might exaggerate or interact with age-specific voice changes. Third, the acoustic environment was consistent for all of the addresses. The talker gave each address to a university audience in one of two settings used for sports and assemblies: a 5,000-seat multi-purpose arena used until 1971 and a 22,700-seat multi-purpose arena used after 1971. Fourth, all of the addresses were long enough (>20 min) to permit removal of the first and last 5 min while still providing a sustained representative speech sample (likely absent of warm-up or closing effects) for analysis.

Another methodological consideration arises from the talker's standing as a religious leader. In this role, he gave biannual addresses in a specifically religious setting, which were widely listened to by the worldwide membership of the church. Because of the proximity of the church's headquarters to the university, it is likely that makeup of the student population from which we recruited would include members of the talker's church. If this were true, such listeners would likely be familiar with his voice as it sounded when he was in his 80s and 90s.

We selected 60 segments from 20 addresses in the archive with the aim of sampling a single address every 2 to 3 years. The average difference in talker age between any two adjacent selected addresses was 2.5 years (s.d. 0.5 years; range: 1.7–3.4 years). The talker gave the earliest address chosen when he was 49 years old; he gave the latest one when he was 97 years old. Any address in which the talker mentioned factors that might have affected his voice (e.g., recovering from a cold, being fatigued from recent travel) was excluded from selection. Assistants excised three segments from the middle 5 min of each of the 20 selected addresses using Cool Edit 2000. Each segment lasted approximately 5 to 10 s, a sample length chosen to be comparable to several previous studies (e.g., Shipp & Hollien, 1969) that used one or two sentences excised from the Rainbow Passage. In addition, while studies have shown that context does not necessarily assist age estimation (Mulac & Giles, 1996), samples were selected to minimize content that would either identify the talker or connect the three segments within a given speech to each other. To minimize other potential clues that might influence perceived age, we chose segments that contained (1) at least one audible breath and (2) no noticeable disfluencies (e.g., voice breaks or restatements). Finally, we scaled all excised segments to have the same average RMS intensity.

2.3 Procedures

Listeners completed the protocol in a single test session during which they sat in a quiet room in front of a computer monitor, keyboard, and mouse. After completing consent procedures and a participant information sheet, the experimenter gave all listeners the following task instructions:

In this experiment, you will hear excerpts from speeches given by the same man over a period of 50 years. The first recordings were made when he was about 50 years old, and he was nearly 100 years old when the last ones were made. Your job is to listen to these excerpts and, for each one, estimate how old you think he was at the time of the speech.

Materials were played from one channel of a Tucker-Davis Technologies (TDT) RP2 real-time processor, attenuated to a comfortable level by a programmable attenuator (TDT PA5), and routed via a headphone buffer (TDT HB7) to Sennheiser circumaural earphones for presentation. Each listener heard the 60 segments in a different random order. For each trial, a segment was presented and the message, “How old do you think the person talking is?” was displayed on the computer monitor. Before responding, listeners had the option to click a box labeled “Listen Again?” to hear the segment one additional time. After the first two trials, the experimenter offered listeners the opportunity to raise or lower the presentation volume level.

The experimenter instructed listeners to guess the talker’s age in whole years, write their guess on a prepared answer sheet, and then press “Enter” on the keyboard to initiate the next trial. After completing the block of 60 trials, the tester offered listeners a short break before completing the block a second time. Upon completion of the second test block, listeners completed a short debriefing questionnaire comprised of the following four questions:

1. How difficult did you find the listening task you just completed?
2. How confident were you in the judgments you were asked to make?
3. What did you listen for to make these judgments?
4. Did you recognize the talker? If so, who do you think it was?

For the present analyses, we only used Question 4.

2.4 Data Analysis

We analyzed data using mixed-effects models, which have several advantages over analyses of variance (Quené & van den Bergh, 2004) and are especially well-suited to studies with repeated-measures designs. Within the experiment, the perceived aged judgments made for the individual segments were likely to covary, as were the judgments made by individual listeners. We therefore modeled segment and listener effects as nested random effects to account for this non-independence. We performed all analyses using Stata 14.1 (StataCorp, 2015).

3 Results

We determined within-listener reliability by computing pairwise correlations between the first and second age estimates that each listener gave for each segment. These correlations were moderate and positive ($r = .63$ to $.88$) for all but one listener, for whom $r = .04$. We excluded data from this listener from further analysis, leaving data from 19 listeners. Responses to the debriefing questionnaire showed that 10 of these listeners recognized the talker while 9 did not, making it feasible for to assess familiarity effects for the direct age estimate data.

Because direct age estimates are ratio data, we were able to analyze the age estimates assigned by each listener to each segment for each of the two test blocks using mixed-effects linear regression models. The first set of mixed-effects models included the talker's chronological age, listeners' familiarity with the talker, and the interaction between these two factors as fixed effects and segment and listener as random effects. Including these random variables significantly improved the fit of the model ($\chi^2 = 711.69$, $p < .001$). Talker age, listener familiarity, and the interaction between them were all significant ($z = 14.32$, $p < .001$; $z = 8.38$, $p < .05$; and $z = 5.76$, $p < .001$, respectively). Figure 1 illustrates the relationship between chronological age and direct age estimates; part (a) of the figure illustrates the average estimated age for the two listener groups for each of the three segments. Pairwise Pearson's correlations between chronological age and estimated age were significant ($p < .001$) for both listener groups, but they were stronger for the listeners who recognized the talker ($r = .75$) than for those who did not ($r = .62$). We found that second-order polynomials fit the data better than simple regression lines, and show these in Figure 1a; both resulting R^2 values are greater than 0.8.

In Figure 1b, which illustrates the average estimated age as a function of talker age decade and listener group, age estimates were roughly constant when the talker was younger, only increasing in his 70s for listeners who recognized him and in his 80s for listeners who did not. Mixed-effects linear regression analyses testing the effects of age decade, listener familiarity, and the interaction between them (with segment and listener as random effects) supported this impression. For listeners who did not recognize the talker, age estimates for segments produced when the talker was in his 50s, 60s, and 70s did not differ significantly from one another ($z = 0.81$ to 1.28 , $p = .06$ to $.415$). Age estimates that listeners who did not recognize the talker made for segments taken from his 80s and 90s, in contrast, differed significantly from each other ($z = 5.92$, $p < .001$) and from the estimates for the earlier decades ($z = 6.67$ to 13.94 , all $p < .001$). For listeners who did recognize the talker, age estimates for segments produced when he was in his 50s and 60s were not significantly different from each other ($z = .83$, $p = .404$), but were significantly lower than estimates for segments taken from the later three decades ($z = 4.09$ to 16.2 , all $p < .001$). Age estimates for segments the talker produced in his 70s, 80s, and 90s also differed significantly from each other ($z = 6.02$ to 10.81 , all $p < .001$) for the listeners who recognized him.

We assessed accuracy for each segment, listener, and observation by subtracting the estimated age from the talker's chronological age. The direction of the error scores varied as a function of talker age, with listeners overestimating the talker's age when he was younger

and underestimating it when he was older. Figure 2a. illustrates the accuracy of the direct age estimates, showing the signed difference (triangles) and absolute difference (squares) in years between the estimates and the talker age for all listeners (both were fit to a 2nd order polynomial). To simplify the interpretation of the age effect, we broke down talker age into decades for the accuracy analysis; Figure 2b. shows average signed difference scores for the two listener groups as a function of talker age decade. We analyzed these signed difference scores using mixed-effects linear regression models assessing the fixed effects of decade and listener familiarity, as well as the interaction between them. All models included segment and listener as random effects. The interaction between talker age decade and listener familiarity was significant ($z = 4.28, p < .001$), so a stratified analysis was carried out to examine the effect of familiarity for each decade and the effect of age decade for listeners who did and did not recognize the talker.

When the talker was in his 50s and 60s, accuracy was similar for the two listener groups ($z = .50, p = .61$ and $z = 1.17, p = .24$, respectively). However, when the talker was older (in his 70s, 80s, and 90s), signed difference scores were significantly more negative (i.e., age was underestimated to a greater extent) for listeners who did not recognize the talker than for listeners who did (respectively, $z = 5.49, 6.54$, and 6.82 , respectively, all $p < .001$). This means that listeners who recognized the talker estimated his age more accurately than listeners who did not recognize him. Signed difference scores were highest when the talker was in his 50s than for any other age decade ($z > 7.4, p < .001$), with both groups overestimating his age by at least 10 years. For listeners who did not recognize the talker, accuracy was best when he was in his 60s, poorest when he was in his 50s and 70s but in opposite directions, and intermediate when he was in his 80s and 90s. Signed difference scores for this listener group did not differ significantly between the talker's 70s, 80s, and 90s ($z = .05$ to $1.35, p = .17$ to $.964$), but difference scores for his 60s were smaller than for all other decades ($z = 2.52$ to 3.77 , all $p < .05$). For listeners who did recognize the talker, signed difference scores were small in magnitude and did not differ statistically from each other in the talker's 60s, 70s, 80s, and 90s ($z = 0.02$ to $1.17, p = .24$ to $.98$).

4.0 Discussion

In the present study, listeners judged the age of a single talker based on speech samples excerpted from speeches given over a period of nearly 50 years. Direct age estimations showed generally equal estimates of age until a chronological age of around 68 years, where the pattern changes slightly. Specifically, the results indicate a piecewise pattern in which accuracy improved as the talker grew older, as seen by the polynomial fit line. When the talker was between 49 and 68 years old, listeners estimated him to be between 58 and 68, with no apparent directional trend during this two-decade period (Figure 1). Estimates in this period were usually higher than the talker's actual age, overestimating his age by 6 years on average. Once the talker reached age 68, however, the estimates began to trend with the age of the talker, albeit underestimating his age by about 5 years (Figure 1a).

In a study analyzing the effects of age on acoustic and other speech production characteristics for this talker, Hunter et al. (2012) showed that voice fundamental frequency decreased 25 Hz from age 48 to about age 64 and increased about the same amount between

ages 72 and 92. Additionally, the length of breath groups when speaking was consistent at an average of 10 words per breath group until about age 66, and then began steadily decreasing. Hunter and Titze (2016) saw similar results in a new single-talker longitudinal sample (52–90 y/o), suggesting the possibility of a fundamental change in the body's maintenance of the speech mechanism occurring sometime between the ages of 68–74. Hunter and Titze also showed that voice breaks increased along with voice pitch, potentially indicating an interaction between the subglottal and supraglottal airway and increased weakness of the vocal folds. This similarity between these speech production and the current perceived age data, where both show a change at around age 68, suggests a relationship between listener age estimates and talker speech fundamental frequency and breath group length. However, whether the relationship was causal or merely a co-occurrence was not tested. Furthermore, note that both Hunter et al. (2012) and Hunter and Titze (2016) analyzed a larger number of speeches than the present study used and also analyzed 5-minute samples which began approximately 5 minutes into the speech, while we took our short segments from the middle 5 minutes of each speech.

As a quick comparison between the acoustic characteristics of our test segments and estimated age, we carried out a brief analysis of the 60 audio samples (a more thorough and complete analysis of the audio will be left to future studies). For each sample, we used custom Praat (Boersma & Weenink, 2016) scripts to extract speaking fundamental frequency (SFF) at 10 msec intervals. Figure 3 illustrates how the SFF of each sample relates to the chronological age of the talker. As was seen in Hunter et al. (2012) and Hunter and Titze (2016), the mean SFF of our test segments appears to slowly decrease with age until age 70, then slowly increase over the next several decades. The linear fit shows that there is a general trend upward (more than 10 Hz) over the course of the recordings. We then conducted a simple cross correlation analysis in Excel between the SFF and the actual age, estimated age, and the signed and absolute difference (e.g. Figure 2a.). As seen in Table 1, the correlation between estimated age and SFF was not strong ($r=0.25$). This suggests that the listeners likely received some information about age from the SFF, but only a small part.

Previous studies have found several other features to be important in judging age, including vocal tract resonances, voice quality, tremor, speaking rate, and articulatory precision (Harnsberger et al., 2010; Harnsberger, Shrivastav, Brown Jr., Rothman, & Hollien, 2008; Sue Ellen Linville, 1988; L. A. Ramig, 1983; Lorraine A. S1_Reference19Ramig, 1983). The listeners here likely made use of several of these, but the goals of the present study did not include determining which acoustic cues the listeners used. Further studies on the longitudinal dataset are planned that will investigate some of these other components.

The crossover point in the present single-talker data was higher (around age 68) than was found in the Hunter et al. (2015) meta-analysis described earlier. There are several possible explanations for this difference. First, our instructions to listeners stated that the recordings were made when the talker was 50 to 100 years old. Three of the studies reviewed in Hunter et al. (Moran et al., 1995; W. J. Ryan & Burk, 1974; Shipp & Hollien, 1969) gave listeners information about the talkers' age ranges, while the others did not. The age overestimations we observed when the talker was in his 50s and 60s could thus reflect a floor effect, which could have shifted the crossover up. However, while some amount of floor effect (and

potentially ceiling effect) may have contributed to the higher crossover point between overestimation and underestimation of the talker's age, it was probably not the primary factor. The absence of a ceiling effect in our data also argue against the presence of a floor effect. While normal life expectancy would likely produce a natural ceiling effect (most would not judge a talker much older than 100), there was no indication of such in either the familiar or unfamiliar group.

A second possible explanation for the higher crossover point observed here is the setting in which the speeches were recorded. The talker in the current study was speaking in an auditorium in a public setting, which may have elicited a voice different from the conversational or laboratory voice samples elicited in previous studies. In addition, when the talker was younger, he may have adopted a presentation-style speaking voice intended to make him sound older (and perhaps more authoritative) than he was. A third possibility is that while the present talker's voice may fall within the range of a cross-sectional average of multiple talkers using non-performance voices (as with any single-talker study), it is nevertheless skewed to one end of the range and not easily generalizable. In this scenario, while the talkers in a given cross-sectional sample may be normal speakers, no one individual talker may actually represent average speech for the group.

Finally, it is possible that some of the listeners who recognized the talker were not only familiar with the individual, but very familiar. The average age of the listeners was 23 years (making them on average, approximately 17 at the time of the talker's death), so those who recognized him and were familiar with his voice may have personal memories of him speaking in the later years of his life. It is possible that this familiarity made their judgments more accurate during those later years (where they had direct memory) than when he was younger. A related possibility is that the familiar listeners knew that the talker was active well into his 90s and thus were comfortable using higher ages in their estimates. However, if this were the case, then we would expect the familiar listeners to have better accuracy in the talker's 90s than in the ages associated with more common life expectancy (e.g. 70–90). In contrast, familiar listeners' accuracy was relatively stable when the talker was in his 60s and older.

It is important to note a couple of caveats regarding how we assessed familiarity with the talker. Our debriefing questionnaire asked listeners whether they recognized the talker only after they had completed the listening task, so it is unknown whether listeners recognized him immediately or at some later point in the listening task. Given the binary nature of the question, it is also unknown just how familiar they were with the talker: we placed any listener who said they recognized the talker and also named him correctly in the "familiar" group. Because of this, further speculation about familiarity may not be prudent. Finally, it is possible that some listeners found the talker familiar but did not actually recognize the voice; such error would add some additional uncertainty to the data. Future studies are planned that will combine the present longitudinal samples with new, cross-sectional recordings in which talkers of various ages and familiarity estimate age of similar speech tokens. This will permit us to explore the effects of familiarity in more depth.

One final point about the representation of direct age estimation is in order. Traditionally when the difference between the chronological age and the estimated age has been compared (Figure 2a, triangles, dashed line), the crossover point where overestimation turns to underestimation (~64 years in this case) has traditionally been defined as the point of greatest accuracy. However, comparing listener responses not in terms of the signed difference but in terms of the absolute magnitude of the difference ($|\text{chronological} - \text{estimated}|$; Figure 2a, boxes, dotted line) offers a slightly different perspective on accuracy. The absolute difference (looking at the polynomial fit dotted line) is the smallest (i.e., the listeners were most accurate) between about 75–90 years, which is not at all where the signed differences is the nearest zero. Thus, while the signed differences show the direction of the estimation error and indicate the crossover point between overestimation and underestimation, the absolute differences show there is always some error (about 6 years) even at the crossover point.

Considering the present age estimates in relation to data on age-related vocal changes could help us identify which vocal characteristics best reveal the age of an individual. This has a variety of potential applications. For example, health care providers may notice a difference between a patient's perceived and chronological ages as they interact with and assess the health of an individual. A patient sounding older than his/her chronological age may be interpreted as a symptom of an underlying health condition, while patients who sound younger than their chronological age might be deemed to be in especially good health. Furthermore, linking estimated age to age-related acoustic changes would be a useful guide for the development of speech signal processing algorithms that estimate age or synthesize aged voices. The results of Harnsberger et al. (2010), who found that normal voices were consistently underestimated on average by 5.7 years, while disordered voices were overestimated by 7.4 years, support this line of thinking.

5.0 Conclusion

We undertook this study to test the hypothesis that perceived vocal age estimates for a longitudinal sample from a single talker would be similar to those found in previous cross-sectional studies that used multiple talkers. In general, the use of the longitudinal sample increases our knowledge of perceived chronological age by giving us a more individualistic look at perceived age than cross-sectional studies. While cross-sectional studies may be more easily generalized, they lack information about individual talkers, which could be useful in understanding individual differences in vocal aging, as well as how these may relate to other aging metrics. Hence, the addition of individual snapshots (from longitudinal studies) to the general broad view (from cross-sectional studies) provides a better insight into the vocal aging process.

Unique to the present study, our listeners estimated age for recordings of a single talker over nearly 50 years of aging. Additionally, the talker was someone recognizable by a portion of the listeners, who may have some memory of listening to the talker's voice before his death in 2008. As in previous studies, which used cross-sections of talkers to represent an age group, listeners overestimated age when the talker was younger and underestimated it when the talker was older. The crossover between overestimation and underestimation happened at

a later talker age than in previous studies using cross-sectional data. Further, the age estimations for the talker corresponded to acoustic changes described in Hunter et al. (2012) and Hunter and Titze (In Press), supporting the relationship between general aging of the speech production system and listeners' perceptions.

Although several possible explanations are entertained above, it is unknown why the results of the current study differed from those of previous studies. These explanations include the recording conditions and speaking style associated with public speeches as well as the possibility that the talker in the current sample was simply not representative of the talkers used in previous cross-sectional studies. Future studies designed to resolve these specific weaknesses are necessary to explore these possible explanations. Further, the fact that some listeners recognized the talker likely skewed the overall results from the average of the cross-sectional studies.

Future studies will compare the degree of variability in age estimations from longitudinal studies with those from cross-sectional studies to provide further understanding of this difference. In order to conduct such a comparison, additional cross-sectional studies, which provide details about accuracy and variability, are necessary. We expect that age estimates from longitudinal studies will be less variable, because cross-sectional designs by necessity include multiple talkers and all of their different speech production characteristics. Some of these talker-specific characteristics, such as voice quality or speaking rate, are known to be related to perceived age (e.g., Shipp & Hollien, 1969; Harnsberger et al., 2010). In contrast, while their phonetic content may vary, the only talker-related difference among the samples in longitudinal studies like the present investigation is chronological age and the resulting voice changes.

Finally, future studies are planned that will combine the present longitudinal samples with new, cross-sectional recordings in which talkers of various ages read portions of speeches given by the longitudinal talker when he was the same age, with the instructions to read as if they were giving a speech. This will permit us to explore the effects of both talker idiosyncrasies and listener familiarity. Listeners of various ages prescreened for high familiarity with the present talker could be also be included to investigate whether age estimation accuracy is related to degree of familiarity. Finally, when dealing with longitudinal samples taken from public databases, there could be an influence of the recording equipment (e.g. microphone, media) from the time of the recording. Future studies could include preprocessing the samples using a spectral normalization which would level out effects from the equipment (Berardi, Hunter, & Gee, 2016).

Acknowledgments

We appreciate the work of several graduate students who assisted with this work: Catherine Newman assisted with the literature review, Patrick Pead gathered the original audio files and performed initial segmentation, and Catherine Mellum and Lydia Rogers assisted with the listening experiments as well as data entry and analysis. Some of the analysis and techniques used in support of this study were part of a larger study of gender differences in speech production, supported by the National Institute On Deafness And Other Communication Disorders of the National Institutes of Health under Award Number R01DC012315. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Berardi ML, Hunter EJ, & Gee KL (2016). Spectral-normalization filter for subjective analysis of the aging voice. *The Journal of the Acoustical Society of America*, 139(4), 1995–1995. 10.1121/1.4949850
- Boersma P, & Weenink D (2016). Praat: doing phonetics by computer (Version 6.0.19). Retrieved from <http://www.praat.org/>
- Brückl M (2007) Women's vocal aging: A longitudinal approach. *Interspeech*, Antwerp, Belgium. 1170–1173.
- Caruso AJ, Mueller PB, & Xue A (1994). Relative contributions of voice and articulation to listener judgements of age and gender: preliminary data and implications. *Voice*, 3, 1–9.
- Eppley BD, & Mueller PB (2001). Chronological age judgments of elderly speakers: the effects of listeners' age. *Contemporary Issues in Communication Science and Disorders*, 28, 5–8.
- Goy H, Pichora-Fuller MK, & van Lieshout P (2016). Effects of age on speech and voice quality ratings. *The Journal of the Acoustical Society of America*, 139(4), 1648–1659. [PubMed: 27106312]
- Harnsberger JD, Brown WS, Shrivastav R, & Rothman H (2010). Noise and Tremor in the Perception of Vocal Aging in Males. *Journal of Voice*, 24(5), 523–530. 10.1016/j.jvoice.2009.01.003 [PubMed: 19815378]
- Harnsberger JD, Shrivastav R, Brown WS, Jr., Rothman H, & Hollien H (2008). Speaking Rate and Fundamental Frequency as Speech Cues to Perceived Age. *Journal of Voice*, 22(1), 58–69. 10.1016/j.jvoice.2006.07.004 [PubMed: 16968663]
- Hartman DE (1979). The perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders*, 12(1), 53–61. [PubMed: 422747]
- Hunter EJ, Ferguson SH, & Newman CA (2015). Listener estimations of talker age: A meta-analysis of the literature. *Logopedics Phoniatrics Vocology*, 1–5. 10.3109/14015439.2015.1009160
- Hunter EJ, & Titze IR (2016). Progression of voice breaks in a non-pathological voice as an indicator of aerodigestive health. *Journal of American Geriatrics Society*, 64(10), 10.1111/jgs.14425.
- Huntley R, Hollien H, & Shipp T (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice*, 1(1), 49–52. 10.1016/S0892-1997(87)80024-3
- Linville SE (1988). Intraspeaker variability in fundamental frequency stability: An age-related phenomenon? *The Journal of the Acoustical Society of America*, 83(2), 741–745. 10.1121/1.396116 [PubMed: 3351132]
- Linville SE, & Fisher HB (1985). Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *The Journal of the Acoustical Society of America*, 78(1 Pt 1), 40–48. [PubMed: 4019906]
- Moran MJ, McCloskey L, & Cady B (1995). Listener age estimates of elderly African American and Caucasian male speakers. *Journal of Cross-Cultural Psychology*, 26(6), 751–758. 10.1177/002202219502600614
- Neiman GS, & Applegate JA (1990). Accuracy of listener judgments of perceived age relative to chronological age in adults. *Folia Phoniatrica*, 42(6), 327–330. [PubMed: 2081585]
- Ptacek PH, & Sander EK (1966). Age recognition from voice. *Journal of Speech and Hearing Research*, 9(2), 273–277. [PubMed: 5925532]
- Ramig LA (1983). Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders*, 16(3), 217–226. 10.1016/0021-9924(83)90035-7 [PubMed: 6875037]
- Ramig LA (1983). Effects of physiological aging on vowel spectral noise. *Journal of Gerontology*, 38(2), 223–225. [PubMed: 6827040]
- Reubold U, Harrington J, & Kleber F (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, 52(7–8), 638–651. 10.1016/j.specom.2010.02.012
- Ryan EB, & Capadano HL (1978). Age perceptions and evaluative reactions toward adult speakers. *Journal of Gerontology*, 33(1), 98–102. [PubMed: 618973]

- Ryan WJ, & Burk KW (1974). Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders*, 7(2), 181–192. [PubMed: 4841332]
- Shipp T, & Hollien H (1969). Perception of the aging male voice. *Journal of Speech and Hearing Research*, 12(4), 703–710. [PubMed: 5385720]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Relationship between chronological age and estimated age. The diagonal dotted line represents where chronological age and estimated age are equal. Both listener groups are represented; lightly filled indicates those subjects unfamiliar with the talker. (a) Average age estimate for each recording sample for both listener groups and fitted to a 2nd order polynomial. (b) Averages of estimated age by talker age decade for the two listener groups; error bars indicate the 95% Confidence Interval.

Figure 2. Accuracy of the direct age estimates. (a) Difference (triangles) and absolute difference (squares) in years for each of the three samples taken from each speech. Dotted lines represent second-order polynomial fits. (b) Difference in years for each of the three samples, averaged over talker age decades; error bars indicate the 95% Confidence Interval. (c) Absolute difference in years for each of the three samples, averaged over talker age decades; error bars indicate the 95% Confidence Interval.

Figure 3.
The speech fundamental frequency (Hz) of each sample over talker age. A linear fit line and a 2nd-order polynomial fit curve are also shown.

Table I.

Cross-correlation coefficient values between chronological age (CA), speech fundamental frequency (SFF), estimated age (EA), and signed and absolute differences (Diff, AbsDiff) between CA and EA.

	CA	SFF	EA	Diff	AbsDiff
CA	1.00				
SFF	0.17	1.00			
EA	0.88	0.25	1.00		
Diff	-0.72	0.03	-0.31	1.00	
AbsDiff	-0.47	0.01	-0.33	0.45	1.00