JSLHR

Research Article

Age Differences in Voice Evaluation: From Auditory-Perceptual Evaluation to Social Interactions

Catherine L. Lortie,[a,b,c] Isabelle Deschamps,[a,c] Matthieu J. Guitton,[b,c] and Pascale Tremblay[a,c]

Purpose: The factors that influence the evaluation of voice in adulthood, as well as the consequences of such evaluation on social interactions, are not well understood. Here, we examined the effect of listeners' age and the effect of talker age, sex, and smoking status on the auditory-perceptual evaluation of voice, voice-related psychosocial attributions, and perceived speech tempo. We also examined the voice dimensions affecting the propensity to engage in social interactions.
Method: Twenty-five younger (age 19–37 years) and 25 older (age 51–74 years) healthy adults participated in this cross-sectional study. Their task was to evaluate the voice of 80 talkers.
Results: Statistical analyses revealed limited effects of the age of the listener on voice evaluation. Specifically, older listeners provided relatively more favorable voice ratings than younger listeners, mainly in terms of roughness. In contrast, the age of the talker had a broader impact on voice evaluation, affecting auditory-perceptual evaluations, psychosocial attributions, and perceived speech tempo. Some of these talker differences were dependent upon the sex of the talker and his or her smoking status. Finally, the results also show that voice-related psychosocial attribution was more strongly associated with the propensity of the listener to engage in social interactions with a person than auditory-perceptual dimensions and perceived speech tempo, especially for the younger adults.
Conclusions: These results suggest that age has a broad influence on voice evaluation, with a stronger impact for talker age compared with listener age. While voice-related psychosocial attributions may be an important determinant of social interactions, perceived voice quality and speech tempo appear to be less influential.
Supplemental Materials: https://doi.org/10.23641/asha.5844102

Voice is the foundation of human verbal communication, carrying prosodic information and information about the talker, including his or her age and sex (Ptacek & Sander, 1966; Shipp & Hollien, 1969), and it is an important component of social interactions. Several studies have shown that listeners excel at identifying a talker's age (Harnsberger, Brown, Shrivastav, & Rothman, 2010; Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008; Huntley, Hollien, & Shipp, 1987; Ryan & Burk, 1974; Shipp & Hollien, 1969) and sex (Amir, Engel, Shabtai, & Amir, 2012; Schvartz & Chatterjee, 2012) on the basis of voice, when no other information is available. In addition to providing information about age and sex, the human voice provides information that is used by listeners to infer psychosocial traits about the talker, such as friendliness and trustworthiness (McAleer, Todorov, & Belin, 2014; Ryan & Johnston, 1987; Zuckerman & Driver, 1989). Interestingly, listeners show high consistency in their voice-related psychosocial attributions from short utterances containing limited information (Aronovitch, 1976; McAleer et al., 2014).

The attribution of psychosocial characteristics on the basis of a rapid evaluation of voice may be important in selecting approach/avoidance behaviors and could thus be a critical component of interpersonal relationships (Lallh & Rochet, 2000; McAleer et al., 2014; Mulac & Giles, 1996; Plank, Schneider, Eysholdt, Schützenberger, & Rosanowski, 2011). Indeed, it is widely known that humans make use of rapid judgments to guide their social interactions (Ambady, Bernieri, & Richeson, 2000; Richeson & Ambady, 2001).

[a]Département de Réadaptation, Faculté de Médecine, Université Laval, Quebec City, Canada
[b]Département d'Ophtalmologie, d'Otorhinolaryngologie et de Chirurgie Cervico-Faciale, Faculté de Médecine, Université Laval, Quebec City, Canada
[c]CERVO Brain Research Centre, Quebec City, Canada
Correspondence to Pascale Tremblay: pascale.tremblay@fmed.ulaval.ca

Editor: Julie Liss
Associate Editor: Nelson Roy
Received May 20, 2016
Revision received December 1, 2016
Accepted October 15, 2017
https://doi.org/10.1044/2017_JSLHR-S-16-0202

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Social encounters, occupational possibilities, and overall quality of life can all be affected by the way an individual is perceived by others (Ambady, Krabbenhoft, & Hogan, 2006; Pittinsky, Shih, & Ambady, 2000; Zuckerman & Driver, 1989). The voice of a person is a contributing factor to these social evaluations, as listeners form judgments about a variety of talker characteristics on the basis of their voices (Banse & Scherer, 1996; Klofstad, Anderson, & Peters, 2012; Mulac & Giles, 1996). Various acoustic cues could be involved in this phenomenon, including pitch, roughness, and intensity, as well as prosodic factors, such as speech tempo, pauses, and melodic contour. For instance, perceived speech tempo has been associated with poor ratings on psychosocial scales in younger adults (Aronovitch, 1976; Ryan & Johnston, 1987). These perceptions may, in turn, influence interactions with a conversational partner, especially if negative stereotypes on the basis of the perceived age of a talker are triggered (Ryan, Giles, Bartolucci, & Henwood, 1986; Ryan & Laurie, 1990). For example, a person interacting with a talker whom they perceive negatively (e.g., as less competent) may act differently (e.g., use exaggerated intonation and more repetition [Kemper, Ferrell, Harden, Billington, & Finter-Urczyk, 1998]) and even refrain from pursuing social interactions with that person (Amir & Levine-Yundof, 2013; Lallh & Rochet, 2000). In other words, talkers whose voices are negatively evaluated on either psychosocial dimensions related to sociability (e.g., likability, extraversion) or auditory-perceptual dimensions related to perceived voice quality (e.g., breathiness and roughness), or both, may also be negatively evaluated on other domains, including the desire to interact with them, as it is the case with the perception of faces (Zebrowitz & Montepare, 2008).

A number of talker and listener factors appear to influence voice psychosocial attribution. Some talker factors may be of particular interest on the basis of their influence on voice acoustics, such as a talker age, sex, and smoking status. For instance, it has been shown that older-sounding talkers are perceived more negatively than younger-sounding talkers (Benjamin, 1986; Mulac & Giles, 1996; Ryan & Capadano, 1978). One study has also observed more negative attitudes toward female voices than male voices in adults with dysphonia (Amir & Levine-Yundof, 2013). Female voices were also attributed more undesirable psychosocial traits associated with gender stereotypes (i.e., more submissive and dependent) than male voices in another study (Mulac & Giles, 1996). In addition to age and sex, cigarette smoking can also influence voice evaluation. Indeed, smoking is associated with wide-ranging impacts on laryngeal structures, such as vocal fold edema and erythema, epithelium thickening, and more severe pathology, such as keratinization and/or necrosis (Auerbach, Hammond, & Garfinkel, 1970; Hirabayashi et al., 1990; Ryan, McDonald, & Devine, 1955). These biological alterations are associated with voice acoustic changes, such as changes in fundamental frequency (f0; Gonzalez & Carpi, 2004) and voice quality as measured by jitter and shimmer (Guimarães & Abberton, 2005; Pinto, Crespo, & Mourão, 2014). Although it has been shown that these changes can be detected by listeners (Vincent & Gilbert, 2012), the influence of a talker smoking status on voice auditory-perceptual and voice-related psychosocial attributions has not been examined all at once. Understanding the factors that affect voice evaluation could facilitate clinical voice interventions with adults by revealing the most functionally relevant targets for voice interventions, that is, those parameters that have the most negative effects on social evaluation.

Differences in voice evaluation may also reflect listener biases rather than, or in addition to, acoustic differences between talkers. Specifically, the age of a listener could play a role in the evaluation of voices (Linville, 1987), as well as in their concept of "oldness" (Eppley & Mueller, 2001). However, previous studies on voice evaluation have mainly used college undergraduates as talkers (Markel, Phillis, Vargas, & Howard, 1972), listeners (Mulac & Giles, 1996; Ryan & Capadano, 1978), or both (Aronovitch, 1976; McAleer et al., 2014; Zuckerman & Driver, 1989), thus limiting the ability to draw conclusions about the impact of listener factors, such as age, on voice evaluation. Nonetheless, a few studies have shown that younger adults were better at estimating age than older adults (Huntley et al., 1987; Linville, 1987). It has also been shown that older adults provide significantly more favorable voice ratings of male and female talkers than younger and middle-aged adults (Hollien, Gelfer, & Carlson, 1991). Hollien et al. (1991) attributed this result to a greater acceptance of older adults toward various voices. It is also possible that older adults have an easier time understanding the utterances of other adults of similar age, with whom they are likely to interact more often, and consequently judge older voices in a more favorable way than younger listeners.

However, the literature is not clear about whether the influence of listeners' age on voice ratings is consistent across talker factors (younger and older, male and female, smoker and nonsmoker) and across voice-related dimensions (auditory-perceptual, psychosocial, speech tempo). In fact, in a recent study, listener age did not affect attitudes toward dysphonic talkers (Amir & Levine-Yundof, 2013). In contrast, one study has found that younger listeners provided more positive evaluations of activeness and intelligence-related dimensions for older adult voices compared with older listeners (Benjamin, 1986). Because a variety of measures of auditory-perceptual (e.g., loudness, roughness, and tremor) and psychosocial trait attribution (e.g., pleasantness, naturalness, flexibility, enthusiasm, and honesty) have been used in voice evaluation studies (Benjamin, 1986; Goy, Pichora-Fuller, & van Lieshout, 2016; Ryan & Capadano, 1978), it remains unclear if some measures are more sensitive to listener or talker differences. Indeed, very few studies have investigated the impact of both talker and listener age on voice evaluations (e.g., Goy et al., 2016). No study has examined the effect of listeners' age on auditory-perceptual and voice-related psychosocial traits attribution simultaneously or the social impact of potential listener differences in voice evaluations. Yet, understanding how aging shapes the perception of the human voice is critical to better understand the factors that affect

communication and social interactions throughout the life span. More generally, it is also unclear whether auditory-perceptual voice evaluation and voice-related psychosocial trait attribution are related to each other and whether this association depends on listener age. In sum, many questions remain about the impact of talker and listener factors on auditory-perceptual voice evaluation and voice-related psychosocial attribution and about the relationship between voice evaluation and social interactions.

The overall objective of this study was to further current understanding of talker and listener factors that affect voice auditory-perceptual evaluation, perceived speech tempo, and voice-related psychosocial attributions. The first specific objective was to examine the role of listeners' age on auditory-perceptual voice evaluation and voice-related psychosocial attribution. The second specific objective was to examine the role of three talker factors, that is, age, sex, and smoking status, on voice auditory-perceptual evaluation and voice-related psychosocial attribution. The third specific objective was to identify the voice-related dimensions associated with the propensity to interact socially with a person, including auditory-perceptual evaluation, perceived speech tempo, and voice-related psychosocial trait attributions, and to determine whether the impact of these dimensions differs as a function of the age of the listeners. The first hypothesis was that younger and older listeners would differ in terms of auditory-perceptual evaluations, perceived speech tempo, and voice-related psychosocial trait attributions across talker types. Specifically, we expected that older listeners would provide generally more positive evaluations of older talkers than younger listeners on both auditory-perceptual and voice-related psychosocial dimensions, irrespective of a talker gender. No differences were expected for the evaluation of younger talkers. Our second hypothesis was that older and smoking talkers would receive more negative auditory-perceptual evaluations and voice-related psychosocial trait attributions than younger and nonsmoking adults, irrespective of a talker gender. Our third hypothesis was that the propensity of a listener to interact with a person would be positively associated with psychosocial trait attributions, perceived speech tempo, and auditory-perceptual evaluations, irrespective of talker age and gender.

Method

Listeners

Fifty-one healthy nonsmoking native speakers of Canadian French were recruited to participate in a voice evaluation task during which they were presented with voice samples collected as part of a previous study (Lortie, Thibeault, Guitton, & Tremblay, 2015). We refer to the participants as the "listeners" and to the voices that were being evaluated as the "talkers." No voice production data were collected on the listeners.

The listeners had no self-reported history of speech, voice, language, hearing, and neurological or neurodegenerative disorder, and no self-reported history of drug or alcohol abuse. They were recruited from the general community in Québec City (QC, Canada) through advertisements placed in local newspapers, flyers distributed at various local events (e.g., public conferences and open house events), and posters in strategic locations within the community (e.g., drugstores and bingo halls). Recruitment e-mails were also sent to the Université Laval and the CERVO Brain Research Center using mailing lists targeting students, professors, and staff. To ensure that listeners would correctly understand the instructions, cognitive functioning was evaluated using the Montreal Cognitive Assessment scale (Nasreddine et al., 2005). One listener was excluded because he did not meet the cutoff score (score < 23). The remaining 50 listeners were divided into two groups: 25 younger adults ($M \pm SD$: 26.96 ± 4.38 years old, range = 19–37 years; 16.4 ± 1.91 years of education, range = 12–20 years) and 25 older adults (62.48 ± 7.52 years old, range = 51–74 years; 15 ± 3.44 years of education, range = 11–23 years). Listeners' characteristics are reported in Table 1. The procedures were approved by the Institutional Ethical Committee of the "Institut Universitaire en Santé Mentale de Québec" (protocol #353-2014). Informed written consent was obtained from all listeners, and they were compensated for their participation ($20 CAD).

The age range for the listeners corresponded to talkers' age range, that is, 19 to 75 years old. Discontinuous age groups were created to examine the relationship between listener age and voice evaluation while maximizing the age difference between the two groups. Previous studies focusing on the relationship between age and voice evaluation have successfully used discontinuous age groups (e.g., Eppley & Mueller, 2001; Goy et al., 2016). The choice of using an older listener group beginning at 50 years of age was based on the well-established finding that several voice changes occur during the fifth decade, including mean f0 in men (e.g., Dehqan, Scherer, Dashti, Ansari-Moghaddam, & Fanaie, 2012; Honjo & Isshiki, 1980; Ma & Love, 2010; Torre & Barlow, 2009). In addition, the average age at menopause, which is associated with important voice changes (e.g., D'haeseleer, Depypere, Claeys, Baudonck, & Van Lierde, 2011; Raj, Gupta, Chowdhury, & Chadha, 2010), is 51 years in Canada (Society of Obstetricians and Gynecologists of Canada, 2002). Thus, 50 years appears to be an appropriate lower cutoff to study age differences in voice evaluation.

Our sample size (N = 50 listeners) was retrospectively validated using a power analysis conducted with G-power software (Faul, Erdfelder, Lang, & Buchner, 2007), which was based on the results from a previous study that found differences in voice evaluations as a function of listener and talker age (Goy et al., 2016). In this study, young and older listeners differently evaluated the pleasantness and roughness of the voice of younger and older talkers. Specifically, younger but not older listeners evaluated older voices more negatively than younger voices. The effect sizes were moderate to high (for pleasantness: d = 0.9, for roughness: d = 0.74). Thus, assuming an alpha (α) level of .05 and a statistical power (1-β) of .80, the sample size needed was

Table 1. Listener characteristics, for each age group and overall.

	Age		Education (in years)		MoCA		Right PTA		Left PTA	
	M ± *SD*	Range	*M* ± *SD*	Range	*M* ± *SD*	Range	*M* ± *SD*	Range	*M* ± *SD*	Range
Young	30.0 ± 4.4	19–37	16.0 ± 1.9	12–20	28.7 ± 1.2	26–30	3.7 ± 3.4	–2.0–12.0	2.8 ± 4.1	–5.3–15.0
Older	62.5 ± 7.5	51–74	15.0 ± 3.4	11–23	27.5 ± 2.1	23–30	14.9 ± 9.1	2.3–36.7	13.0 ± 8.0	0.3–39.3
Total	44.7 ± 18.9	19–74	15.7 ± 2.8	11–23	28.1 ± 1.8	23–30	9.3 ± 8.8	–2.0–36.7	7.9 ± 8.1	–5.3–39.3

Note. MoCA = Montreal Cognitive Assessment scale. PTA = pure-tone average. The MoCA score ranges from 0 to 30, and a cutoff score between 20 to 23 optimizes sensitivity and specificity of detection of impairment (Waldron-Perrine & Axelrod, 2012).

25 participants per group to be able to detect differences between listeners (Cohen's *d*) of a minimal magnitude of 0.74 (the smallest reported effect size of interest) using Mann–Whitney tests for independent samples. The current sample of 50 adults, divided into two groups of 25 participants, was thus adequately powered to test hypotheses about listeners' age on voice evaluation.

Stimuli (Talker Voices)

Eighty native talkers of Canadian French were recruited as part of previous experiments (Lortie et al., 2015) from the general community in Québec City (QC, Canada) through advertisements placed in local newspapers, flyers distributed at various local events (e.g., public conferences and open house events), and posters in strategic locations within the community (e.g., drugstores and bingo halls). Recruitment e-mails were also sent to the Université Laval and the CERVO Brain Research Center using mailing lists targeting students, professors, and staff. Exclusion criteria for the talkers included self-reported history of diagnosed respiratory, speech, voice, language, swallowing, hearing, neurological, or neurodegenerative disorder, a history of acute or chronic respiratory disorder (asthma, bronchitis, etc.), dysphagia, or any self-reported laryngeal trauma or disorders, including surgically treated nodules, polyps, or Reinke's edema. At the time of recording, the talkers reported no voice ailments, allergy, or respiratory infection.

As part of our previous experiment, talkers produced a sustained vowel /a/ twice at comfortable frequency and amplitude levels, then narrated, using their own words, two popular stories (i.e., "Red Riding Hood" and "Three Little Pigs"). The talkers were given an illustration of the story to help recall. The stories were narrated at a comfortable voice frequency, amplitude, and rate. All recordings were performed by the same examiner under identical conditions in a quiet (but not soundproof) testing room, using a headset microphone (Microflex Beta 53, Shure) placed at 5 cm (45° angle) from the talker's mouth to decrease aerodynamic noise from the mouth. A headset microphone was used to ensure that the distance between the microphone and the mouth was kept constant throughout the procedure. The microphone was connected to an analog interface (Edirol UA-25EX, Roland) connected to a laptop computer. The recordings were performed using the Audacity software (Version 2.0.3; Free Software Foundation) at a sampling rate of 44.1 kHz and 32 bits of quantization.

Voice Assessment

To determine whether talkers' voices were within normal limits, the vocal samples were first analyzed using Praat, Version 5.3.39 (Boersma & Weenink, 2013). All voice recordings were visually inspected to identify segments with artifacts, such as extraneous noise, laughter or coughing, and caricatured voices. These segments were excluded from the analysis.

1. For the sustained vowel analysis, the longest and most stable central segment of the vowel was manually selected, excluding voice onset and offset (mean duration ± *SD*: 12.24 ± 6.7 s). A Praat script was applied on that central section to automatically extract all acoustical measures, that is, mean f0 (Hz) relative jitter (%) and shimmer (dB).
2. For the connected speech analysis, a fixed 10-s central segment of each original recording was manually selected. A Praat script was applied to that section to extract mean f0 (Hz). This was done because the duration of the speech samples varied considerably across participants. Moreover, this procedure allowed us to choose a stable portion of the discourse, excluding the beginning and end. Using a portion of a long recording is also a relatively standard procedure (Amir & Levine-Yundof, 2013; Goy et al., 2016).

The analysis of talkers' voices revealed that their voices were within normal limits (Goy, Fernandes, Pichora-Fuller, & van Lieshout, 2013). The details of this analysis are provided in the Supplemental Material. In addition, all talkers scored well within the normal limits on the Voice Handicap Index (Jacobson et al., 1997; threshold for mild voice handicap: 33/120), meaning that their voices did not have a negative impact on their daily activities (younger: 5.22 ± 5.6, older: 4.45 ± 6.3).

Stimuli Preparation

The 80 connected speech recordings were separated into four categories (N = 20 each) on the basis of age (younger: 20–49 years old, older: 50–75 years old) and smoking status (smoking, nonsmoking). Smokers were regular

tobacco consumers, with an average of 17.4 ± 10.9 cigarettes consumed per day and 25.1 ± 14.8 years of regular smoking. Nonsmokers had never smoked or had stopped smoking at least 3 years prior to the study. Each talker category was composed of 10 women and 10 men (Table 2). The choice of using an older group beginning at 50 years follows the same rationale described in the Listeners section.

To create the 80 connected speech recordings that served as stimuli in this study, first, for each talker, an uninterrupted and semantically coherent segment of 20 to 25 words was manually selected from one of the stories and saved in .wav format using the Audacity software. Selected segments varied in length from 3.4 to 14.3 s, with mean ± *SD* of 8.4 ± 2.4 s. The number of words was controlled rather than the duration (in seconds) of the recording on the basis of previous studies in which the number of words was kept constant by including only selected sentences from a reading passage irrespective of the precise duration of the sentence (Amir & Levine-Yundof, 2013; Goy et al., 2016; Huntley et al., 1987; Ryan & Capadano, 1978). To ensure that semantic content was coherent in all samples, we avoided breaking down sentences. This is important because semantic content can influence voice evaluation (Ryan & Johnston, 1987). Finally, to avoid intensity effects on voice evaluation, the amplitude of the segments was normalized to 70 dB SPL using a Praat script. This is important because previous studies have shown that intensity is a dominant dimension in listeners' judgments of dysphonic voices (Kempster, Kistler, & Hillenbrand, 1991) and that most listeners prefer medium intensity voices (Hollien et al., 1991).

Procedure

The study consisted of two parts: (a) hearing assessment and (b) voice evaluation task.

Hearing Assessment

To ensure that listeners' hearing was normal, an audiometric evaluation was performed in a double-walled soundproof room (Génie Audio Inc.). This was necessary given the auditory nature of the main task. Pure-tone audiometry was performed using a clinical audiometer (AC40, Interacoustic) and TDH-39 earphones (Telephonics) for each ear separately, for the following frequencies: 0.25, 0.5, 1, and 2 kHz. For each listener, a standard pure-tone average (average of thresholds at 0.5, 1, and 2 kHz) was computed for the left and right ear (Stach, 2010). Listeners' pure-tone average hearing thresholds were under 20 dB HL (young) or 40 dB HL (older), which confirmed that their hearing capacities were within normal limits according to their age (Blanchet et al., 2008).

Voice Evaluation Task

Listeners evaluated 80 voice recordings through a rigorous computer-based procedure. For this task, listeners were comfortably seated in a double-walled soundproof room, facing a computer screen. The stimuli were presented auditorily using Presentation (Version 18.1, Neuro-Behavioural Systems) through a Quartet soundcard (Apogee Electronics) and a high-quality, closed, and dynamic headset (DT 770 Pro, Beyerdynamic). All voice samples were presented binaurally.

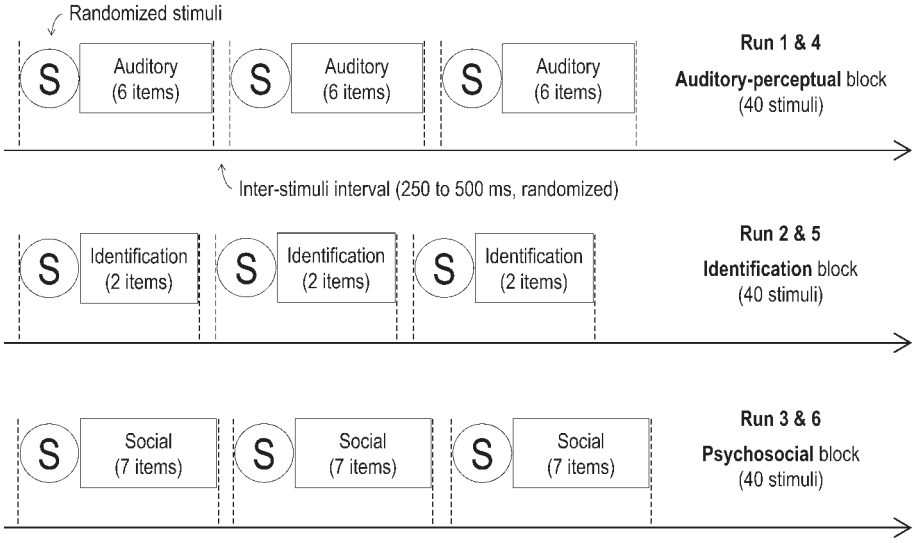
Standardized instructions were presented on a computer monitor. Before beginning the experiment, a training session was conducted, during which listeners evaluated a sample of 10 voices that were not included in the main experiment. This was done to ensure that participants understood the instructions. Listeners could ask the experimenter to increase or decrease the intensity of the stimuli during the training session until they could hear the stimuli effortlessly. This was done to ensure that they could hear well, a procedure that is often used for tests of auditory perception. After the initial adjustments, the intensity remained constant. The evaluation of talker voices was done through 15 different questions, organized into three blocks: (a) talker identification, (b) auditory-perceptual evaluation, and (c) psychosocial trait attributions (see Figure 1). In the identification block, listeners identified the age of the talkers on a 7-point scale (20–29; 30–39; 40–49; 50–59; 60–69; 70–79; and 80+ years old), as well as their sex (male, female). In the auditory-perceptual block, which was adapted from the clinical Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009), listeners scored each talker's voice on six dimensions (i.e., presence of a disorder [disorder], roughness, strain, high pitch, low pitch, breathiness). The loudness dimension of the CAPE-V was not included because the voice samples were intensity normalized. The pitch dimension of the CAPE-V was split into two dimensions. The first documented the extent to which a voice was perceived as low pitched, and the other documented

Table 2. Talker characteristics.

	Talker categories			
	Younger talkers		Older talkers	
Features	Smoking	Nonsmoking	Smoking	Nonsmoking
Mean age ± *SD*	33.9 ± 8.9	34.1 ± 8.9	58.3 ± 5.8	62.2 ± 4.8
Male/female ratio	10/10	10/10	10/10	10/10

Note. Age and gender ratio for each of the talker groups (younger and older).

Figure 1. Experimental design. Listeners evaluated 240 samples, organized into six runs containing 40 voice samples each. Within a run, only one question block (identification [talker age and sex], auditory-perceptual, or voice-related psychosocial attribution) was used to avoid confusion.



the extent to which a voice was perceived as high pitched. This was done to facilitate rating and to determine the direction of any perceived pitch-related group difference. Finally, the overall severity dimension of the CAPE-V, which evaluates the severity of the voice disorder assessed (from mild to severe), was slightly modified to ask participants to indicate the extent to which voices were perceived as presenting a disorder or not (from a scale of normal to disordered). This was done so that participants would not assume that talkers suffered from a voice disorder. Five-point asymmetric Likert scales, ranging from 1 = *not at all* to 5 = *extremely,* were used instead of the visual analog scales that are used in the CAPE-V. This was done to standardize the voice evaluation procedure across blocks (auditory-perceptual evaluation and psychosocial trait attributions) and, thus, facilitate rating. Given that each participant evaluated many voices (80) as part of a long procedure, we felt that it was important to simplify the rating process. Having one scale across the entire procedure made the experiment simpler and faster. To improve clarity, the orientation of the auditory-perceptual scores was flipped so that all large numbers reflected positive, advantageous evaluations. For example, a high score on the roughness dimension means that a voice was not perceived as rough and was hence given a positive roughness evaluation. The specific questions are provided in the Appendix.

In the Psychosocial block, which was adapted from Oswald's situational empathy scale (Oswald, 1996), listeners evaluated six "social" aspects of the voices (i.e., warmness, agreeableness, work affinity [easy to work with], cheerfulness, confidence, conversational affinity [enjoyable to speak with]), also using 5-point Likert scales but of a symmetric nature, ranging from 1 = *very difficult* to 5 = *very easy* or similar scales. Listeners also evaluated the perceived speech tempo using a symmetric 5-point Likert scale, ranging from 1 = *very slow* to 5 = *very fast*. Therefore, a low score on psychosocial dimensions reflects a negative evaluation. Perceived speech tempo was included in the psychosocial block rather than in the auditory-perceptual block to group questions on the basis of scale types (symmetric vs. asymmetric) to avoid confusion, but it was analyzed separately. The words used in the scales (e.g., warmness) were selected from everyday language to avoid misunderstandings; thus, no further definitions were provided to the listeners.

During the main experiment, participants first listened to a voice sample, and then, they rated it using a computer keyboard. Listeners were asked to try and use the entire scales for all the questions. Throughout the experiment, water and short breaks were provided to the listeners as needed. The task was self-paced, and listeners typically completed the experiment within 120 to 150 min.

Each voice sample was evaluated three times (once for each block of questions: auditory-perceptual, identification, and psychosocial) across six different experimental runs (Figure 1). Each run contained half the stimuli (40 voice samples) and only one type of block to avoid confusion. Thus, each listener performed a total of 240 evaluations. The rationale for having no voice segment presented twice in the same run was based on practical considerations with regard to testing time and fatigue. Stimuli were randomized for each listener in each run. Interstimulus intervals varied in length (ranging from 250 to 500 ms, randomized). The order of the runs was also randomized for each listener. Listeners were not aware that age and smoking status were of interest in the study.

Eight listeners (one younger listener and seven older listeners) did not complete the task within the allocated time (180 min) due to a slow response rate. For

these listeners, we were able to collect on average 75.78% of the data.

Statistical Analyses

All data were analyzed using SPSS 23 (IBM SPSS Statistics, Armonk, NY). Perceived age corresponded to the age category selected from 1 to 7, and perceived sex corresponded to the attributed sex (male, female). Listeners' evaluations of sociality measures (work affinity, conversational affinity) and the 11 remaining dimensions (disorder, roughness, strain, high pitch, low pitch, breathiness, warmness, agreeableness, perceived speech tempo, cheerfulness, and confidence) were analyzed separately. Because listener sex was shown to have little or no influence on a listener's perceptual evaluation in previous studies (Amir & Levine-Yundof, 2013; Aronovitch, 1976; Hollien et al., 1991), the sex of the listener was not analyzed here. The statistical analyses were conducted in four steps, described below. For all statistical procedures, $\alpha = .05$ was used to establish significance. Measures of effect sizes are provided in the form of Cohen's *d* statistics when comparing means.

Age and Sex Identification

To determine whether talker age and sex were correctly identified by listeners, a series of correlations were performed between perceived and real age of the talkers for each listener. An error analysis was also performed on sex identification (percentage of correctly identified voices).

Intrarater and Interrater Reliability

First, *outliers,* defined as values that were 3 *SD*s away from the mean in each listener group (age grouping) and talker category (age, sex, and smoking status grouping), were removed. After excluding outliers (337 out of 7,440 values), 95.47% of the original data were included in the analyses. Next, the interrater reliability was assessed using Cronbach's alpha coefficients. This involves measuring the relationship between each listener's rating for each voice and the group mean of all the other listeners. The intrarater reliability was also assessed. However, because no voice segment was evaluated twice on the same dimension due to time constraints, a split-half reliability procedure on the basis of Cronbach (Cronbach, 1951) was used. This procedure was based on a previous study from our group (Deschamps, Hasson, & Tremblay, 2016). That is, for each listener, the evaluations of each talker category (composed of 10 different voice stimuli) were divided in two halves each containing five voices, resulting in two sets of voices each containing the same number of stimuli per talker category. We then calculated the split-half coefficient for each listener.

Voice Evaluation

To dissociate the effect of listener- and talker-related variables on voice evaluation, a series of nonparametric Mann–Whitney tests for independent samples was conducted. For each listener, voice evaluations were averaged within each of the eight talker categories (see Table 2). Work affinity and conversational affinity were not included in these analyses. A second and more stringent outlier removal procedure was applied to the within-subject averaged data to ensure the normality of the group distribution. *Outliers,* defined as values that were 3 median absolute deviations away from the median of each dependent variable in each age group, were removed from the statistical analyses (Huber, 1981; Leys, Ley, Klein, Bernard, & Licata, 2013). The median absolute deviation is a robust alternative to the classic mean ± 3 or 2 *SD*s. This method is immune to sample size (Leys et al., 2013). After excluding outliers (712 out of 5,096 values), 86.03% of the data were included in the analyses.

Younger versus older listeners. To identify age differences in voice evaluation, listeners were divided into two age groups (young: 20–37, older: 51–74 years). Voice ratings on 11 dimensions (disorder, roughness, strain, high pitch, low pitch, breathiness, warmness, agreeableness, perceived speech tempo, cheerfulness, and confidence) were compared using Mann–Whitney tests for independent samples (younger and older listeners) separately for each talker category.

These results are summarized in Figure 2 in the form of radar charts. Each radius in the radar chart corresponds to a measured dimension of voice (e.g., warmness and strain). The average rating for each group on each dimension is marked by a point located within each dimension. When a point is located closer to the outer line, regardless of the dimension being evaluated, evaluation is more positive. For each group, the ratings are connected through a line forming a geometrical figure. Note that for speech tempo, higher (more positive) scores indicate greater speed.

Younger versus older talker. To determine whether younger and older talker voices were differently evaluated, voice evaluations from the two listener groups were collapsed. Differences in voice ratings on 11 dimensions (disorder, roughness, strain, high pitch, low pitch, breathiness, warmness, agreeableness, perceived speech tempo, cheerfulness and confidence) were compared using Mann–Whitney tests for independent samples (younger and older talkers) separately for each talker category (female nonsmoking, female smoking, male nonsmoking, male smoking). These results are summarized in Figure 3 in the form of radar charts. The organization of these charts is described in section Younger versus older listeners.

Voice Dimensions

Relationship between voice dimensions. To determine whether listeners' evaluations of different voice dimensions are independent, a series of correlation analyses was performed, separately for the younger and older listeners. Because the dependent variables were on an ordinal scale (the ratings), nonparametric (Spearman) correlations were used to examine the strength of the relationships between the 11 voice dimensions that were evaluated in this study (disorder, roughness, strain, high pitch, low pitch, breathiness, warmness, agreeableness, perceived speech tempo, cheerfulness, and confidence).

Figure 2. Age differences in voice evaluation as a function of listener age. The radar chart illustrates the result of the Mann–Whitney tests for the differences between younger and older listeners' evaluations of 11 voice-related dimensions (e.g., agreeableness and strain), separately for each of the eight talker categories (younger/older, female/male, smoking/nonsmoking). The average rating for each listener group on each dimension is marked by a point (black for the younger listeners and gray for the older listeners). When a point is located closer to the outer line, regardless of the dimension being evaluated, evaluation is more positive. For speech tempo, a higher (more positive) score indicates greater speed. For each group, the ratings are connected through a line forming a geometrical figure (this figure is black for the younger group and gray for the older group). If all dimensions received similar evaluations, the geometrical figure formed would be a hendecagon (i.e., an eleven-sided polygon). The more heterogeneous the ratings, the less regular-shaped the figure is. If a group of speakers was evaluated globally more positively than another group, such as was expected (young > older adults) the geometrical figure formed by the ratings of the second group (e.g., older adults) would be completely enclosed within the geometrical figure of the first group (e.g., younger adults). However, as shown in the figure, most of the differences between the younger and older listeners' evaluations concerned the roughness dimension, which was evaluated more positively by the older listeners compared with the younger listeners. Asterisks indicate a significant group difference (young listeners ≠ older listeners) for a specific voice-related dimension at $p < .05$.

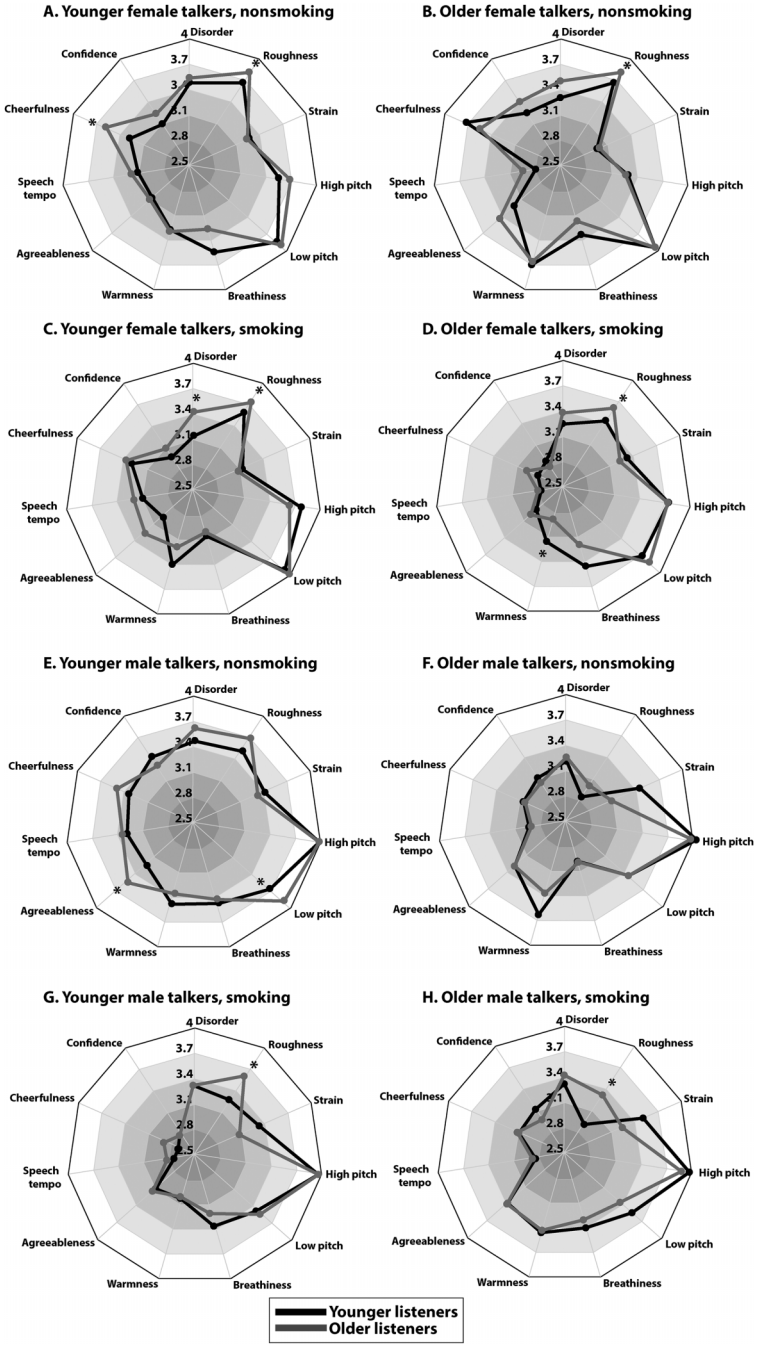
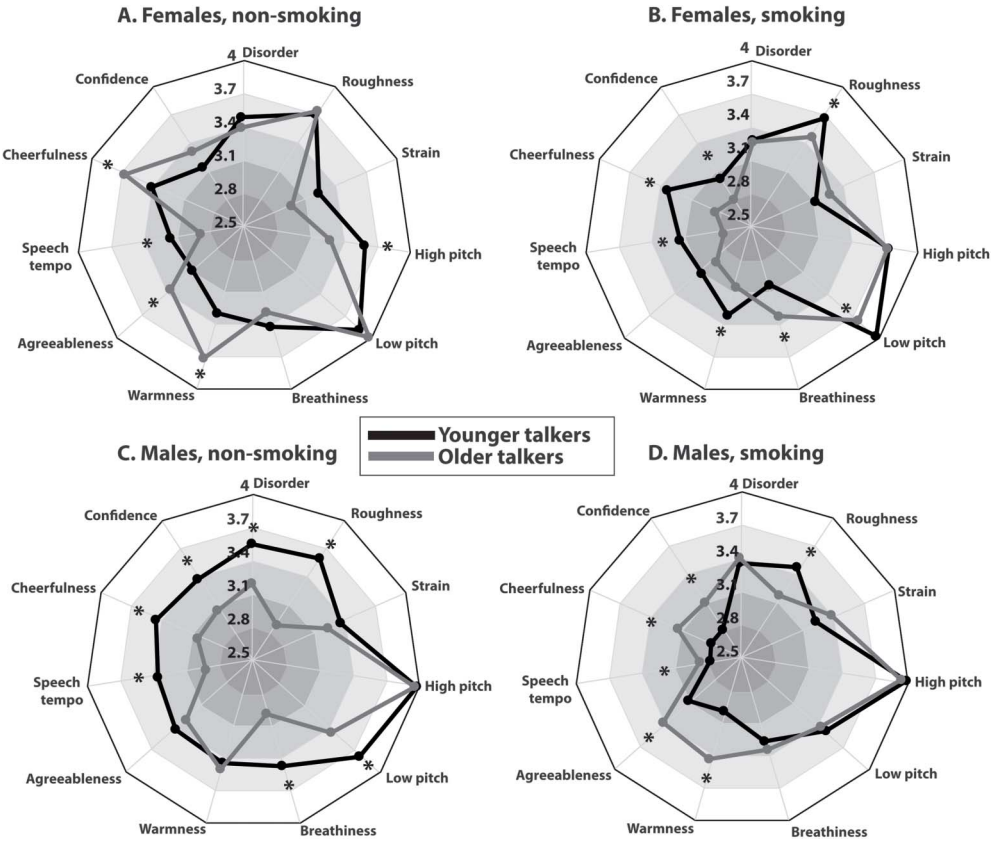


Figure 3. Age differences in voice evaluation as a function of talker age. The radar chart illustrates the result of the Mann–Whitney tests for the differences between younger and older talker voices on 11 voice-related dimensions (e.g., agreeableness and strain), separately for each talker category (female/male, smoking/nonsmoking). The average rating for each talker group on each dimension is marked by a point. In each plot, the ratings for one group are connected through a line forming a geometrical figure (this figure is black for the younger talkers and gray for the older talkers). As can be seen in the figure, the younger and older talkers were evaluated distinctly on several dimensions (including cheerfulness, speech tempo, and agreeableness); these age differences varied as a function of talker sex and smoking status. Asterisks indicate a significant group difference (young talker ≠ older talker) at $p < .05$.



Propensity to engage in social interactions. Because we were interested in identifying the voice dimensions that may influence social interactions, a series of nonparametric correlation analyses was performed to identify the voice dimensions that were associated with a listener's propensity to engage in social interactions with a person, separately for the younger and the older listeners. Because the dependent variables were on an ordinal scale (the ratings), nonparametric (Spearman) correlations were used to examine the strength of the relationships between sociality measures (work affinity, conversational affinity) and the 11 voice dimensions (disorder, roughness, strain, high pitch, low pitch, breathiness, warmness, agreeableness, perceived speech tempo, cheerfulness, and confidence).

Results

Age and Sex Identification

Perceived talker age was strongly correlated with the real talker age ($r = .71$, $r^2 = .52$, $p < .05$). The association between perceived and real age was stronger for younger listeners ($r = .78$, $r^2 = .61$, $p < .001$) than older listeners ($r = .63$, $r^2 = .43$, $p < .05$), but it was positive and significant in both groups. Listeners correctly identified the sex of the talker 98.24% of the time, with a mean error rate of 1.36 talkers misidentified over 80 samples in total. The percentage of talkers' sex correctly identified was similar between younger (98.4%) and older listeners (98.2%; $t(79) = -0.45$, $p = .65$).

Intrarater and Interrater Reliability

A high interrater reliability was observed for auditory-perceptual dimensions (Cronbach α ranging from .85 to .95) and social dimensions (Cronbach α ranging from .93 to .97). Split-half coefficients revealed good internal consistency for each listener for auditory-perceptual dimensions (Cronbach α ranging from .78 to .96) and psychosocial dimensions (Cronbach α ranging from .93 to .98).

Voice Evaluation

Younger Versus Older Listeners

Age differences in voice evaluations are provided as a function of talker category in Table 3 and illustrated as radar charts in Figure 2. In the figure, each of the eight categories of talker is illustrated separately. As can be seen in the figure, there were only a few age differences, most of which (10/11) reflected more positive evaluations made by the older listeners. Most of the differences between younger and older listeners' evaluations (6/11) concerned the roughness dimension (d = 0.67 to 1.22), which was evaluated more positively by the older listeners compared with the younger listeners (Figures 2A, 2B, 2C, 2D, 2G, and 2H). In addition, older listeners gave a more positive evaluation of cheerfulness for the younger nonsmoking female talkers (d = 1.26) and agreeability for the younger nonsmoking male talkers (d = 0.79) than younger listeners (Figures 2A and 2E). There was only one other instance of a more positive score attributed by the younger listeners; the younger listeners provided a more positive rating of the warmness of older smoking female talkers than the older listeners (d = 0.97; Figure 2D).

Younger Versus Older Talkers

Voice evaluations are provided as a function of talker category in Table 4 and illustrated as radar charts in Figure 3. In the radar charts, each category of talker is illustrated separately. The lines on the radar charts represent the average scores for each talker group (young, older) across all listeners. As can be seen in the figure, there were several differences in how younger and older talkers were perceived overall, with important sex differences. All types of voices exhibited age differences, but the female nonsmoking talkers were the group with the least perceived age differences. For the female talkers, the Mann–Whitney tests revealed age differences that varied significantly across smoking status. Specifically, for nonsmoking talkers (Figure 3A), older voices were perceived more negatively than younger voices on two dimensions: high pitch (d = 0.68) and speech tempo (d = 0.98; lower scores for the older talkers); in contrast, they were perceived more positively than younger voices on three dimensions: warmness (d = 1.29), agreeableness (d = 0.6), and cheerfulness (d = 0.94). For the smoking female talkers (Figure 3B), older voices were evaluated more negatively than younger voices on roughness (d = 0.87), low pitch (d = 1.16), warmness (d = 0.85), speech tempo (d = 1.78), cheerfulness (d = 1.4), and confidence (d = 0.51). Only breathiness was evaluated more positively for the older smoking female talkers compared with the younger smoking female talkers (d = 0.57).

A remarkably different pattern was found for the male talkers. For nonsmoking male talkers (Figure 3C), Mann–Whitney tests revealed that older voices were perceived more negatively than younger voices on seven dimensions: disorder (d = 0.79), roughness (d = 1.59), low pitch (d = 1), breathiness (d = 0.99), perceived speech tempo (d = 1.88), cheerfulness (d = 1.36), and confidence (d = 0.71). In contrast, for the smoking male talkers (Figure 3D), older voices were perceived more positively than younger voices on five dimensions: warmness (d = 1.32), agreeableness (d = 0.95), cheerfulness (d = 0.97), confidence (d = 0.71), and perceived speech tempo (d = 0.49). Only roughness was evaluated more negatively for the older male smoking talkers compared with the younger male smoking talkers (d = 0.75).

Voice Dimensions

Relationship Between Voice Dimensions

First, we examined the relationship between all pairs of voice dimensions using Spearman correlations, separately for the younger and the older listeners. These analyses, shown in Table 5, revealed that, in the older listeners, most auditory-perceptual dimensions were positively associated. Eleven out of 15 (73%) pairs of auditory-perceptual dimensions were significantly associated. In contrast, in the younger listeners, only one significant association was found between auditory-perceptual dimensions, namely, strain and breathiness.

The analyses also revealed that only a few auditory-perceptual dimensions were associated with voice-related psychosocial attributions. Specifically, for the younger listeners, higher scores on low pitch and roughness dimensions were associated with higher scores on cheerfulness and confidence, respectively, meaning that a positive auditory-perceptual evaluation was associated with a positive psychosocial attribution. In addition, the analyses revealed that, in younger listeners, higher perceived speech tempo (i.e., the perception of a faster speech rate) was associated with lower warmness scores. In contrast, in the older listeners, we did not find any association between speech tempo and any voice dimension.

Dimensions Associated With the Propensity to Engage in Social Interactions

Next, we examined the relationship between voice dimensions and sociality measures (work affinity, conversational affinity) using Spearman correlations. The results of these analyses revealed that, for the young listeners, higher perceived warmness was associated with higher perceived conversational affinity. Confidence ratings were positively associated with both sociality measures. Moreover, higher agreeableness and cheerfulness ratings were positively associated with sociality measures in both the younger and older listener groups. Speech tempo was not associated with sociality measures in any of the listener groups.

For auditory-perceptual voice dimensions, the analyses revealed that, for the younger listeners, positive evaluations of disorder and strain were associated with higher work affinity scores and that positive evaluations of low pitch were associated with higher conversational affinity scores. Finally, we examined the relationship between work and conversational affinity (see Table 5, bottom lines). This analysis revealed that work affinity and conversational

Table 3. Voice evaluations (*M* and *SD*) by listener age and talker categories.

			Female talkers								Male talkers							
		Talkers	Younger nonsmoking		Older nonsmoking		Younger smoking		Older smoking		Younger nonsmoking		Older nonsmoking		Younger smoking		Older smoking	
Voice evaluations		**Listeners**	**Y**	**O**	**Y**	**O**	**Y**	**O**	**Y**	**O**	**Y**	**O**	**Y**	**O**	**Y**	**O**	**Y**	**O**
Auditory-perceptual dimensions	Disorder	*M*	3.47	3.53	3.31	3.52	**3.16**	**3.45**	3.22	3.35	3.50	3.65	3.19	3.24	3.34	3.34	3.34	3.45
		SD	0.27	0.44	0.37	0.51	0.38	0.45	0.41	0.56	0.31	0.33	0.47	0.63	0.29	0.46	0.38	0.51
	Roughness	*M*	**3.65**	**3.79**	**3.68**	**3.83**	**3.61**	**3.77**	**3.41**	**3.58**	3.54	3.72	2.83	2.99	**3.30**	**3.63**	**2.93**	**3.34**
		SD	0.22	0.16	0.25	0.20	0.15	0.19	0.26	0.24	0.36	0.25	0.47	0.62	0.41	0.26	0.27	0.40
	Strain	*M*	3.28	3.24	2.98	3.00	3.14	3.08	3.31	3.21	3.42	3.32	3.42	3.07	3.36	3.11	3.53	3.26
		SD	0.25	0.62	0.57	0.74	0.55	0.63	0.36	0.55	0.45	0.54	0.37	0.66	0.42	0.69	0.28	0.61
	High pitch	*M*	3.56	3.69	3.32	3.29	3.79	3.65	3.72	3.71	4.00	4.00	4.00	3.94	4.00	4.00	4.00	3.91
		SD	0.40	0.27	0.53	0.57	0.22	0.36	0.23	0.37	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.13
	Low pitch	*M*	3.87	3.93	4.00	4.00	3.92	4.00	3.70	3.81	**3.68**	**3.90**	3.44	3.45	3.50	3.55	3.57	3.37
		SD	0.14	0.08	0.00	0.00	0.09	0.00	0.26	0.20	0.33	0.12	0.27	0.46	0.31	0.39	0.21	0.60
	Breathiness	*M*	3.55	3.27	3.36	3.20	3.05	3.00	3.44	3.19	3.49	3.44	2.97	2.98	3.37	3.20	3.41	3.31
		SD	0.33	0.53	0.45	0.46	0.60	0.57	0.30	0.50	0.40	0.34	0.66	0.50	0.38	0.56	0.33	0.41
Psychosocial dimensions	Warmness	*M*	3.28	3.31	3.74	3.70	3.40	3.19	**3.15**	**2.88**	3.50	3.37	3.61	3.36	3.02	3.01	3.47	3.44
		SD	0.39	0.43	0.22	0.25	0.27	0.33	0.29	0.26	0.32	0.35	0.45	0.31	0.32	0.35	0.39	0.28
	Agreableness	*M*	3.08	3.11	3.24	3.47	2.98	3.26	2.89	2.98	**3.26**	**3.57**	3.28	3.28	3.09	3.13	3.40	3.41
		SD	0.45	0.14	0.46	0.50	0.39	0.45	0.14	0.38	0.38	0.38	0.36	0.17	0.33	0.32	0.29	0.32
	Speech tempo	*M*	3.11	3.18	2.81	2.96	3.11	3.22	2.75	2.79	3.33	3.39	2.93	2.91	2.72	2.81	2.85	2.88
		SD	0.20	0.36	0.28	0.22	0.17	0.27	0.21	0.23	0.32	0.27	0.15	0.15	0.16	0.24	0.23	0.16
	Cheerfulness	*M*	**3.27**	**3.57**	3.74	3.57	3.30	3.38	2.81	2.96	3.37	3.53	3.04	3.02	2.70	2.88	3.10	3.12
		SD	0.30	0.17	0.28	0.25	0.29	0.25	0.29	0.47	0.26	0.24	0.29	0.40	0.30	0.37	0.28	0.39
	Confidence	*M*	3.08	3.22	3.25	3.41	2.98	3.11	2.85	2.78	3.46	3.34	3.10	3.03	2.79	2.79	3.13	2.99
		SD	0.25	0.55	0.54	0.32	0.40	0.34	0.54	0.41	0.50	0.48	0.47	0.44	0.43	0.22	0.43	0.41

Note. Auditory-perceptual dimensions are measured on 5-point asymmetric scales, ranging from 1 = *not at all* to 5 = *extremely*. The direction of the scale was flipped for the analyses so that all large numbers reflected positive, advantageous evaluations. Psychosocial dimensions are measured on 5-point symmetric scales, ranging from 1 = *very difficult* to 5 = *very easy* (or similar scales). Significant age differences (young listeners vs. older listeners), assessed via Mann–Whitney tests, are in bold numbers. Y = young listeners; O = older listeners.

Table 4. Voice evaluations (*M* and *SD*) by talker categories, collapsed across listener age groups.

Voice evaluations			Female talkers				Male talkers			
			Younger nonsmoking	Older nonsmoking	Younger smoking	Older smoking	Younger nonsmoking	Older nonsmoking	Younger smoking	Older smoking
Auditory-perceptual dimensions	Disorder	*M*	3.50	3.40	3.30	3.28	**3.57**	**3.21**	3.34	3.39
		SD	0.35	0.45	0.43	0.48	0.33	0.55	0.38	0.44
	Roughness	*M*	3.72	3.75	**3.68**	**3.49**	**3.63**	**2.91**	**3.45**	**3.15**
		SD	0.21	0.24	0.18	0.26	0.32	0.55	0.38	0.40
	Strain	*M*	3.26	2.99	3.11	3.26	3.37	3.26	3.25	3.40
		SD	0.47	0.65	0.58	0.46	0.49	0.55	0.57	0.49
	High pitch	*M*	**3.62**	**3.30**	3.72	3.71	4.00	3.97	4.00	3.96
		SD	0.35	0.54	0.30	0.30	0.00	0.07	0.00	0.10
	Low pitch	*M*	3.90	4.00	**3.96**	**3.75**	**3.78**	**3.45**	3.52	3.46
		SD	0.12	0.00	0.08	0.24	0.28	0.38	0.35	0.47
	Breathiness	*M*	3.42	3.28	**3.02**	**3.32**	**3.47**	**2.98**	3.29	3.37
		SD	0.45	0.46	0.58	0.42	0.37	0.59	0.47	0.37
Psychosocial dimensions	Warmness	*M*	**3.29**	**3.72**	**3.30**	**3.04**	3.43	3.50	**3.01**	**3.46**
		SD	0.40	0.24	0.32	0.30	0.34	0.41	0.33	0.35
	Agreeableness	*M*	**3.09**	**3.35**	3.10	2.94	3.41	3.28	**3.11**	**3.40**
		SD	0.37	0.49	0.44	0.29	0.41	0.30	0.32	0.30
	Speech tempo	*M*	**3.14**	**2.87**	**3.16**	**2.77**	**3.36**	**2.92**	**2.77**	**2.86**
		SD	0.29	0.26	0.23	0.22	0.29	0.15	0.20	0.20
	Cheerfulness	*M*	**3.39**	**3.66**	**3.34**	**2.88**	**3.44**	**3.03**	**2.78**	**3.11**
		SD	0.30	0.28	0.27	0.38	0.26	0.34	0.34	0.33
	Confidence	*M*	3.15	3.32	**3.04**	**2.82**	**3.40**	**3.07**	**2.79**	**3.07**
		SD	0.42	0.46	0.37	0.48	0.49	0.46	0.36	0.42

Note. Auditory-perceptual dimensions are measured on 5-point asymmetric scales, ranging from 1 = *not at all* to 5 = *extremely*. The direction of the scale was flipped for the analyses so that all large numbers reflected positive, advantageous evaluations. Psychosocial dimensions are measured on 5-point symmetric scales, ranging from 1 = *very difficult* to 5 = *very easy* (or similar scales). Significant age differences (young talkers vs. older talkers), assessed via Mann–Whitney tests, are in bold numbers.

affinity were highly positively correlated in both listener groups.

Discussion

In this study, we examined the role of talker and listener factors on auditory-perceptual evaluation of voice, perceived speech tempo, and voice-related psychosocial attributions. Our main finding is that talker-related factors (age, sex, and smoking status) have an overall stronger impact on voice evaluation than listener age. We also demonstrated that voice-related psychosocial attributions (e.g., cheerfulness and agreeableness), but few auditory-perceptual dimensions, are positively associated with the desire to engage in social interactions with a person. These findings are discussed in the following sections.

Influence of Listener Age on Voice Evaluation

On the basis of the assumption that older adults may have an easier time understanding the utterances of adults of similar age, with whom they are likely to interact more often, our hypothesis was that voice evaluation would differ between younger and older listeners and, in particular, that older listeners would judge older voices in a more favorable way compared with younger listeners. Our results support this hypothesis, at least to some extent. Indeed, we found that older listeners evaluated older talkers more favorably than younger listeners. However, our results also demonstrate that older listeners provided overall slightly more positive evaluations, regardless of talker age. Most differences between listener groups were observed for auditory-perceptual dimensions, mainly roughness, rather than psychosocial attributions. This finding is consistent with a previous study that observed that listener age affected auditory-perceptual ratings of nonsmoking voices more so than speech ratings, such as clarity and naturalness, and that younger listeners tended to give more negative roughness ratings to older talkers (Goy et al., 2016). Another study also observed differences in dysphonic voice evaluation depending on the listener age, whereby older listeners rated talkers more favorably than younger listeners (Amir & Levine-Yundof, 2013). In this study, a similar influence of listener age was observed on the evaluation of nonsmoking and smoking normal voices using a comprehensive set of voice-rating scales. It is thus possible that older listeners are relatively more accepting toward all types of voices or that their voice ratings are based on different perceptual characteristics or criteria than younger listeners. However, though older listeners tended to be more positive than younger listeners, this finding did not generalize to all dimensions (e.g., warmness in older smoking female talkers). Moreover, the ability to estimate a talker's age on the basis of the voice was not influenced by listener age, which is consistent with previous work on the same issue (Eppley & Mueller, 2001; Huntley et al., 1987). These findings may

Table 5. Spearman correlation matrix.

Dimensions		Roughness	Strain	High pitch	Low pitch	Breathiness	Warmness	Agreeableness	Speech tempo	Cheerfulness	Confidence	Work affinity	Conversational affinity
Disorder	Y	–.07	.29	.31	.12	.22	–.18	.05	.02	.35	.19	**.54**	.37
	O	**.45**	**.50**	**.56**	.37	**.55**	–.05	–.10	.11	–.26	.39	–.19	–.21
	Roughness	Y	.25	.16	.26	.38	–.13	.24	–.09	.10	**.52**	.07	.17
		O	.30	**.53**	**.55**	.37	–.23	–.35	–.29	–.44	.21	–.34	–.40
		Strain	Y	.07	.15	**.72**	–.05	.25	–.19	.39	.36	**.41**	.38
			O	**.63**	**.54**	**.72**	.04	–.11	–.36	–.23	.16	–.31	–.26
			High pitch	Y	.15	.18	–.14	.23	.31	–.14	.12	.30	.30
				O	**.72**	**.48**	.11	–.24	–.29	–.14	.03	–.19	–.24
				Low pitch	Y	.24	.15	.12	.09	**.46**	.06	.33	**.44**
					O	.33	–.10	–.45	–.26	–.38	–.04	–.38	–.44
					Breathiness	Y	.08	.30	–.24	.17	.39	.35	.37
						O	–.06	–.37	.01	–.21	.24	–.45	–.35
						Warmness	Y	.40	**–.43**	.35	.09	.16	**.41**
							O	.35	–.24	**.58**	.07	.24	.26
							Agreeableness	Y	–.31	.38	**.65**	**.56**	**.79**
								O	–.09	.29	.05	**.65**	**.64**
								Speech tempo	Y	–.34	–.26	–.08	–.26
									O	.27	.19	.07	.06
									Cheerfulness	Y	.38	**.60**	**.68**
										O	–.18	**.57**	**.51**
										Confidence	Y	**.50**	**.49**
											O	–.18	–.30
											Work affinity	Y	**.75**
												O	**.88**

Note. Significant correlations at $p < .05$ are in bold. Y = young listeners, O = older listeners.

indicate that younger and older listeners respond similarly on some measures, for example chronological age, but differently on other measures, such as perceived voice quality. In sum, these results confirm that listener factors, such as age, may influence the perception of a talker but that this influence depends upon the specific dimension being evaluated.

Influences of Talker Factors on Voice Evaluation

While our second hypothesis was that older and smoking talkers would receive more negative evaluations than younger and nonsmoking adults, the results reveal a more complex scenario. For the nonsmoking talkers, consistent with the literature, older male talker voices were perceived as more rough and breathy than younger male talkers (Benjamin, 1986; Gorham-Rowan & Laures-Gore, 2006; Harnsberger et al., 2010; Mulac & Giles, 1996; Prakup, 2012; Ryan & Burk, 1974; Ryan & Capadano, 1978). This is likely a direct consequence of the physiological and acoustic changes associated with aging (Dehqan et al., 2012; Stathopoulos, Huber, & Sussman, 2011; Xue & Deliyski, 2001). However, older female voices were not perceived differently than younger female talkers with one exception, perceived high pitch, which was rated more negatively for older female talkers compared with younger female talkers.

In general, perceived speech tempo was also rated lower in older talkers compared with younger talkers, meaning that older talkers were perceived as slower than younger talkers, consistent with prior studies (Harnsberger et al., 2010, 2008; Ryan & Burk, 1974). In the literature, slow speech tempo has been associated (at least in western societies) with poor ratings of self-confidence, extraversion, boldness, energy, dominance, and emotion in younger adults (Aronovitch, 1976), and poor psychosocial attribution on benevolence and competence dimensions in younger and older adults (Ryan & Johnston, 1987), thus suggesting that slower perceived speech tempo negatively affects psychosocial voice evaluation. Here, however, speech tempo in the older talkers did not correlate with voice-related psychosocial evaluations, suggesting that the interpretation of speech tempo may depend on the task or context and that a slower tempo in older adults is not always negatively perceived.

In this study, listeners were unaware of the fact that half of the talkers were regular smokers, which was done to avoid negative social biases. Yet, the results reveal that smoking voices received more negative psychosocial attributions (ratings closer to zero) compared with nonsmoking voices. A study that examined auditory-perceptual evaluation of smoking and nonsmoking voices also observed that smoking voices were perceived as being more strained and breathy than nonsmoking voices (Dedivitis et al., 2004). Given the numerous acoustic changes caused by cigarette smoking, including a decrease in fundamental frequency and voice stability measured by jitter, shimmer, and harmonics-to-noise ratio indices (Gonzalez & Carpi, 2004; Guimarães & Abberton, 2005; Pinto et al., 2014; Vincent & Gilbert, 2012), it is not surprising that cigarette smoking influences voice evaluation. This is the first study, to the best of our knowledge, to investigate differences in voice-related psychosocial attribution for nonsmoking and smoking talkers. Future work is needed to identify the acoustical parameters that are associated with the voice evaluation in smoking adults. A better understanding of the relationship between acoustical characteristics and perceptual evaluations is important to guide clinical interventions in patients with voice-related communication disorders. Indeed, the selection of voice-related objectives should consider the social impacts of different voice dimensions. If one parameter, for example, roughness, was found to be negatively associated with the quality or quantity of social interaction, treating roughness would become a priority for the treatment of voice disorders.

Importantly, our results demonstrate that perceived differences between younger and older talkers were different depending on talker sex and smoking status. Specifically, our results show that, in nonsmoking talkers, age differences were moderated by sex. Older male voices were perceived less positively than younger male voices on all scores, whereas this difference was limited to explicit scores (i.e., auditory-perceptual and perceived speech tempo scores) in female voices. Moreover, older female voices were more positively evaluated on warmness, agreeableness, and cheerfulness than younger female voices.

Instead, we found that, in smoking female talkers, aging was negatively associated with the evaluation of all voice dimensions except low pitch and that, in contrast, in smoking male talkers, aging was positively associated with voice evaluations on all psychosocial dimensions. In sum, whereas older female nonsmoking voices are evaluated more positively than younger voices, this pattern is reversed in smoking female talkers. For male voices, a similar reversal effect is also found, whereby younger male nonsmoking voices are evaluated more positively than older voices, a pattern that is flipped in smoking male talkers.

Dimensions Associated With the Propensity to Engage in Social Interactions

On the basis of the hypothesis that the perception of a talker may influence the propensity to engage in social interactions with that person (Amir & Levine-Yundof, 2013; Kemper et al., 1998; Lallh & Rochet, 2000), our third hypothesis was that the propensity to interact with a person would be associated with auditory-perceptual evaluations and even more so with voice-related psychosocial attributions, independently of age. Our results partially confirmed this hypothesis. Indeed, psychosocial attributions were positively associated with the propensity to engage in social interactions, but this relationship was observed twice as often in the younger listeners than in the older listeners. Moreover, in younger but not older listeners, a few auditory-perceptual dimensions (disorder, strain, and low pitch) were

positively associated with the propensity to engage in social interactions, meaning that in the younger listeners, an advantageous auditory-perceptual evaluation was associated with a higher propensity to engage in social interactions. Perceived speech tempo did not affect the inclination of a listener to engage in social interactions with a person. Thus, the propensity to interact with a person is associated with auditory-perceptual and voice-related psychosocial attributions to some extent and with some important age differences. Importantly, our paradigm seems to have better captured voice dimensions that influence younger adults' propensity to engage socially with a person than those that influence older adults. Additional studies are therefore needed to identify the dimensions more relevant to older adults.

The correlation analysis also revealed that most auditory-perceptual dimensions were positively associated with each other, but, surprisingly, only few auditory-perceptual dimensions were associated with voice-related psychosocial attributions, suggesting that voice-related psychosocial attribution relies on acoustic dimensions different from those that are used for auditory-perceptual evaluation. One such dimension could be the use of prosodic variations that are used to signal emotional and linguistic factors. The relationship between prosodic variations and psychosocial attribution will need to be explored in future research to further current understanding of the acoustical information that is used for psychosocial attributions during social interactions.

Surprisingly, a moderate correlation between high-pitch and low-pitch ratings was observed but only for older listeners. In a previous validation study, pitch ratings on the CAPE-V have been found to exhibit strong intrarater reliability coefficient (Zraick et al., 2011). In this study, however, unlike what is done in the CAPE-V, the evaluation of pitch was divided into a low-pitch and a high-pitch rating. This was done to avoid confusion, but it is possible that the two questions on pitch did confuse the older listeners. Alternatively, it is possible that both high-pitch and low-pitch voices occurred at different instances in the voice samples, thus resulting in a conflictual evaluation ("too high and too low pitch"). However, this pattern was only found in the older listeners, who evaluated the same samples as the young listeners. Replication of this finding is necessary to interpret it with more certainty.

Taken together, these results suggest that listeners may base psychosocial attribution on distinct perceptual characteristics depending on their age (Eppley & Mueller, 2001) and shed new light on the factors that influence the inclination to engage in social interactions with a person.

Limitations

This study does present a few limitations worth discussing. These limitations include a moderate sample size consisting mainly of highly educated individuals, a cross-sectional design with discontinuous age groups, the use of an adaptation of the CAPE-V to evaluate semispontaneous speech samples, and a focus on listener age. Though our sample included 50 adults, it was broken down into two age groups, each comprising 25 listeners. As detailed in the method section, on average, both groups were highly educated (~15 years), which corresponds approximately to a university degree in Québec. However, only approximately 50% of the population in Québec holds a college degree (Institut de la statistique du Québec, 2016). Additional studies are therefore needed to determine whether our results can generalize to the entire population, including individuals with lower education level. For both scientific and practical reasons, we did not break down our sample into gender groups. Scientifically, our main interest was in examining whether listener age affects voice evaluations. On a practical level, including more factors would have required a larger sample. Because listener sex was shown to have little or no influence on a listener's perceptual evaluation in previous studies (Amir & Levine-Yundof, 2013; Aronovitch, 1976; Hollien et al., 1991), the sex of the listener was not analyzed here.

Another limitation is the use of a cross-sectional design. Because of this design, we cannot exclude that other factors may contribute to explaining the lack of an association between listener age and voice evaluation. Further studies are needed with larger sample sizes, including older adults (80+ years) and a prospective longitudinal design with multiple measurements made over several decades. Such a design would allow researchers to draw causal inferences about voice evaluation and age.

In this study, we used semispontaneous speech samples instead of sustained vowels or completely unconstrained speech samples. Voice evaluation is often restricted to the use of sustained vowels. Yet, voice in connected speech is not only produced in a different way from sustained vowels (Lortie et al., 2015) but may also be perceived differently by listeners (Fourcin & Abberton, 2008; Maryn, Roy, De Bodt, Van Cauwenberge, & Corthals, 2009), as it offers more information about a person's voice quality and perceived psychosocial traits (Berry, 1991; Hughes & Rhodes, 2010; McAleer et al., 2014). Nonetheless, these semispontaneous speech samples may have introduced variability in the ratings. Indeed, although the vocabulary used by the talkers was constrained by the nature of the tales, the choice of words, the language register (e.g., informal), the quality of articulation, and the amount of prosodic variations differed across talkers. Thus, variability in the signal itself may have introduced variability in the ratings.

Another methodological choice that was made was to use an adaptation of the CAPE-V that was more suited for our intensive (~2 h) computer-based voice evaluation protocol than the original CAPE-V. Because three dimensions were modified (severity, pitch) or discarded (intensity) and the scale changed, the voice evaluation protocol used in this study does not have the validity and fidelity of the CAPE-V. However, the high interrater agreement (see Voice Evaluation section) strongly suggests that the voice evaluations that were conducted were reliable and are replicable. These modifications were necessary to facilitate the evaluation of a large number of samples (80) on a large

number of dimensions that extended beyond those evaluated by the CAPE-V to include voice-related trait attributions, perceived age, and perceived sex.

Finally, talkers in this study were not evaluated clinically (e.g., through indirect laryngoscopy) to assess the normality of their voice, which was self-reported. Although talkers were carefully selected through an interview process, we cannot exclude that some of them may have been suffering from undiagnosed voice disorders that could have influenced the results. However, voice analysis revealed that talker voice parameters were within normal limits. Moreover, the stringent outlier exclusion procedures reduced this risk. Finally, the large effect sizes for both listener (Cohen's *d* from 0.67 to 1.32, average of 0.91) and talker differences (Cohen's *d* from 0.5 to 1.99, average of 1.02) suggest that these results may be of clinical importance in selecting the most appropriate voice-related treatment objectives (in terms of voice dimensions) to facilitate social interactions by reducing the impact of factors detrimental to social communication.

Conclusions

This study demonstrates that voice evaluation is affected by the age, sex, and smoking status of the talker and, to a lesser extent, by the age of the listener, with older adults providing relatively more positive evaluations than younger adults. The results also reveal, for the first time, that most voice-related psychosocial attributions but only few auditory-perceptual scores are associated with the desire to interact socially with a person and that these two classes of dimensions are only weakly correlated. Understanding the dimensions that are used to determine whether a voice is considered warm or agreeable and the dimensions that have an impact on the quality of social communication is key to guide clinical interventions. The present results are a step toward this objective. Indeed, interventions should focus on the most functionally relevant acoustical dimensions of voice to improve social communications in patients with voice-related communication difficulties. Future studies are therefore needed to identify the acoustical parameters that form the basis of voice-related psychosocial attributions to help guide clinical interventions.

Acknowledgments

This study was supported by grants from the "Fonds de la Recherche du Québec Société-Culture" and the "Fonds de la Recherche du Québec en Santé" to Pascale Tremblay and from a graduate scholarship from the Natural Sciences and Engineering Research Council of Canada to Catherine L. Lortie. Pascale Tremblay and Matthieu J. Guitton hold career awards from the "Fonds de Recherche du Québec–Santé." The authors thank all the participants.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32,* 201–271.

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology, 16*(1), 4–13. https://doi.org/10.1207/s15327663jcp1601_2

Amir, O., Engel, M., Shabtai, E., & Amir, N. (2012). Identification of children's gender and age by listeners. *Journal of Voice, 26*(3), 313–321. https://doi.org/10.1016/j.jvoice.2011.06.001

Amir, O., & Levine-Yundof, R. (2013). Listeners' attitude toward people with dysphonia. *Journal of Voice, 27*(4), 524.e1–524.e10. https://doi.org/10.1016/j.jvoice.2013.01.015

Aronovitch, C. D. (1976). The voice of personality: Stereotyped judgments and their relation to voice quality and sex of speaker. *The Journal of Social Psychology, 99,* 207–220. https://doi.org/10.1080/00224545.1976.9924774

Auerbach, O., Hammond, E. C., & Garfinkel, L. (1970). Histologic changes in the larynx in relation to smoking habits. *Cancer, 25*(1), 92–104. https://doi.org/10.1002/1097-0142

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Benjamin, B. J. (1986). Dimensions of the older female voice. *Language & Communication, 6*(1–2), 35–45.

Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information. *Journal of Personality and Social Psychology, 61*(2), 298–307. https://doi.org/10.1037/0022-3514.61.2.298

Blanchet, C., Pommie, C., Mondain, M., Berr, C., Hillaire, D., & Puel, J.-L. (2008). Pure-tone threshold description of an elderly French screened population. *Otology & Neurotology, 29*(4), 432–440. https://doi.org/10.1097/MAO.0b013e3181719746

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer*. Amsterdam, the Netherlands: University of Amsterdam.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/https://doi.org/10.1007/BF02310555

Dedivitis, R. A., Barros, A. P. B., Queija, D. S., Alexandre, J. C. M., Rezende, W. T. M., Corazza, V. R., . . . Nishimoto, I. N. (2004). Interobserver perceptual analysis of smokers voice. *Clinical Otolaryngology and Allied Sciences, 29*(2), 124–127. https://doi.org/10.1111/j.1365-2273.2004.00793.x

Dehqan, A., Scherer, R. C., Dashti, G., Ansari-Moghaddam, A., & Fanaie, S. (2012). The effects of aging on acoustic parameters of voice. *Folia Phoniatrica et Logopaedica, 64*(6), 265–270. https://doi.org/10.1159/000343998

Deschamps, I., Hasson, U., & Tremblay, P. (2016). The structural correlates of statistical information processing during speech perception. *PLoS ONE, 11*(2), 1–19. https://doi.org/10.1371/journal.pone.0149375

D'haeseleer, E., Depypere, H., Claeys, S., Baudonck, N., & Van Lierde, K. (2011). The impact of menopause on vocal quality. *Menopause, 18*(3), 267–272. https://doi.org/10.1016/j.jvoice.2011.11.011

Eppley, B. D., & Mueller, P. B. (2001). Chronological age judgments of elderly speakers: The effects of listeners' age. *Contemporary Issues in Communication Science and Disorders, 28,* 5–8.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191.

Fourcin, A., & Abberton, E. (2008). Hearing and phonetic criteria in voice measurement: Clinical applications. *Logopedics, Phoniatrics, Vocology, 33*(1), 35–48. https://doi.org/10.1080/14015430701251574

Gonzalez, J., & Carpi, A. (2004). Early effects of smoking on the voice: A multidimensional study. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research, 10*(12), 649–656.

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders, 39*(3), 171–184. http://doi.org/10.1016/j.jcomdis.2005.11.005

Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of Voice, 27*(5), 545–555. https://doi.org/10.1016/j.jvoice.2013.03.002

Goy, H., Pichora-Fuller, M. K., & 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. https://doi.org/10.1121/1.4945094

Guimarães, I., & Abberton, E. (2005). Health and voice quality in smokers: An exploratory investigation. *Logopedics, Phoniatrics, Vocology, 30*(3–4), 185–191. https://doi.org/10.1080/14015430500294114

Harnsberger, J. D., Brown, W. S., Shrivastav, R., & Rothman, H. (2010). Noise and tremor in the perception of vocal aging in males. *Journal of Voice, 24*(5), 523–530. https://doi.org/10.1016/j.jvoice.2009.01.003

Harnsberger, J. D., Shrivastav, R., Brown, W. S., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *Journal of Voice, 22*(1), 58–69. https://doi.org/10.1016/j.jvoice.2006.07.004

Hirabayashi, H., Koshii, K., Uno, K., Ohgaki, H., Nakasone, Y., Fujisawa, T., . . . Hirabayashi, K. (1990). Laryngeal epithelial changes on effects of smoking and drinking. *Auris Nasus Larynx, 17*(2), 105–114. https://doi.org/10.1016/S0385-8146(12)80192-1

Hollien, H., Gelfer, M. P., & Carlson, T. (1991). Listening preferences for voice types as a function of age. *Journal of Communication Disorders, 24*(2), 157–171. https://doi.org/10.1016/0021-9924(91)90019-F

Honjo, I., & Isshiki, N. (1980). Laryngoscopic and voice characteristics of aged persons. *Archives of Otolaryngology, 106*(3), 149–150. https://doi.org/10.1001/archotol.1980.00790270013003

Huber, P. J. (1981). *Robust statistics*. New York: Wiley.

Hughes, S. M., & Rhodes, B. C. (2010). Making age assessments based on voice: The impact of the reproductive viability. *Journal of Social, Evolutionary, and Cultural Psychology, 4*(4), 290–304. https://doi.org/10.1037/h0099282

Huntley, R., Hollien, H., & Shipp, T. (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice, 1*(1), 49–52. https://doi.org/10.1016/S0892-1997(87)80024-3

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66–69. https://doi.org/10.1044/1058-0360.0603.66

Kemper, S., Ferrell, P., Harden, T., Billington, C., & Finter-Urczyk, A. (1998). Use of Elderspeak by young and older adults to impaired and unimpaired listeners. *Aging, Neuropsychology, and Cognition, 5*(1), 43–55.

Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Kempster, G. B., Kistler, D. J., & Hillenbrand, J. (1991). Multidimensional scaling analysis of dysphonia in two speaker groups. *Journal of Speech and Hearing Research, 34*(3), 534–543. https://doi.org/10.1044/jshr.3403.534

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences, 279*(1738), 2698–2704.

Lallh, A. K., & Rochet, A. P. (2000). The effect of information on listeners' attitudes toward speakers with voice or resonance disorders. *Journal of Speech, Language, and Hearing Research, 43*(3), 782–795.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766.

Linville, S. E. (1987). Acoustic-perceptual studies of aging voice in women. *Journal of Voice, 1*(1), 44–48. https://doi.org/10.1016/S0892-1997(87)80023-1

Lortie, C. L., Thibeault, M., Guitton, M. J., & Tremblay, P. (2015). Effects of age on the amplitude, frequency and perceived quality of voice. *Age, 37*(6), 1–24. https://doi.org/10.1007/s11357-015-9854-1

Ma, E. P. M., & Love, A. L. (2010). Electroglottographic evaluation of age and gender effects during sustained phonation and connected speech. *Journal of Voice, 24*(2), 146–152. https://doi.org/10.1016/j.jvoice.2008.08.004

Markel, N. N., Phillis, J. A, Vargas, R., & Howard, K. (1972). Personality traits associated with voice types. *Journal of Psycholinguistic Research, 1*(3), 249–255. https://doi.org/10.1007/BF01074441

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. *The Journal of the Acoustical Society of America, 126*(5), 2619–2634. https://doi.org/10.1121/1.3224706

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say "hello"? Personality impressions from brief novel voices. *PLoS ONE, 9*(3), 1–9. https://doi.org/10.1371/journal.pone.0090779

Mulac, A., & Giles, H. (1996). "You're only as old as you sound": Perceived vocal age and social meanings. *Health Communication, 8*(3), 199–215. https://doi.org/10.1207/s15327027hc0803

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Oswald, P. A. (1996). The effects of cognitive and affective perspective taking on empathic concern and altruistic helping. *The Journal of Social Psychology, 136*(5), 613–623. https://doi.org/10.1080/00224545.1996.9714045

Pinto, A. G. L., Crespo, A. N., & Mourão, L. F. (2014). Influence of smoking isolated and associated to multifactorial aspects in vocal acoustic parameters. *Brazilian Journal of Otorhinolaryngology, 80*(1), 60–67. https://doi.org/10.5935/1808-8694.20140013

Pittinsky, T., Shih, M., & Ambady, N. (2000). Will a category cue affect you? Category cues, positive stereotypes and reviewer recall for applicants. *Social Psychology of Education, 4*(1), 53–65.

Plank, C., Schneider, S., Eysholdt, U., Schützenberger, A., & Rosanowski, F. (2011). Voice- and health-related quality of life in the elderly. *Journal of Voice, 25*(3), 265–268. https://doi.org/10.1016/j.jvoice.2009.11.002

Prakup, B. (2012). Acoustic measures of the voices of older singers and nonsingers. *Journal of Voice, 26*(3), 341–350. https://doi.org/10.1016/j.jvoice.2011.05.007

Ptacek, P. H., & Sander, E. K. (1966). Age recognition from voice. *Journal of Speech and Hearing Research, 9*(2), 273–277. https://doi.org/10.1044/jshr.0902.273

Quebec Statistical Institute. (2016). *Répartition de la population de 25 à 64 ans selon le plus haut niveau de scolarité atteint, la région administrative, l'âge et le sexe, Québec* [Distribution of the population aged 25 to 64 by highest level of schooling, administrative region, age and sex. Quebec]. Retrieved from http://www.bdso.gouv.qc.ca/pls/ken/ken213_afich_tabl.page_tabl?p_iden_tran=REPERH7N9UA342130349047429Ax3W&p_lang=1&p_m_o=ISQ&p_id_ss_domn=824&p_id_raprt=3012

Raj, A., Gupta, B., Chowdhury, A., & Chadha, S. (2010). A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *Journal of Voice, 24*(3), 363–368. https://doi.org/10.1016/j.jvoice.2008.10.005

Richeson, J. A., & Ambady, N. (2001). Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles, 44*(9–10), 493–512.

Ryan, E. B., & Capadano, H. L. (1978). Age perceptions and evaluative reactions toward adult speakers. *Journal of Gerontology, 33*(1), 98–102.

Ryan, E. B., Giles, H., Bartolucci, G., & Henwood, K. (1986). Psycholinguistic and social psychological components of communication by and with the elderly. *Language & Communication, 6*(1–2), 1–24. https://doi.org/10.1016/0271-5309(86)90002-9

Ryan, E. B., & Johnston, D. G. (1987). The influence of communication effectiveness on evaluations of younger and older adult speakers. *Journal of Gerontology, 42*(2), 163–164.

Ryan, E. B., & Laurie, S. (1990). Evaluations of older and younger adult speakers: Influence of communication effectiveness and noise. *Psychology and Aging, 5*(4), 514–519. https://doi.org/10.1037/0882-7974.5.4.514

Ryan, R. F., McDonald, J. R., & Devine, K. D. (1955). The pathologic effects of smoking on the larynx. *AMA Archives of Pathology, 60*(5), 472–480.

Ryan, W. J., & Burk, K. W. (1974). Perceptual and acoustic correlates of aging in the speech of males. *Journal of Communication Disorders, 7*(2), 181–192. https://doi.org/10.1016/0021-9924(74)90030-6

Schvartz, K. C., & Chatterjee, M. (2012). Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech. *Ear and Hearing, 33*(3), 411–420. https://doi.org/10.1097/AUD.0b013e31823d78dc

Shipp, T., & Hollien, H. (1969). Perception of the aging male voice. *Journal of Speech and Hearing Research, 12*(4), 703–710. https://doi.org/10.1044/jshr.1204.703

Society of Obstetricians and Gynecologists of Canada. (2002). The Canadian Consensus Conference on Menopause and Osteoporosis 2002 Update. *Journal of Obstetrics and Gynaecology Canada, 24*(10), 1–91.

Stach, B. A. (2010). *Clinical audiology: An introduction* (Delmar). NY: Clifton Park.

Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research, 54*(4), 1011–1021. https://doi.org/10.1044/1092-4388(2010/10-0036)

Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*(5), 324–333. https://doi.org/10.1016/j.jcomdis.2009.03.001

Vincent, I., & Gilbert, H. R. (2012). The effects of cigarette smoking on the female voice. *Logopedics, Phoniatrics, Vocology, 37*(1), 22–32. https://doi.org/10.3109/14015439.2011.638673

Waldron-Perrine, B., & Axelrod, B. N. (2012). Determining an appropriate cutting score for indication of impairment on the Montreal Cognitive Assessment. *International Journal of Geriatric Psychiatry, 27*(11), 1189–1194. https://doi.org/10.1002/gps.3768

Xue, S. A., & Deliyski, D. D. (2001). Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. *Educational Gerontology, 27*(2), 159–168. https://doi.org/10.1080/03601270151075561

Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*(3), 1497.

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., … Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior, 13*(2), 67–82. https://doi.org/10.1007/BF00990791

Appendix

List of questions used in the voice evaluation protocol.

1.1. Auditory-perceptual block
- To what extent does this voice present a **disorder**?
(Selon vous, cette voix est-elle atteinte d'un trouble ?)
- To what extent is this voice **rough**?
(Selon vous, cette voix est-elle rauque ?)
- To what extent is this voice **strained**?
(Selon vous, cette voix est-elle tendue sous l'effort ?)
- Is this voice too **high pitched**?
(Selon vous, cette voix est-elle trop aigue ?)
- Is this voice too **low pitched**?
(Selon vous, cette voix est-elle trop grave ?)
- To what extent is this voice **breathy**?
(À quel point entendez-vous le souffle de la personne dans la voix ?)
1 = *not at all,* 2 = *a little,* 3 = *moderately,* 4 = *a lot,* 5 = *extremely*

1.2. Identification block
- How **old** do you think this person is?
(Selon vous, quel est l'âge de cette personne ?)
1 = 20–29, 2 = 30–39, 3 = 40–49, 4 = 50–59, 5 = 60–69, 6 = 70–79, 7 = 80+
- What **sex** do you think this person is?
(Selon vous, quel est le sexe de cette personne ?)
1 = male, 2 = female

1.3. Social potential block
- Is this voice **warm**?
(Selon vous, cette voix est-elle chaleureuse ?)
1 = *very cool,* 2 = *a little cool,* 3 = *neutral,* 4 = *a little warm,* 5 = *very warm*
- Is this voice **agreeable**?
(Selon vous, cette voix est-elle agréable ?)
1 = *very irritating,* 2 = *a little irritating,* 3 = *neutral,* 4 = *a little agreeable,* 5 = *very agreeable*
- Is the **speech tempo** adequate?
(Selon vous, la vitesse est-elle adéquate ?)
1 = *very slow*, 2 = *a little slow,* 3 = *neutral,* 4 = *a little fast,* 5 = *very fast*
Do you think it would be **easy to work** with this person?
(Croyez-vous qu'il serait facile de travailler avec cette personne ?)
1 = *very difficult,* 2 = *a little difficult,* 3 = *neutral,* 4 = *a little easy,* 5 = *very easy*
- Is this voice **cheerful**?
(Selon vous, cette voix est-elle enjouée ?)
1 = *very dull,* 2 = *a little dull,* 3 = *neutral,* 4 = *a little cheerful,* 5 = *very cheerful*
- Is this voice **confident**?
(Selon vous, cette vois est-elle hésitante ?)
1 = *very hesitant,* 2 = *a little hesitant,* 3 = *neutral,* 4 = *a little confident,* 5 = *very confident*
- Do you think it would be agreeable to **have a conversation** with this person?
(Croyez-vous qu'une conversation avec cette personne serait agréable ?)
1 = *very irritating,* 2 = *a little irritating,* 3 = *neutral,* 4 = *a little agreeable,* 5 = *very agreeable*