Behavior Research Methods (2020) 52:990–1007
https://doi.org/10.3758/s13428-019-01296-0

The Jena Speaker Set (JESS)—A database of voice stimuli from unfamiliar young and old adult speakers

Romi Zäske[1,2,3] • Verena Gabriele Skuk[1,2,3] • Jessika Golle[4] • Stefan R. Schweinberger[1,2,5,6]

Published online: 21 October 2019

Abstract

Here we describe the Jena Speaker Set (JESS), a free database for unfamiliar adult voice stimuli, comprising voices from 61 young (18–25 years) and 59 old (60–81 years) female and male speakers uttering various sentences, syllables, read text, semi-spontaneous speech, and vowels. Listeners rated two voice samples (short sentences) per speaker for attractiveness, likeability, two measures of distinctiveness ("deviation"-based [DEV] and "voice in the crowd"-based [VITC]), regional accent, and age. Interrater reliability was high, with Cronbach's α between .82 and .99. Young voices were generally rated as more attractive than old voices, but particularly so when male listeners judged female voices. Moreover, young female voices were rated as more likeable than both young male and old female voices. Young voices were judged to be less distinctive than old voices according to the DEV measure, with no differences in the VITC measure. In age ratings, listeners almost perfectly discriminated young from old voices; additionally, young female voices were perceived as being younger than young male voices. Correlations between the rating dimensions above demonstrated (among other things) that DEV-based distinctiveness was strongly negatively correlated with rated attractiveness and likeability. By contrast, VITC-based distinctiveness was uncorrelated with rated attractiveness and likeability in young voices, although a moderate negative correlation was observed for old voices. Overall, the present results demonstrate systematic effects of vocal age and gender on impressions based on the voice and inform as to the selection of suitable voice stimuli for further research into voice perception, learning, and memory.

Keywords Vocal age · First impressions · Personality · Voice space · Voice database · Speech corpus

Similar to faces, voices carry a wealth of information about a speaker, even if the semantic content of a message is absent or incomprehensible. Among these nonlinguistic vocal signals are, for instance, cues to speaker identity, gender, age, affective state or personality impressions (for a review, see Schweinberger, Kawahara, Simpson, Skuk, & Zäske, 2014). Although the human capacity to extract, process and retain these signals is essential to navigate in our social environment, relatively little is known about human voice cognition abilities. The relatively scarce representation of voice research (as compared to face research) may be partly due to the difficulty to access a sufficient number of standardized stimuli. By contrast, equivalent databases for faces have been available for a couple of years (e.g., CAL/PAL Face Database by Minear & Park, 2004) and have fueled face research ever since. Although the linguistic community offers a broad and versatile range of speech corpora (e.g., Linguistic Data Consortium [LDC], University of Pennsylvania; Bavarian Archive for Speech Signals [BAS], Ludwig-Maximilians-Universität of

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13428-019-01296-0) contains supplementary material, which is available to authorized users.

✉ Romi Zäske
romi.zaeske@gmail.com

[1] DFG Research Unit Person Perception, Friedrich Schiller University of Jena, Jena, Germany

[2] Department for General Psychology and Cognitive Neuroscience, Institute of Psychology, Friedrich Schiller University of Jena, Jena, Germany

[3] Department of Otorhinolaryngology, Jena University Hospital, Jena, Germany

[4] Hector Research Institute for Education Sciences and Psychology, Eberhard Karls Universität Tübingen, Tübingen, Germany

[5] Michael Stifel Center Jena for Data-Driven and Simulation Science, Friedrich Schiller University of Jena, Jena, Germany

[6] Swiss Center for Affective Sciences, Campus Biotech, Geneva, Switzerland

Munich), these databases are typically unsuitable for experimental voice research, which often requires large numbers of highly standardized voice samples in terms of speech content and recording conditions. Furthermore, existing databases usually contain only a few different speakers, or offer a relatively limited corpus of different utterances (Ferdenzi et al., 2015). Other recent databases, such as the Montreal Affective Voices (Belin, Fillion-Bilodeau, & Gosselin, 2008), the Oxford Vocal Sounds Database (Parsons, Young, Craske, Stein, & Kringelbach, 2014), or a recent corpus of affective vocalizations from online amateur videos (Anikin & Persson, 2017) are highly specialized and only offer nonlinguistic affective sounds. Although such sounds may communicate signals beyond mere affect (e.g., Raine, Pisanski, & Reby, 2017), they are more typically used for research into the processing of affective prosody. By providing a substantial set of voice recordings from 120 adult speakers, we here aim to provide a significant new resource to voice research.

The present research considers other vocal signals beyond affective prosody that have attracted considerable scientific attention in the past few years. For instance, studies on voice adaptation uncovered the mental representation of speaker gender (Schweinberger et al., 2008; Skuk, Dammann, & Schweinberger, 2015), identity (Latinus & Belin, 2011; Zäske, Schweinberger, & Kawahara, 2010), and age (Zäske & Schweinberger, 2011; Zäske, Skuk, Kaufmann, & Schweinberger, 2013). Others have started to target the acoustic and perceptual correlates of attractiveness (Babel, McGuire, & King, 2014; Bestelmeyer et al., 2012; Bruckert et al., 2010; Feinberg, DeBruine, Jones, & Perrett, 2008), trustworthiness and dominance (McAleer, Todorov, & Belin, 2014), or investigated the interplay in the processing of nonlinguistic and linguistic information (Formisano, De Martino, Bonte, & Goebel, 2008; Perrachione, Del Tufo, & Gabrieli, 2011; Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015; von Kriegstein & Giraud, 2004; Zarate, Tian, Woods, & Poeppel, 2015; Zäske, Volberg, Kovacs, & Schweinberger, 2014). Importantly, voice researchers currently devote large efforts into understanding not only (1) the neural correlates of voice perception, learning and memory (e.g., Babel et al., 2014; Latinus, Crabbe, & Belin, 2011; Schweinberger et al., 2008; Zäske et al., 2014), but also (2) the role of specific characteristics of speakers´ voices as expressed either in acoustic measurements or perceptual ratings (e.g., Baumann & Belin, 2010; Skuk & Schweinberger, 2014), and (3) individual differences between listeners in their ability to perceive and recognize voices (Aglieri et al., 2017; Garrido et al., 2009; Mühl, Sheil, Jarutyte, & Bestelmeyer, 2018; Skuk & Schweinberger, 2013).

Due to these developments, we anticipate a growing demand for a well-documented set of high-quality voice stimuli suitable for the study of a whole range of social signals. Such a resource should be highly valuable for researchers who wish to assess the acoustic, perceptual and neural correlates of voice perception and memory across a large number of different speakers. To optimally investigate the manifold aspects of voice cognition, we believe it is desirable to have a range of standardized yet ecologically valid stimuli (e.g., including not only brief vowels but also naturalistic speech) to choose from.

The Jena speaker set (JESS) offers the scientific community a freely available and exceptionally large corpus of various utterances from 120 young and old adult speakers. Recordings for all speakers include a variety of utterances—that is, sustained vowels, syllables, sentences, read standardized text, and semi-spontaneous speech from picture descriptions. The data reported in this article are based on a subset of utterances of two sentence stimuli spoken by all speakers. For each of 120 speakers, we provide these two sentences along with further sentences (in total $N = 31$, and for young speakers $N = 14$ additional sentences), two vowel–consonant–vowel (VCV) syllables, five sustained vowels, one read text, and two samples of semi-spontaneous speech by the same speakers (cf. Supplemental Table S1) via the following link: https://osf.io/m5zdf/. We anticipate that vowel and syllable stimuli void of semantic content may be particularly useful for cross-cultural studies. Moreover, German sentence stimuli may also be used in nonnative listeners of German: For instance, although voice recognition is more difficult for voices in an unfamiliar language (Goggin, Thompson, Strube, & Simental, 1991; Perrachione & Wong, 2007), performance remains transferable between languages with high degrees of phonological overlap such as German and English (Zarate et al., 2015).

Importantly, apart from young adults, JESS offers voice samples of *old* adult speakers as well, which are otherwise hard to obtain. By including two age categories of adult voices we particularly hope to foster research into vocal age, a relatively neglected yet important social signal (for a recent overview, see Latinus & Zäske, 2019). The few studies that exist today, suggest moderate accuracy of vocal age judgments overall, with accuracy depending critically on the response format for age judgments (exact age estimation vs. coarse age categorization), the listener´s age and gender, and the kind and length of stimulus material (Linville, 1996). As a general consensus, perceived and chronological speaker age correlate (e.g., Bruckert, Lienard, Lacroix, Kreutzer, & Leboucher, 2006; Harnsberger, Brown, Shrivastav, & Rothman, 2010; Huntley, Hollien, & Shipp, 1987; Moyse, 2014; Ryan & Burk, 1974; Shipp & Hollien, 1969).

Studies investigating perceptually relevant acoustic cues to age reported reduced speaking rate (Brown, Morris, & Michel, 1989; Harnsberger, Shrivastav, Brown, Rothman, & Hollien, 2008) and increased variability of fundamental frequency (F0; Gorham-Rowan & Laures-Gore, 2006; Linville, 1996; Ramig et al., 2001; Torre & Barlow, 2009) as markers for older age, in both female and male voices. By contrast, acoustic parameters that have been reported to change in a gender-specific fashion include the mean F0 (see, e.g., Baken, 2005, for a review), first-formant frequency (F1; e.g., Reubold, Harrington, & Kleber, 2010), harmonics-to-noise ratio (HNR), and voice onset time (VOT; Linville, 1996; Stathopoulos, Huber, & Sussman, 2011; Torre & Barlow, 2009). We expect that the present voice database can be used to further contribute to the identification of acoustic parameters associated with chronological and perceived age.

In the present report, we focus on putative effects of speaker age on various person impressions from the voice, in order to exemplify possible research questions that can be addressed with the JESS. Specifically, we explore both effects of chronological speaker age, and their possible modulations by speaker sex, on perceptions of attractiveness, likeability, distinctiveness, strength of regional accent, and age. Selection of these dimensions has been partially inspired by previous relevant research into the perception and memory of faces. Specifically, a study on the CAL/PAL faces revealed that old as compared to young adult faces are perceived as being less attractive and less likeable (Ebner, 2008). As a qualification, face research has operationalized distinctiveness using various measures, the most common of which are deviation-based distinctiveness (DEV) and "face-in-the-crowd" distinctiveness (FITC). In deviation-based ratings, participants are asked to estimate how much a face differs from other typical faces they know. In FITC ratings, participants estimate how likely they would spot a face in a crowd of people. Importantly, because these two measures do not necessarily reflect the same construct in the face and voice domain (Wiese, Altmann, & Schweinberger, 2014; Zäske, Schweinberger, & Skuk, 2018b), the present investigation considers both types of distinctiveness ratings as adapted for voices.

To determine whether perceived distinctiveness is associated with regional accent, we also assessed ratings for the strength of regional accent. Moreover, regional accent may be regarded as a subtle marker of group membership, similar to facial cues of ethnicity (although the biological origin of facial ethnicity contrasts with the sociocultural origin of accent in voices). Although we are unaware of studies indicating differential *perception* of face ethnicity as a function of face age, both facial attributes appear to interactively modulate recognition *memory* for faces (Wallis, Lipp, & Vanman, 2012; Wiese, 2012). Because most of the above social cues are transmitted by both faces and voices, which are believed to be processed and coded in a similar fashion (Yovel & Belin, 2013), a default expectation could be that effects of speaker age on voice perception would broadly parallel previously reported age effects on face perception. The above dimensions (and distinctiveness, age, and attractiveness in particular) are known to systematically modulate face memory (e.g., Bruce & McDonald, 1993; Meissner & Brigham, 2001; Schulz, Kaufmann, Kurt, & Schweinberger, 2012; Wiese et al., 2014) and, to some extent, voice memory (Zäske, Limbach, et al., 2018). We anticipate that the present database will enable researchers to test analogous predictions regarding the role of these dimensions for voice memory.

Taken together, the first objective of the present article is to describe the JESS in order to promote research on voice perception and person perception in general, as well as related fields. The present results are based on young adult listeners who provided extensive ratings for two sentences per speaker on attractiveness, likeability, deviation-based distinctiveness (DEV) and "voice-in-the-crowd"-based distinctiveness (VITC), strength of regional accent, and age. Note that we assessed two common measures of distinctiveness, because both might measure slightly different aspects of distinctiveness (Wickham & Morris, 2003), and because "in the crowd" measures of distinctiveness may be more prone to bias in response to attractive stimuli (Wiese et al., 2014; Zäske, Schweinberger, & Skuk, 2018b). We report inter-rater consistency measures as well as mean ratings for speakers, both on the individual and group level, in order to provide future studies with a basis for stimulus selection. The second objective of this article is to provide an example for how the database can be used in combination with rating results in order to explore open questions in voice research. Specifically, we use multilevel regression analyses and inter-dimension correlations in order to explore the combined roles of vocal age and sex for perceptions of the above speaker attributes by male and female listeners.

Method

Recording

Speakers Overall we recorded voices, videos and took photographs of 64 young and 62 old adult speakers who we recruited from the local community. Note that here we describe a subset of 120 speakers who agreed to provide their voice recordings for other researchers, although the experiments involved all 126 speakers. The subset of 120 speakers constitutes the Jena Speaker Set (JESS) and includes voice recordings of 61 young (*M* = 21.8 years, *SD* = 2.2, range: 18–25 years) and 59 old (*M* = 67.6 years, *SD* = 4.8, range: 60–81 years) female and male native speakers of German. Note that we selected target age ranges of the two age groups (7-year range in young, 21-year range in old speakers) to be in approximate proportion to their absolute mean target ages. The young sample was comprised of only students, whereas the old sample included both pensioners and employees. For detailed speaker information including sex, age, body size and weight, profession, occupation, personality traits, places of living or smoking habits, confer Supplemental Table S3. Young female (*N* = 30, *M* = 21.8 years, *SD* = 2.4, range:

18–25 years) and male speakers ($N = 31$, $M = 21.8$ years, $SD = 2.1$, range: 18–25 years) did not significantly differ with respect to age [$t(59) = – 0.009$, $p = .992$]. The same was true for the old female ($N = 29$, $M = 67.5$ years, $SD = 2.4$, range: 60–77 years) and male ($N = 30$, $M = 67.7$ years, $SD = 4.8$, range: 60–81 years) speakers [$t(57) = – 0.146$, $p = .884$]. Due to the novelty of the present research, we were unable to calculate a priori power for finding significant correlations between rating dimensions. However, we used G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) Version 3.1.9.2 to determine the number of speakers required to find significant correlations between rating dimensions (when averaged across individual raters, using speakers as cases). To detect correlations of $\rho \geq .5$ with a power of .8, a sample of 29 speakers would be required (actual power with 30 speakers = .83). To detect smaller correlations of $\rho \geq .3$ with a power of .8, a sample of 84 speakers would be required (actual power with 120 speakers = .92). Accordingly, we considered the number of speakers in the dataset as adequate for the purposes of the present study.

Procedure From each of the young speakers we recorded 42 sentences that had the same syntactic structure and consisted of seven or eight syllables. A third of the sentences started with the German articles “der,” “die,” and “das,” respectively. We also recorded 12 further sentences, six VCV syllables, 12 consonant–vowel–consonant–vowel (CVCV) syllables, seven sustained vowels, a standardized text (“Der Nordwind und die Sonne,” Aesop), and spontaneous descriptions of two line-drawn pictures showing a farmyard (http://ausmalbildertop.com/bauernhof-3/, last accessed June 14, 2019) and a kitchen scene, the so-called “Boston Cookie Theft” scene (Goodglass & Kaplan, 1983). The line-drawings were unknown to the speakers who were instructed to describe them in detail. The protocol for the old speakers was shortened to keep the duration of the recording sessions within reasonable limits (~2 to 2.5 h, including 30–45 min for questionnaires and breaks). The protocol was identical to that of the young speakers with the exception that it did not contain “das” sentences, six of the additional sentences (of variable syntactic structure), and the CVCV syllables. A complete list of the recording protocol and the final stimulus set of the JESS can be found in the supplemental materials (https://osf.io/m5zdf/) along with all supplemental tables (S1–S6). Although we provide an extensive set of stimuli for other researchers (per speaker, two VCV syllables, five sustained vowels, one read text, two samples of semi-spontaneous speech, and 31 sentences, plus 14 additional sentences for young speakers only), the present evaluation of the JESS database, including acoustic measurements (see the supplemental material), was performed on two exemplary sentences: (1) “Der Fahrer lenkt den Wagen.” (“The driver steers the car.”) and (2) “Die Kundin kennt den Laden.” (“The customer knows the shop.”).

Prior to the recording session speakers filled out consent forms and various questionnaires as summarized in Supplemental Table S3. These data include self-reports on body height, weight, smoking habits, and assessments of personality traits of each speaker as well as occupation/profession, self-reported regional accents and places of living. Young speakers filled out the German version (Borkenau & Ostendorf, 2008) of the 60-item Big-Five inventory by McCrae and Costa (1987). To shorten the procedure for the old speaker group, a 10-item Big-Five inventory was used instead (Rammstedt, Kemper, Klein, Beierlein, & Kovaleva, 2013). All speakers performed a German version of the 50-item autism spectrum quotient questionnaire (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001; Freitag et al., 2007), and of a 10-item shyness and sociability inventory (Asendorpf & Wilpers, 1998; Neyer & Asendorpf, 2001).

The recordings were obtained in a quiet and semi-anechoic room. First we took three photographs of each speaker’s face (one frontal, two profiles) by means of a Sony DCR-DVD403E camcorder. For that purpose speakers sat on a chair in front of a green background and were illuminated by a three-point lighting system. To standardize visual stimuli, speakers were asked to take off glasses, jewelry and make-up and, if applicable, to shave before the session. All wore a black cape. Using the same setup, audio and video recordings were then obtained simultaneously. Video recordings were obtained for the entire protocol with the exception of the standardized text and the picture descriptions for which some speakers required glasses. To standardize intonation and duration of the utterances (sentences, syllables, vowels) and to keep regional accents to a minimum, speakers were encouraged to repeat utterances from a pre-recorded model speaker presented via loudspeakers. Voices were recorded with a Sennheiser MD 421-II microphone with a pop protection and a Zoom H4n audio interface (16-bit resolution, mono, 48 kHz). The audio interface was connected to a computer in the neighboring room at which the audio manager monitored the recordings via Audobe Audition 3.0. Speakers were instructed to intonate these utterances as emotionally neutral as possible and to close their mouths between utterances while directly facing the camera and looking at a predetermined point above the camera lens. Each utterance was recorded several times (usually four or five times) until the session manager was satisfied with the vocal and facial performance. Speakers were encouraged to take self-paced breaks and to drink still water whenever needed.

The best audio recordings of each of the two sentences were chosen (in terms of artifacts, background noise, clear pronunciation). Using PRAAT software (Boersma, 2001) voice recordings were cut to contain one sentence starting exactly at plosive onset of “Der”/“Die.” Voice recordings were then resampled to 44.1 kHz and RMS normalized to 70dB. Mean stimulus duration was 1,762 ms (*SD* = 184 ms, range: 1,396–2,413 ms). An analysis of variance (ANOVA) on mean sentence durations with repeated measures on sentence (#1 vs. #2) and with vocal age group (VA: young vs. old) and voice sex (VS: male vs. female) as between-subjects factors revealed no significant effects. Please note that comprehensive acoustic analyses of our stimuli can be found in the supplemental materials as well as in Supplemental Tables S4–S6.

Validation

Raters Twenty-four student raters (12 female, four left-handed, all native speakers of German, mean age = 23.0 years, range: 18–30 years) contributed rating data. Raters were tested in two sessions of ~ 60 min duration each. They came back for the second session after 7.5 days on average (*SD* = 1.4 days, range: 6–11 days). None reported hearing difficulties or prior familiarity with any of the voices used in the experiment. Data from eight additional raters who were either familiar with voices (*N* = 4), did not return for the second session (*N* = 2), responded extremely fast (*N* = 1 who on average gave responses during rather than after voice presentation, suggesting premature responses) or highly similar across trials (*N* = 1) were excluded from the analyses. Raters received a payment of €10 or course credit. All gave written informed consent. The study was conducted in accordance with the Declaration of Helsinki, and was approved by the Faculty Ethics Committee (Ethics Vote FSV 12/02) of the Friedrich Schiller University of Jena.

Procedure Raters were tested individually in a sound-attenuated chamber. To avoid interference from the experimenter's voice, the experimenter did not talk to raters during the testing breaks, and all instructions were presented in writing on the computer screen. Voice stimuli were presented diotically via Sennheiser HD 212Pro headphones with an approximate peak intensity of 60 dB(A) as determined with a Brüel & Kjær Precision Sound Level Meter Type 2206. In each session, raters rated all voices on the basis of one sentence on each of the six dimensions: attractiveness, likeability, DEV and VITC distinctiveness, strength of regional accent, and age. Note that separate ratings for each dimension (rather than ratings for multiple dimensions on a single trial) were obtained to avoid spill-over or generalization effects across dimensions. Rating dimensions were presented blockwise with each block containing all 126 speakers uttering one of the two sentences. Overall, 756 trials (126 × 6) were presented per session and rater. Individual breaks were allowed after each young and old voice block. Within a given session young and old speakers uttered different sentences. To accommodate two sentences per speaker the assignment of speakers to sentences was reversed in the second session and counterbalanced across raters.

Trials started with a black fixation cross for 1,000 ms on a gray background. Upon voice presentation (one sentence) the fixation cross disappeared and a reminder of the current task and the response alternatives appeared at the bottom of the screen. Raters were instructed to enter their response via number keys in the upper row of a computer keyboard. Age ratings were entered as direct estimates using two-digit responses (10–99 years). For the remaining traits, Likert scales were used on which the numbers 1 and 6 marked the lower and higher ends of a given trait spectrum (e.g., 1 = *very unattractive* vs. 6 = *very attractive*). For the exact instructions and response labels, please refer to the Appendix, Table 5. There was no time limit for the responses, but raters were encouraged to respond spontaneously and as accurately as possible.

The order of rating dimensions was randomized across raters. Each scale contained a young and an old voice block, which were further subdivided into a female and a male voice block, respectively. The order of young and old, female and male blocks was counterbalanced across raters and remained the same for a given rater for all rating dimensions and in both sessions. Within blocks voices were presented randomly.

Results

Because we were interested in spontaneous ratings as instructed, only responses given within 200 to 6,000 ms from voice onset were considered. For the age ratings this window was extended to 8,000 ms, to consider increased time demands for a double-digit response, as determined preexperimentally in pilot runs. In total 2.03% of trials of the 120 speakers were excluded from the analyses (i.e., 1.25%, 0.83%, 1.77%, 1.84%, 1.67%, and 4.81% for attractiveness, likeability, DEV distinctiveness, VITC distinctiveness, strength of regional accent, and age, respectively). Mean rating scores for each speaker on each rating dimension can be found in the supplemental material (Table S2). Moreover, substantial exploratory acoustic analyses can be found in supplemental materials (Figs. S1–S3 and Tables S4–S6).

Internal consistency

Interrater reliability (Cronbach's α)—that is, the agreement between raters—was computed separately for each of the six dimensions, with ratings collapsed across the two sentences. Cronbach's α ranged between .818 and .994 for young voices, and between .832 and .990 for old voices. To determine the intrarater reliability across sentences—that is, the correspondence of ratings within raters (Sentences #1 and #2)—we also computed Cronbach's α on the raw—that is, noncollapsed—data (Table 1).

Correlations between and within dimensions

We correlated the mean ratings of the six dimensions separately for young and old voices, resulting in 30 interdimension correlations. Rank correlations (Spearman's ρ) are depicted in Tables 2 and 3 (cf. also Supplemental Fig. S1). Furthermore, to explore the extent to which voice properties are perceived reliably across sentence contents, we correlated ratings between Sentences #1 and #2 for each dimension (see Tables 2 and 3, bottom rows). All correlations in this study, including the supplemental materials, are reported with uncorrected p-values. To account for a possible inflation of Type I error due to multiple testing, we applied Bonferroni correction to the correlations of main interest (N = 42 correlations reported in Tables 2 and 3). The majority (33/35) of the significant correlations survived Bonferroni correction, but two correlations were rendered nonsignificant, as described below.

The correlations between dimensions pointed to similarities and differences in the perception of young and old voices: Rated likeability and attractiveness were strongly correlated in both vocal age groups. Also, more likeable and attractive voices were rated as being less distinctive. As a qualification, for young voices this was only true for DEV (but not VITC) distinctiveness. In both age groups, DEV and VITC distinctiveness correlated positively; that is, the more a voice was rated as differing from known voices, the more likely it was considered to be spotted in a crowd of people speaking simultaneously. Note that for the young speakers only, this correlation did not survive Bonferroni correction for 42 tests (p_{uncorr} = .006). Strong perceived regional accents were associated with lower attractiveness and likeability ratings, but with higher distinctiveness ratings, for both young and old voices. Whereas a strong regional accent was associated with higher ratings on both distinctiveness measures in old voices, this association was only significant with DEV (but not VITC) distinctiveness in young voices. The most notable difference between young and old voices was that perceived age within the group of old voices correlated with all dimensions, such that increased perceptions of age were associated with increased ratings of distinctiveness and regional accent, and with decreased ratings of attractiveness and likeability. By contrast, perceived age within the group of young voices did not correlate significantly with those dimensions.[1] With respect to the correlations within dimensions, ratings for Sentences #1 and #2 were positively correlated on all dimensions and for both vocal age groups. Note that one between-sentence correlation (DEV distinctiveness ratings for young speakers) did not survive Bonferroni correction (p_{uncorr} = .003).

Precision of the age ratings

On average, the age of young voices was overestimated ($M_{\text{rated age}}$ = 28.9 years, SD = 2.2 vs. chronological age of $M_{\text{chron\_age}}$ = 21.8 years, SD = 5.6), whereas the age of old voices was underestimated ($M_{\text{rated age}}$ = 56.2 years, SD = 4.8 vs. chronological age of $M_{\text{chron\_age}}$ = 67.6 years, SD = 5.8); compare Fig. 1 and Table 6 (in the Appendix). Of note, the age ratings span the entire age continuum between our highly circumscribed age groups. Spearman rank correlations performed separately for each vocal age group revealed that perceived age and chronological age were significantly and positively correlated for old male and young female speakers (ρ = .402, p = .028, and ρ = .442, p = .015), but not for old female and young male speakers (ρ = .132, p = .495, and ρ = .001, p = .994); compare Fig. 1. For further exploratory acoustic analyses, please refer to the supplemental material.

Mean ratings of young and old voices

Figure 2 summarizes the mean rating results for all dimensions. We used multilevel regression for repeated measurements (Hoffman & Rovine, 2007; Hox, 2002) to determine whether the ratings of attractiveness, likeability, distinctiveness (DEV, VITC), strength of accent, and age depended on the age group of the speakers (young vs. old), the sex of the speakers, and/or the sex of the raters. For each of the six rating dimensions, we computed a separate model using mean ratings (i.e., collapsed across the two sentences) as the dependent variable.[2] On the within-person level (Level 1), we entered voice age (VA: young vs. old) and voice sex (VS: male vs. female) as predictor variables. On the between-person level (Level 2), we used rater sex (RS: male vs. female) as a predictor variable. All possible Level 1 and cross-level interactions were also entered into the analysis. A random-coefficient model was calculated (see Eq. 1) without any constraints regarding the estimation of variances and covariances of the Level 2 residuals. Robust estimators were used for statistical inference with respect to the fixed effects and variance components, to account for possible violations of the model assumptions, such as normality of the Level 2 residuals. Degrees of freedom were computed on the basis of Satterthwaite's approximation, to account for the moderate sample size at Level 2 (Satterthwaite, 1946). Therefore, the degrees of freedom were not necessarily integers and could vary across tests independently of the number of parameters.

[1] Note that the variances of the age ratings were similar for young and old voices (Table 6), and therefore can not account for the lack of significant correlations for young voices.

[2] To explore potential effects of sentence content, we performed the same analysis with sentence (1 vs. 2) as an additional first-level predictor variable. The only significant effects involving the factor sentence were found for ratings of VITC (interaction of VA × VS × Sentence × RS [$\beta_{\text{VITC}} = -.433$, p = .043] and regional accent (interactions of VS × Sentence [$\beta_{\text{acc}} = -.284$, p < .001] and of VS × Sentence × RS [β_{acc} = .398, p = .003]). Note that the effects found in the original analysis (without sentence as a variable) remained similar and significant overall. The only exception was that the effect of VS in the ratings of regional accent disappeared when sentence was added as a variable.

Table 1 Intrarater reliability across sentences (Cronbach's α) for all voices, and separately for young and old voices

Voices	*N*	Attractiveness	Likeability	Distinctiveness (DEV)	Distinctiveness (VITC)	Accent	Age
All	120	.610	.554	.410	.449	.611	.882
Young	61	.507	.565	.429	.433	.544	.655
Old	59	.583	.516	.372	.415	.575	.599

Table 4 shows the results for all rating dimensions, with fixed effects (VA, VS, and RS, the Level 1 interaction VA × VS, and the cross-level interactions VA × RS, VS × RS, and VA × VS × RS), and random effects (VA, VS, and VA × VS).

$$\begin{aligned} &\mathit{Level}\ \mathrm{I}: Y_{it} = \alpha_{0t} + \alpha_{1t}VA + \alpha_{2t}VS + \alpha_{3t}VA \times VS + \varepsilon_{it} \\ &\mathit{Level}\ \mathrm{II}: \forall \alpha_{jt}(j = 0, \dots, 3): \alpha_{jt} = \beta_{j0} + \beta_{j1}RS + \upsilon_{jt} \end{aligned} \tag{1}$$

Attractiveness The fixed effect of VA revealed that male raters rated male voices as being more attractive when voices were young than when they were old ($\beta_{\text{attr10}} = -.51, p < .001$). The same was true for female raters, due to a lack of an interaction of VA with RS. The effect of VA on voice attractiveness was qualified by a Level 1 interaction of VA × VS ($\beta_{\text{attr11}} = -.78, p < .001$), suggesting that the effect of age was more pronounced when male raters rated female as opposed to male voices. The cross-level interaction effect of VA × VS × RS ($\beta_{\text{attr31}} = .72, p = .001$) indicated that the interaction of voice age and voice sex was less pronounced for female raters.

Likeability We found no effect of VA, suggesting that likeability did not differ significantly between young and old male voices ($\beta_{\text{like10}} = -.13, p = .224$). However, an effect of VS ($\beta_{\text{like20}} = .18, p = .022$) suggested that raters liked young female voices better than young male voices. The Level 1 interaction effect of VA and VS ($\beta_{\text{like30}} = -.35, p = .001$) suggested a significantly larger effect of VA in female than in male voices, such that young female voices were more likeable than old female voices, as compared to the difference between young and old male voices. None of these effects were further modulated by RS, suggesting that female raters showed a comparable pattern of results.

Distinctiveness In terms of DEV distinctiveness, old voices were rated as being more distinctive than young voices ($\beta_{\text{dev10}} = .19, p = .043$). This was true for both voice sexes and rater groups, as no further interactions were observed. Moreover, no significant effects were observed for VITC distinctiveness (Fig. 2).

Regional accent An effect of VA revealed that regional accent was perceived as stronger in old than in young voices ($\beta_{\text{acc10}} = .72, p = .007$). No interaction effects involving VA were statistically significant. An effect of VS ($\beta_{\text{acc20}} = -.25, p = .003$) suggested that female voices had less pronounced accents than male voices. No interaction effects involving VS were statistically significant.

Age As expected, the effect of VA suggested that male raters judged old male voices as older than young male voices ($\beta_{\text{age10}} = 25.38, p < .001$). The lack of significant two-way interactions of VA with VS or RS suggests that this was also true when male raters judged female voices and when female raters judged male voices. Furthermore, the effect of VS suggested that male raters perceived young male voices as being older than young female voices ($\beta_{\text{age20}} = -5.12, p = .001$). The lack of two-way interactions of VS with VA or RS suggests that this was also true when male raters rated old voices and when female raters rated young voices. As a qualification, the cross-level interaction effect of VA × VS × RS ($\beta_{\text{age31}} = 5.08, p = .038$) indicated that the effect of voice age was most pronounced in female raters rating female as compared to male voices.

Discussion

Here we present a novel voice database from 120 young (61) and old (59) unfamiliar adult speakers (half male) suitable for experimental voice research and related disciplines. The Jena Speaker Set (JESS) contains various standardized high-quality voice samples that are freely available for the scientific community via the following link: https://osf.io/m5zdf/. For all speakers, the material comprises 31 sentences (plus 14 additional sentences for young speakers), two VCV syllables, five sustained vowels, one read text and two samples of semi-spontaneous speech as documented in supplemental Table S1. We thereby hope to overcome the present lack of databases offering a sufficiently large and standardized, as well as validated corpus of utterances and speakers, particularly old speakers. Apart from vocal age, this database can be used to study acoustic, perceptual and neural correlates of various vocal signals including identity, gender, attractiveness, distinctiveness, or perceived personality traits. Similar to existing databases for faces (e.g., CAL/PAL Face Database by Minear & Park, 2004) we expect that the JESS will further stimulate research into voice perception, learning and memory.

To characterize the new database, we provide ratings of the JESS on six socially relevant dimensions, separately for individual speakers (cf. the supplemental material, Table S2) and for groups of young and old male and female speakers (cf. the

Table 2 Young voices: Correlations (Spearman's ρ) between rating dimensions (collapsed across sentences) and between sentences (S1 and S2) for each dimension (bottom row)

Young voices ($N = 61$)	Attractiveness	Likeability	Distinctiveness (DEV)	Distinctiveness (VITC)	Accent	Perceived Age
Likeability	.87***					
Distinctiveness (DEV)	– .67***	– .65***				
Distinctiveness (VITC)	– .13	– .15	.35(**)			
Accent	– .75***	– .74***	.54***	.08		
Perceived Age	– .05	.04	.02	.17	.11	
Between S1 and S2	.50***	.58***	.38(**)	.49***	.43***	.86***

$^{***}p < .001$ (two-tailed, uncorrected); (**) not significant following Bonferroni correction

Appendix, Table 6). These ratings can help researchers to select voice stimuli for individual research purposes. Specifically, in the present study listeners rated two different sentences of each speaker for attractiveness, likeability, deviation-based distinctiveness (DEV) and "voice-in-the-crowd" distinctiveness (VITC), strength of regional accent, and age. Although raters highly agreed on mean voice ratings, *intra*rater reliability (i.e., the correspondence between the two ratings by the same rater based on the two different sentences) was only moderate to high. Although this result could simply reflect variability of intra-individual voice ratings, it may be more likely that impressions elicited by a voice depend not only on speaker characteristics that are stable across utterances, but also to a substantial extent on acoustic properties of a specific utterance or sample judged. In fact, this finding can be related to recent research on impression formation from faces, which established remarkably strong influences of image characteristics that may even exceed the influence of robust person characteristics (Jenkins et al., 2011; Todorov & Porter, 2014). The notion that impression formation depends to some extent on the voice samples used, may be related to other findings indicating that memory for individual voices also depends on sample characteristics such as the type, length and/or phonetic variability of an utterance (Cook & Wilding, 1997; Schweinberger, Herholz, & Sommer, 1997; Skuk & Schweinberger, 2013).

Irrespective of acoustic voice sample characteristics, it is also plausible that speech content per se may exert systematic effects on ratings of speakers´ voices. In the present case, we tried to control for this by using two sentences with comparable length, phonetic variability, syntactic structure, and neutral content. In addition, multilevel regression analysis indicated that mean voice ratings were largely unaffected by sentence content (with small exceptions for VITC and strength of regional accent). Nevertheless, systematic effects of speech content on ratings for voices may well be demonstrated with material that involves stronger and more systematic variations of content (for potentially relevant findings, cf. Ben-David, Multani, Shakuf, Rudzicz, & van Lieshout, 2016), and this should be taken into account when selecting stimuli from the JESS.

Researchers should keep in mind that our speakers were seated during voice recordings. Although sitting and standing may be the two most common global positions for speaking in natural interaction, our choice was governed by practical considerations related to the duration of recording sessions. Posture can affect vocal production, although more detailed aspects of posture (e.g., head position) may be even more relevant than global posture per se (e.g., Gilman & Johns, 2017).

The role of speaker age and sex for voice ratings

To exemplify the type of research questions that can be addressed with the JESS, we have explored the role of vocal age in the perception of various social signals in voices, by testing vocal age differences in mean ratings of the above six

Table 3 Old voices: Correlations (Spearman's ρ) between rating dimensions (collapsed across sentences) and between sentences (S1 and S2) for each dimension (bottom row)

Old voices ($N = 59$)	Attractiveness	Likeability	Distinctiveness (DEV)	Distinctiveness (VITC)	Accent	Perceived Age
Likeability	.84***					
Distinctiveness (DEV)	– .81***	– .76***				
Distinctiveness (VITC)	– .56***	– .54***	.67***			
Accent	– .82***	– .70***	.73***	.45***		
Perceived Age	– .73***	– .45***	.65***	.54***	.63***	
Between S1 and S2	.68***	.76***	.65***	.66***	.63***	.68***

Note that all correlations were highly significant ($^{***}p < .001$, two-tailed, uncorrected); all correlations remained significant following Bonferroni correction

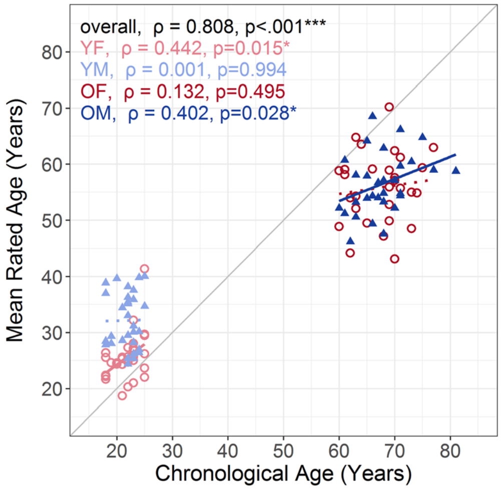

Fig. 1 Spearman rank correlations between chronological age and mean age ratings, for all speakers (overall) and separately for each speaker group (Y = young; O = old; F = female; M = male). Asterisks and solid regression lines mark significant correlations

dimensions. We also explored correlations between these dimensions separately for both age groups, and conducted exploratory acoustic analyses for the convenience of future users (cf. the supplemental material) who may find this information useful for their own (follow-up) research.

Old voices were perceived as less attractive than young voices, paralleling findings in the face domain (Ebner, 2008). In addition, old female voices were rated as least attractive, but only when rated by male listeners. In the present article, we do not wish to make strong claims regarding the potential sociobiological implications of these findings. However, we direct interested readers to other research that is broadly in line with the present results, and that argues more specifically that voices may signal mate quality and reproductive potential in humans (e.g., Apicella, Feinberg, & Marlowe, 2007; Feinberg, 2008; Hughes, Dispenza, & Gallup, 2004), that perceived age plays an important role for mate selection (Kenrick & Keefe, 1992), and that perceived vocal age and attractiveness are intertwined (Collins, 2000; Collins & Missing, 2003; Feinberg, Jones, Little, Burt, & Perrett, 2005).

With respect to the age range of speakers in the JESS, our decision for recording speakers in just two relatively homogeneous age groups has the advantage to provide substantial numbers of speakers per group. Although more recordings from continuous age ranges of adult speakers would have been desirable, this was simply beyond the scope of this project. However, users of the JESS should note that convincing impressions of middle aged voices can be elicited by morphing voices between young and old speakers (e.g., Zäske & Schweinberger, 2011). Thus, researchers have the option to use the JESS stimuli to create morphed "middle-aged voices" for their own specific research questions, where appropriate (e.g. when the focus is on creating impressions of variability in age).

In some contrast with face likeability ratings, which were reported to be generally lower for old than for young faces (Ebner, 2008), voice likeability was modulated by age group *and* by the sex of the speakers: Young female voices were rated as more likeable than old female voices, whereas no such difference was found for male voices. This pattern was independent of rater sex, and therefore differed from voice attractiveness ratings.

In terms of distinctiveness only the DEV measure yielded an effect of vocal age, with old voices rated as more distinctive than young voices. It remains possible that this could reflect different levels of perceptual exposure of the present listeners. According to an influential account of face memory (MDFS; Valentine, 1991), experience shapes our perception of which faces look typical—that is, nondistinctive. A similar notion has been put forward for voices (Latinus, McAleer, Bestelmeyer, & Belin, 2013; Papcun, Kreiman, & Davis, 1989), and it is therefore possible that young listeners are perceptually tuned to young voices, and that old voices appear distinctive to young listeners as a result of lower levels of exposure. Alternatively, voices generally might become more distinctive as speakers age. The results of a recent study, in which old voices were judged as more distinctive than young voices both by young *and* old adult listeners (Zäske, Limbach, et al., 2018) are more in line with that latter possibility.

For the domain of faces, the deviation-based measure of face distinctiveness has been used interchangeably with the "in-the-crowd" measure, assuming that norm-deviant, distinctive faces would pop out of a crowd. Because this notion has recently been challenged (Wiese et al., 2014), we also included a "voice-in-the-crowd" measure in this study. We found no effects of vocal age, voice sex, or rater sex on VITC ratings, which may be an indicator that the present groups of old and young stimuli do not contain systematic differences in features that would cause them to easily pop out of a crowd of speakers talking simultaneously. It is possible that the most salient cue in this sort of situation would be sound intensity, a cue we had rendered uninformative by normalizing the present voice samples for intensity. In Ebner's (2008) study face distinctiveness was unaffected by face age when rated by young observers. However, in Ebner's study instructions did not further specify "distinctiveness" such that those results are not easily related to the present findings in the voice domain.

Regional accent was perceived as stronger in old than in young voices, and in male than in female voices. This result may well reflect natural variation. We speculate that group differences in accent intensity could be related to different levels of previous mobility in speakers, and note that whereas the old speakers in our study were predominantly long-term local residents, young speakers were mostly students with more variable origins within Germany. Original regional accents of young speakers may well have been attenuated due to experience with other accents. The degree to which mobility could also explain the smaller sex differences in perceived

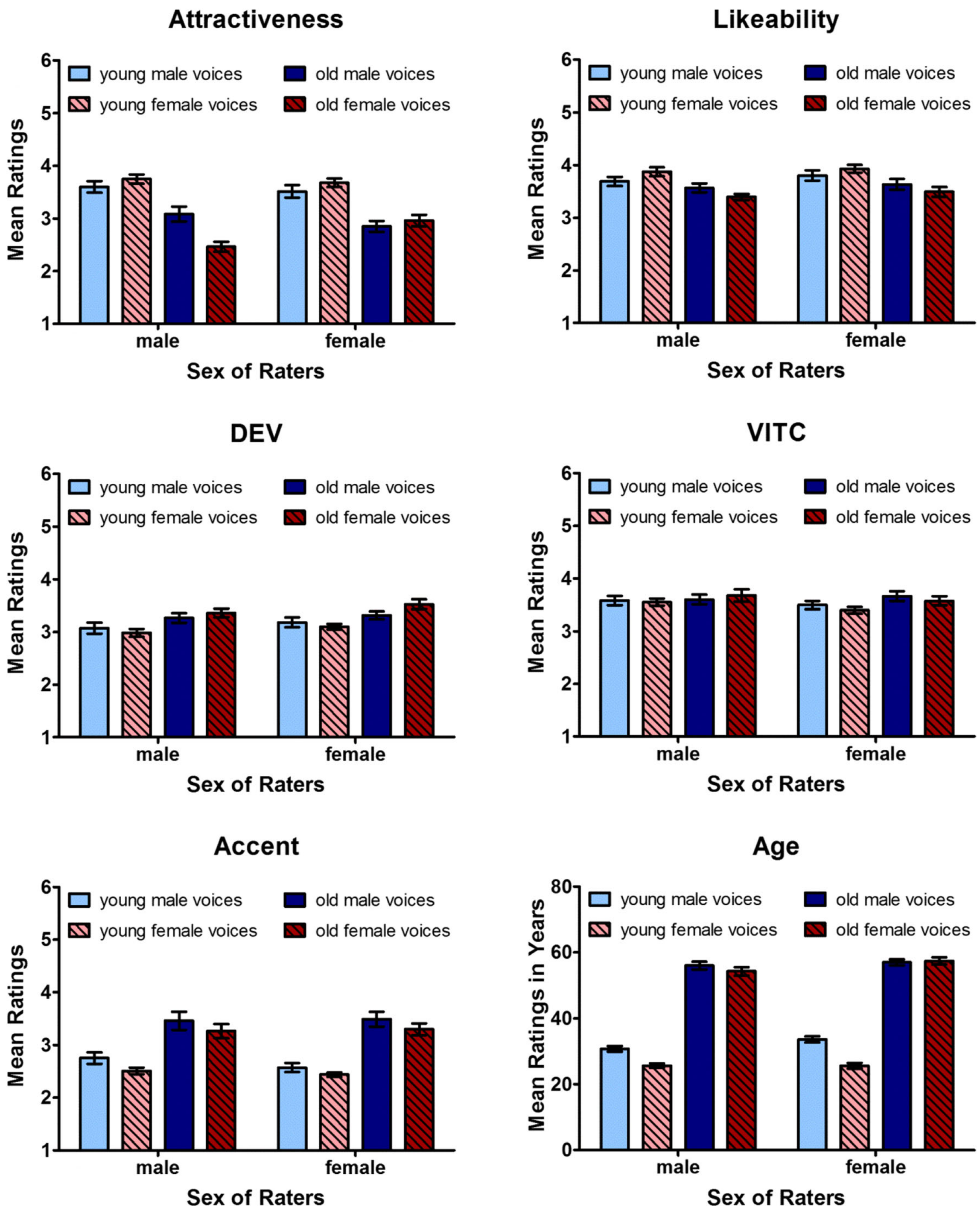

Fig. 2 Mean ratings on each rating dimension (attractiveness, likeability, deviation-based distinctiveness [DEV], "voice-in-the-crow distinctiveness [VITC], strength of regional accent, and age), depicted for young and old voices, male and female raters as well as male and female voices. Error bars are standard errors of the means (*SEM*s), based on the variance between speakers

accent is less clear. Alternatively, specific groups of speakers with stronger accent may have been less successful, able or willing to conceal their accents during recordings.

Although raters were generally able to distinguish old and young voices, precise age estimations in both groups were relatively poorer: Old voices were underestimated (by ~ 12 years) relative to their mean chronological age, whereas young voices were overestimated (by ~ 7 years). This regression to the mean is a common finding that may reflect a statistical artifact (Goy, Pichora-Fuller, & van Lieshout, 2016; Huntley et al., 1987; Shipp & Hollien, 1969) or an underdeveloped concept of age in the young (Hartman, 1979; Huntley et al., 1987). In our view, a degree of regression to the mean will likely have occurred because vocal age estimations by human listeners are generally somewhat imprecise (Linville, 1996), and because age estimations under uncertainty will tend to produce errors into the more "plausible" direction. Accordingly, the present age ratings varied substantially spanning the whole age spectrum between groups. This was despite a chronological age difference of 35 years between the oldest speaker of the young group and the youngest speaker of the old group.[3]

The analyses of mean age ratings further revealed that young male voices were perceived as older than young female voices, a finding that replicates previous results with different sets of speakers and raters (Zäske et al., 2010; Zäske et al., 2013). We tentatively propose that this difference could be because young female (as compared to male) voices are more similar in pitch to children's voices. This interpretation may be in line with the observation that perceived age in young female voices appears to be particularly strongly related to F0/pitch (cf. Supplemental Fig. S2).

Finally, age effects differed dependent on voice sex and rater sex, such that the effect of vocal age was more pronounced for female than for male voices, particularly when rated by female listeners. Within speaker groups, chronological and rated age correlated positively only for young female and old male speakers, with no significant relationship for the other two groups. Although this could lead to suggest that young female voices and old male voices convey more perceptually relevant cues to speaker age than the other groups, these findings should not be over-interpreted due to the small variability in chronological age within speaker groups. As we outlined above, vocal age perceptions and chronological age have both been linked to a number of acoustic cues (e.g., Harnsberger et al., 2008; Linville & Fisher, 1985; Shipp, Qi, Huntley, & Hollien, 1992), some of which are gender-specific (Linville, 1996; Stathopoulos et al., 2011; Torre & Barlow, 2009). Among the most prominent acoustic cues that change across the lifespan may be mean F0. We correlated F0 with both vocal age measures, to exemplify the kind of research questions that JESS stimuli allow to address. Previous research had emphasized a significant decrease of female F0 only after menopause due to hormonal changes, an effect we also observed here. Remarkably, our data additionally indicated that F0 in young adult women drops with increasing age even within this small age range (18–25 years). In the other speaker groups, F0 did not significantly correlate with chronological age. This may relate to the limited age ranges of speakers tested here, to the possibility that acoustic cues other than F0 would show more prominent age-related changes, or to a combination of these and other factors. It has been argued, for instance, that vocal changes beyond 60 years of age are often due to disease rather than mere physiological ageing (Woo, Casper, Colton, & Brewer, 1992). It is possible that the present sample of old speakers has been particularly healthy, thus providing only relatively subtle cues to physiological age. Alternatively, or in addition, sociocultural norms related to speaking styles and vocal pitch (e.g., Loveday, 1981; Pemberton, McCormack, & Russell, 1998; Starr, 2015) may have contributed to the presence or absence of correlations between chronological age and F0 in the present cross-sectional data. It may be noted that, according to recent longitudinal data, inter-individual differences in vocal pitch appear stable after puberty and throughout adulthood in men (Fouquet, Pisanski, Mathevon, & Reby, 2016) and in young pre-pubertal children (Levrero, Mathevon, Pisanski, Gustafsson, & Reby, 2018).

Although F0 and chronological age was uncorrelated within the other speaker groups, listeners did appear to use F0 as an indicator for age estimations (cf. the supplementary material, Fig. S2B). Within young voices (both female and male), perceived age increased with decreasing F0, consistent with the natural slope of F0 during early and middle adulthood (Kreiman & van Lancker Sidtis, 2011). Similarly, our finding that perceived age of old men (but not women) increased with increasing F0 mirrors reports that F0 increases during older age in men (e.g., Torre & Barlow, 2009)—even though this relationship between F0 and chronological age was not statistically significant in the present data (supplemental Fig. S2A). Overall, our findings suggest that listeners use F0 to estimate speaker age, in ways that tend to reflect chronological age effects on F0, but that perceived age also depends on a number of other parameters beyond F0 (also cf. Linville, 1996).

Relationships between perceptions on different rating dimensions

Correlations between all six rating dimensions revealed striking similarities and differences in the perception of young and old voices. For both groups positive correlations between likeability and attractiveness were similar and high ($\rho_{\text{young}} = .87$ and $\rho_{\text{old}} = .84$), but not perfect, confirming that these rating dimensions measure similar but nonidentical aspects. This interpretation is in line with differences in the pattern of mean ratings for attractiveness versus likeability (Fig. 2). Where attractiveness may refer to physical attributes, likeability may pertain more to a global affective response to a person (Zajonc, 1980). Of note, strong correlations between likeability and attractiveness were also reported for faces ($r = .76$; Ebner, 2008) suggesting that impressions of attractiveness and likeability can be similarly

[3] Note that in the present experiment we blocked stimuli by age and sex, rather than presenting them in an entirely randomized fashion. At the same time, raters were permitted to use the entire age scale from 0 to 99 in each block. Although we prevented any systematic biases of block order by counterbalancing blocks, slightly different results might have been obtained with randomized presentation. This is because age ratings for both faces (Clifford, Watson, & White, 2018; Schweinberger et al., 2010) and voices (Zäske & Schweinberger, 2011) show sequential dependencies, in the form of contrastive adaptation. Note also that blocking stimuli by age (as compared to randomized presentation) would be expected to promote contrastive adaptation, and thus might have counteracted (rather than produced) the abovementioned biases of overestimating young speakers' ages and underestimating old speakers' ages.

Table 4 Ratings as a function of voice age, voice sex, and rater sex: Results of the multilevel regression for a repeated measures analysis

	Attractiveness	Likeability	Distinctiveness (DEV)	Distinctiveness (VITC)	Accent	Age
Fixed effects						
Intercept	3.590***	3.690**	3.070***	3.590***	2.750***	30.730**
Voice age (VA)	– 0.510***	– 0.130	0.190*	0.020	0.720**	25.380***
Voice sex (VS)	0.160	0.180*	– 0.090	– 0.040	– 0.250**	– 5.120**
Rater sex (RS)	– 0.080	0.110	0.110	– 0.090	– 0.180	3.000
VA × VS	– 0.780***	– 0.350**	0.190	0.120	0.050	3.460
VA × RS	– 0.150	– 0.040	– 0.050	0.150	0.200	– 2.140
VS × RS	– 0.001	– 0.060	0.010	– 0.060	0.110	– 2.970
VA × VS × RS	0.720**	0.080	0.090	– 0.130	– 0.100	5.080*
Random effects						
Residual $s^2(\varepsilon)$	1.190***	1.061***	1.246***	1.268***	1.276***	85.845***
Intercept $s^2(\upsilon_0)$	0.085**	0.200**	0.127**	0.143**	0.333**	24.462**
VA $s^2(\upsilon_1)$	0.121*	0.140**	0.057	0.079**	0.265**	44.268**
VS $s^2(\upsilon_2)$	0.055	0.014	0.031	0.026	0.013	12.876**
VA × VS $s^2(\upsilon_3)$	0.149*	–	0.092	–	–	25.415**
$s(\upsilon_{0,1})$	0.002	– 0.114*	– 0.015	– 0.019	– 0.164*	– 25.551**
$s(\upsilon_{0,2})$	0.007	0.026	0.016	0.006	0.031	– 11.639*
$s(\upsilon_{0,3})$	– 0.017	–	– 0.021	–	–	14.277*
$s(\upsilon_{1,2})$	– 0.050	– 0.014	0.005	– 0.014	– 0.050*	11.281
$s(\upsilon_{1,3})$	0.014	–	– 0.023	–	–	– 12.732
$s(\upsilon_{2,3})$	– 0.057	–	– 0.045	–	–	– 14.250*

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; voice age (0 = young, 1 = old), voice sex (0 = male, 1 = female), rater sex (0 = male, 1 = female). We report regression coefficients for fixed effects, and estimated variance components (s^2) and covariances (s) for random effects. Missing values (–) relate to coefficients that were not estimated due to the complexity of the model

derived from both stimulus domains. Increasing attractiveness and likeability were also linked to decreasing DEV distinctiveness in both vocal age groups. This aspect of the present data is in perfect agreement with an averageness account of attractiveness that predicts a negative relationship between attractiveness and distinctiveness both for faces (Langlois & Roggman, 1990) and voices (Bruckert et al., 2010; Latinus et al., 2013). Accordingly, a given face or voice appears more attractive the more it resembles the current perceptual norm—that is, the less distinctive it is. Correlations between these ratings and acoustic distance-to-mean (DTM) measures also tend to corroborate this notion (cf. Supplemental Fig. S3).

It may also be noted that relationships between attractiveness/likeability and perceived distinctiveness were generally smaller and less systematic when considering VITC (rather than DEV) distinctiveness, and significant negative correlations were confined to old voices. As we mentioned in the introduction, in-the-crowd measures may be problematic to quantify distinctiveness. Specifically, they may be more susceptible to distortions, based on heuristics such as “Surely I would spot such an attractive voice,” thus making extremely attractive voices seem distinctive (for a similar argument for faces, cf. Wiese et al., 2014). This may potentially explain why facial attractiveness and distinctiveness were positively correlated in Ebner’s (2008) study. Not surprisingly, DEV distinctiveness was positively linked to perceived strength of regional accent in both vocal age groups, and the same applied to VITC distinctiveness when old (but not young) voices were rated. In parallel, we observed larger correlations between both distinctiveness measures in old (ρ = .67) than young (ρ = .35) voices. The reason for this pattern is not completely clear at present. We speculate that the higher prevalence of voice pathologies in older speakers (Linville, 1996) could be a contributing factor that exerts concordant effects on DEV and VITC distinctiveness ratings. This issue warrants further investigation.

Strong negative correlations were found between strength of regional accent and both attractiveness and likeability. This is in line with findings that personal impressions of speakers with a noticeable regional accent tend to be more negative than those of speakers without an accent (Rakic, Steffens, & Mummendey, 2011). Those authors argued that negative impressions may arise from the perceived unwillingness of speakers and their lack of effort to hide the accent. An additional possibility that remains to be assessed is whether such ratings could be influenced by a bias toward a listener´s own accent (Bestelmeyer, Belin, & Ladd, 2015). Note that Rakic et al’s. findings were also modulated by the type of regional accent and the type of speaker attribute tested. Although we acknowledge that type of accent may be a relevant factor, the present study was not designed to address this issue in more detail.

Although perceived age in young voices was remarkably uncorrelated with the other rating dimensions, perceived age in old voices correlated with all dimensions (cf. Tables 2 and 3 and Fig. S1). Specifically, increasing perceived age in old voices was associated with decreasing ratings of attractiveness and likeability, and with increasing ratings of distinctiveness and strength of accent. Essentially, for young voices, perceptions of attractiveness, likeability, distinctiveness and accent do not vary systematically with perceived age. By contrast, perceived age appears to matter for ratings of old voices. At present, we can only speculate why this might be the case. One possibility is that perceived voice age is relatively unimportant as long as a speaker is perceived as being in a range that would qualify this speaker as a potential mate. A second possibility involves the concept of biological as opposed to chronological age. For instance, various indicators such as telomere length as measured in white blood cells (Benetos et al., 2001), or parameters of brain age as derived from structural magnetic resonance imaging (Franke, Ziegler, Klöppel, Gaser, & the Alzheimer's Disease Neuroimaging Initiative, 2010) have been argued to represent valid estimators of biological age. To the extent that discrepancies between chronological and biological age tend to increase with increasing chronological age, and that biological age is reflected in voice characteristics, it might therefore disproportionately influence ratings of old voices. To further assess these possibilities, future research could additionally consider independent indicators of biological age, and investigate voices across a larger and more continuous range of adult ages.

Of note, correlational analyses were only interpreted at the level of age groups (as they are presented in Tables 2 and 3), as these have been corrected for multiple tests. Note that other correlations at the level of speaker age AND gender (as depicted in supplemental Fig. S1) were not adjusted, due to the exploratory nature of these tests and because we planned to focus on effects of speaker age (rather than gender). In that sense, although these data may serve to generate hypotheses for future studies, they also call for replication in a more focused design and analysis approach. It also needs to be kept in mind that the present findings were obtained with a group of young adult raters only. Although at least some rating dimensions (in particular, measures of distinctiveness) for voices appear to be largely independent of listener age (Zäske, Limbach, et al., 2018), the age of observers has been shown to modulate ratings for faces (Ebner, 2008). Because our main aim was to present and describe a novel database on voices, a consideration of any effects of rater age was beyond the scope of the present study, but may be an interesting option for future research. We note that the possible impact of age-related hearing loss on voice perception will potentially complicate such research, and will have to be carefully considered when interpreting findings.

Potential applications and future directions of research

We believe that the present data and analyses pertain to timely topics in voice research (for a recent key publication see Frühholz & Belin, 2019), such as perceptual and acoustic correlates of vocal ageing, impression formation, and the representation of voices in memory. Notably, we tried to tap into a wide range of social signals conveyed by voices, such that our results can serve as a starting point for more detailed and focused research programs. In addition to the research questions highlighted above, the JESS may be utilized to construct standardized tests in order to assess individual differences in voice cognition abilities. There is a need for good diagnostic tools measuring voice perception and memory to understand, for instance, the altered mechanisms of social cognition in individuals with autism or phonagnosia, or so-called super-recognizers (e.g., Aglieri et al., 2017; Roswandowitz, Schelinski, & von Kriegstein, 2017; Schelinski, Roswandowitz, & Von Kriegstein, 2017; Skuk, Palermo, Broemer, & Schweinberger, 2019). Although the existing tests mainly use simple vowel or syllable stimuli (Aglieri et al., 2017; Mühl et al., 2018) tests constructed with the JESS stimuli can compare voice cognition abilities across a range of utterance types, including more ecologically valid sentence stimuli and semi-spontaneous speech. Notably, the majority of the JESS stimuli—that is, sentences—were designed for both behavioral and neuroscientific research. Specifically, time-sensitive techniques such as electroencephalography (EEG) require a clear onset of speech samples (such as the [d:] plosive in the German articles "der," "die" or "das") to obtain high quality EEG data. Moreover, many stimuli are optimized for the creation of voice morphs (e.g., syllables), which are increasingly used to systematically study the acoustic properties underlying the perception of various social signals in voices (reviewed in Kawahara & Skuk, 2019). The JESS may also serve other speech-related disciplines, such as speech sciences, medicine, or forensics. For instance, by means of the information provided here, it would be possible to study acoustic properties of regional accents or age and their interplay with speaker memory, a topic relevant for research on the reliability of "earwitness" testimony. Possible clinical applications could include using the JESS to study altered speaker and speech perception in individuals with autism spectrum disorder, schizophrenia, dementia or stroke. Another promising application may entail the evaluation of nonverbal voice perception in hearing-impaired patients, to optimize the technology of hearing aids, such as cochlear implants, for a wider spectrum of vocal signals including speech content, age, gender, or identity.

Conclusion

Taken together, we present and describe the Jena Speaker Set (JESS), a database of 120 adult voices of young and old female and male speakers, which can be used to study acoustic, perceptual and neural correlates of various vocal signals such as age, identity, gender, attractiveness, distinctiveness, or perceived personality traits. The present research includes extensive ratings of individual voices, and specifies the role of vocal age as an important factor that affects voice perception. Accordingly, age is a factor that needs to be considered when designing voice studies. We expect that future research will show that impressions based on voices play a similarly important role for person memory as has been established for their facial counterparts (e.g., Bruce & McDonald, 1993; Meissner & Brigham, 2001; Schulz et al., 2012; Wiese et al., 2014).

Acknowledgments This work was funded by the Deutsche Forschungsgemeinschaft (Grant ZA 745/1-1; 1-2; Schw 511/10-2; FOR1097) and by the Programm zur Förderung der Drittmittelfähigkeit (2012) of the University of Jena. We thank all of the speakers and participants, as well as our student research assistants (Leonie Fresz, Achim Hötzel, Christoph Klebl, Katrin Lehmann, Carolin Leistner, Constanze Mühl, Denise Humble, Finn Pauls, Marie-Christin Perlich, Johannes Pfund, Mathias Riedel, Saskia Rudat, Meike Wilken, and Lukas Erchinger), for stimulus acquisition and editing. The entire stimulus set of the JESS is available from the following link: https://osf.io/m5zdf/. This study has not been preregistered.

Open Practice Statement The data of the experiment, including acoustic analyses of the stimuli, are available as (electronic) supplemental materials to this article.

Appendix

Table 5 Instructions and response alternatives for ratings of each dimension

Dimension	Instruction	Response alternatives
Attractiveness[a]	Please assess how unattractive/attractive the voices are. You will listen to female voices now. If you are a woman, please assess how attractive the respective voice may sound to heterosexual men. If you are a man, please assess how attractive you personally find the respective voice. (Or: You will listen to male voices now. If you are a man, please assess how attractive the respective voice may sound to heterosexual women. If you are a woman, please assess how attractive you personally find the respective voice.)	1 very unattractive 2 unattractive 3 a little unattractive 4 a little attractive 5 attractive 6 very attractive
Likeability	Please assess how unlikeable/likeable the voices are.	1 very unlikeable 2 unlikeable 3 a little unlikeable 4 a little likeable 5 likeable 6 very likeable
Distinctiveness[b] (DEV)	Please assess how untypical/typical the voices are. Ask yourself how the respective voice differs from other voices that you know.	1 very untypical 2 untypical 3 a little untypical 4 a little typical 5 typical 6 very typical
Distinctiveness (VITC)	Please assess how undistinctive/distinctive the voices are. Imagine yourself on a busy square. You are surrounded by many people who are talking simultaneously. A voice is distinctive if it stands out of the crowd.	1 very undistinctive 2 undistinctive 3 a little undistinctive 4 a little distinctive 5 distinctive 6 very distinctive
Intensity of regional accent	Please assess how weak/strong the regional accent of the following speakers is.	1 very weak 2 weak 3 a little weak 4 a little strong 5 strong 6 very strong
Age	Please assess the age of the voices. Use the keys 0 to 9 to enter your response. Please enter a two-digit response. There are no right or wrong answers. Just assess how old you think the speakers are.	10–99

[a] We decided to specify the attractiveness task such that we asked raters judging same-sex voices to take the perspective of a heterosexual opposite-sex rater. This was in order to emphasize the sexual connotation of "attractiveness" and to separate it more clearly from the related concept of "likeability," which we assessed in a separate task. This was done in order to prevent potential confounds between these different connotations of attractiveness, which have complicated the interpretation of previous findings (Babel et al., 2014). [b] For convenience of interpretation, ratings of DEV distinctiveness were recoded such that low and high values reflect nondistinctive (typical) and distinctive (untypical) ratings, respectively. All data reported here are based on recoded DEV ratings

Table 6 Descriptive information about ratings of young and old voices in the total sample and in subsamples of young and old, female and male raters: size of speaker set (*N*), mean rating (*M*) and standard deviation (*SD*) between ratings of different speakers

		Attractiveness		Likeability		Distinctiveness (DEV)		Distinctiveness (VITC)		Accent		Perceived Age	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
All raters (*N* = 24)													
All voices	120	3.25	0.69	3.68	0.47	3.22	0.46	3.57	0.43	2.97	0.73	42.31	14.80
Young voices	61	3.63	0.54	3.82	0.46	3.08	0.44	3.51	0.36	2.57	0.43	28.93	5.61
Female voices	30	3.72	0.44	3.90	0.41	3.47	0.33	3.04	0.32	2.47	0.27	25.59	4.13
Male voices	31	3.56	0.62	3.75	0.50	3.54	0.39	3.13	0.54	2.66	0.53	32.16	4.94
Old voices	59	2.85	0.60	3.52	0.44	3.37	0.44	3.63	0.49	3.38	0.75	56.15	5.83
Female voices	29	2.72	0.54	3.44	0.38	3.63	0.51	3.44	0.46	3.28	0.65	55.83	6.30
Male voices	30	2.97	0.64	3.60	0.48	3.63	0.47	3.29	0.43	3.48	0.84	56.45	5.42
Female raters (*N* = 12)													
All voices	120	3.26	0.67	3.72	0.52	3.28	0.46	3.53	0.45	2.94	0.71	43.18	15.03
Young voices	61	3.60	0.57	3.87	0.49	3.14	0.42	3.45	0.39	2.50	0.39	29.65	6.39
Female voices	30	3.68	0.44	3.93	0.40	3.40	0.35	3.10	0.29	2.44	0.24	25.57	4.66
Male voices	31	3.51	0.67	3.80	0.56	3.50	0.43	3.18	0.52	2.57	0.49	33.59	5.29
Old voices	59	2.91	0.58	3.57	0.52	3.42	0.46	3.62	0.48	3.40	0.69	57.16	5.45
Female voices	29	2.96	0.59	3.49	0.49	3.58	0.46	3.52	0.50	3.30	0.60	57.36	5.98
Male voices	30	2.85	0.57	3.64	0.54	3.66	0.50	3.32	0.41	3.49	0.76	56.98	4.98
Male raters (*N* = 12)													
All voices	120	3.24	0.77	3.63	0.47	3.16	0.51	3.60	0.49	2.99	0.78	41.42	14.68
Young voices	61	3.67	0.55	3.78	0.49	3.03	0.51	3.57	0.42	2.63	0.51	28.19	4.95
Female voices	30	3.75	0.49	3.87	0.47	3.55	0.37	2.98	0.40	2.51	0.34	25.60	3.73
Male voices	31	3.60	0.61	3.69	0.49	3.58	0.48	3.07	0.59	2.75	0.61	30.70	4.73
Old voices	59	2.78	0.71	3.48	0.40	3.31	0.48	3.64	0.56	3.36	0.85	55.09	6.54
Female voices	29	2.47	0.50	3.39	0.31	3.68	0.62	3.36	0.44	3.26	0.73	54.20	6.80
Male voices	30	3.09	0.75	3.56	0.46	3.60	0.50	3.27	0.51	3.46	0.95	55.95	6.28

References

Aglieri, V., Watson, R., Pernet, C., Latinus, M., Garrido, L., & Belin, P. (2017). The Glasgow Voice Memory Test: Assessing the ability to memorize and recognize unfamiliar voices. *Behavior Research Methods*, *49*, 97–110. https://doi.org/10.3758/s13428-015-0689-6

Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, *49*, 758–771. https://doi.org/10.3758/s13428-016-0736-y

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters*, *3*, 682–684.

Asendorpf, J. B., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*, 1531–1544. https://doi.org/10.1037/0022-3514.74.6.1531

Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLoS ONE*, *9*, e88616. https://doi.org/10.1371/journal.pone.0088616

Baken, R. J. (2005). The aged voice: A new hypothesis (Reprinted from *Voice*, Vol. 3, pp. 57–73, 1994). *Journal of Voice*, *19*, 317–325. https://doi.org/10.1016/j.jvoice.2004.07.005

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. https://doi.org/10.1023/a:1005653411471

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research*, *74*, 110–120.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*, 531–539. https://doi.org/10.3758/brm.40.2.531

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech Language and Hearing Research, 59*, 72–89. https://doi.org/10.1044/2015_jslhr-h-14-0323

Benetos, A., Okuda, K., Lajemi, M., Kimura, M., Thomas, F., Skurnick, J., . . . Aviv, A. (2001). Telomere length as an indicator of biological aging—The gender effect and relation with pulse pressure and pulse wave velocity. *Hypertension, 37,* 381–385.

Bestelmeyer, P. E. G., Belin, P., & Ladd, D. R. (2015). A neural marker for social bias toward in-group accents. *Cerebral Cortex*, *25*, 3953–3961. https://doi.org/10.1093/cercor/bhu282

Bestelmeyer, P. E. G., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., & Belin, P. (2012). Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. *Cerebral Cortex*, *22*, 1263–1270. https://doi.org/10.1093/cercor/bhr204

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und McCrae (NEO-FFI): Manual* (2nd ed.). Göttingen, Germany: Hogrefe.

Brown, W. S., Morris, R. J., & Michel, J. F. (1989). Vocal jitter in young adult and aged female voices. *Journal of Voice*, *3*, 113–119.

Bruce, A. J., & McDonald, B. G. (1993). Face recognition as a function of judgments of likability or unlikability. *Journal of General Psychology*, *120*, 451–462.

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., . . . Belin, P. (2010). Vocal Attractiveness Increases by Averaging. *Current Biology, 20,* 116–120.

Bruckert, L., Lienard, J. S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society B*, *273*, 83–89.

Clifford, C. W. G., Watson, T. L., & White, D. (2018). Two sources of bias explain errors in facial age estimation. *Royal Society Open Science*, *5*, 180841. https://doi.org/10.1098/rsos.180841

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*, 773–780.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour*, *65*, 997–1004.

Cook, S., & Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology*, *11*, 95–111.

Ebner, N. C. (2008). Age of face matters: Age-group differences in ratings of young and old faces. *Behavior Research Methods*, *40*, 130–136. https://doi.org/10.3758/brm.40.1.130

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/brm.41.4.1149

Feinberg, D. R. (2008). Are human faces and voices ornaments signaling common underlying cues to mate value? *Evolutionary Anthropology*, *17*, 112–118.

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, *37*, 615–623.

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*, 561–568.

Ferdenzi, C., Delplanque, S., Mehu-Blantar, I., Cabral, K. M. D., Felicio, M. D., & Sander, D. (2015). The Geneva Faces and Voices (GEFAV) database. *Behavior Research Methods*, *47*, 1110–1121. https://doi.org/10.3758/s13428-014-0545-0

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, *322*, 970–973.

Fouquet, M., Pisanski, K., Mathevon, N., & Reby, D. (2016). Seven and up: Individual differences in male voice fundamental frequency emerge before puberty and remain stable throughout adulthood. *Royal Society Open Science*, *3*, 160395. https://doi.org/10.1098/rsos.160395

Franke, K., Ziegler, G., Klöppel, S., Gaser, C., & the Alzheimer's Disease Neuroimaging Initiative. (2010). Estimating the age of healthy subjects from T_1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage*, *50*, 883–892. https://doi.org/10.1016/j.neuroimage.2010.01.005

Freitag, C. M., Retz-Junginger, P., Retz, P., Seitz, C., Palmason, H., Meyer, J., . . . von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ)—die Kurzversion AQ-k. *Zeitschrift für Klinische Psychologie und Psychotherapie, 36,* 280–289.

Frühholz, S., & Belin, P. (2019). *The Oxford handbook of voice perception* (1st ed.). Oxford, UK: Oxford University Press.

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., . . . Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia, 47,* 123–131.

Gilman, M., & Johns, M. M. (2017). The effect of head position and/or stance on the self-perception of phonatory effort. *Journal of Voice*, *31*, 24. https://doi.org/10.1016/j.jvoice.2015.11.024

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*, 448–458. https://doi.org/10.3758/BF03199567

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia, PA: Lea & Febiger.

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders*, *39*, 171–184.

Goy, H., Pichora-Fuller, M. K., & van Lieshout, P. (2016). Effects of age on speech and voice quality ratings. *Journal of the Acoustical Society of America*, *139*, 1648–1659. https://doi.org/10.1121/1.4945094

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*, 523–530.

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*, 58–69.

Hartman, D. E. (1979). Perceptual identity and characteristics of aging in normal male adult speakers. *Journal of Communication Disorders*, *12*, 53–61.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*, *39*, 101–117. https://doi.org/10.3758/bf03192848

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Hughes, S. M., Dispenza, F., & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, *25*, 295–304.

Huntley, R., Hollien, H., & Shipp, T. (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice*, *1*, 49–52.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition 121*(3), 313-323. https://doi.org/10.1016/j.cognition.2011.08.001

Kawahara, H., & Skuk, V. G. (2019). Voice morphing. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception* (pp. 685–706). Oxford, UK: Oxford University Press.

Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex-differences in reproductive strategies. *Behavioral and Brain Sciences*, *15*, 75–91.

Kreiman, J., & van Lancker Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception* (1st ed.). Chichester, UK: Wiley-Blackwell.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1*, 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x

Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, *2*, 175:1–12. https://doi.org/10.3389/fpsyg.2011.00175

Latinus, M., Crabbe, F., & Belin, P. (2011). Learning-induced changes in the cerebral processing of voice identity. *Cerebral Cortex*, *21*, 2820–2828. https://doi.org/10.1093/cercor/bhr077

Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, *23*, 1075–1080.

Latinus, M., & Zäske, R. (2019). Perceptual correlates and cerebral representation of voices—Identity, gender, and age. In S. Frühholz & P.

Belin (Eds.), *The Oxford handbook of voice perception* (pp. 561–584). Oxford, UK: Oxford University Press.

Levrero, F., Mathevon, N., Pisanski, K., Gustafsson, E., & Reby, D. (2018). The pitch of babies' cries predicts their voice pitch at age 5. *Biology Letters*, *14*(7). https://doi.org/10.1098/rsbl.2018.0065

Linville, S. E. (1996). The sound of senescence. *Journal of Voice*, *10*, 190–200.

Linville, S. E., & Fisher, H. B. (1985). Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *Journal of the Acoustical Society of America*, *78*, 40–48.

Loveday, L. (1981). Pitch, politeness and sexual role—An exploratory investigation into the pitch correlates of English and Japanese politeness formulas. *Language and Speech*, *24*, 71–89. https://doi.org/10.1177/002383098102400105

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS ONE*, *9*, e90779. https://doi.org/10.1371/journal.pone.0090779

McCrae, R. R., & Costa, P. T. (1987). Validation of the 5-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90. https://doi.org/10.1037/0022-3514.52.1.81

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces—A meta-analytic review. *Psychology Public Policy and Law*, *7*, 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*, 630–633. https://doi.org/10.3758/bf03206543

Moyse, E. (2014). Age estimation from faces and voices: A review. *Psychologica Belgica*, *54*, 255–265. https://doi.org/10.5334/pb.aq

Mühl, C., Sheil, O., Jarutyte, L., & Bestelmeyer, P. E. G. (2018). The Bangor Voice Matching Test: A standardized test for the assessment of voice perception ability. *Behavior Research Methods*, *50*, 2184–2192. https://doi.org/10.3758/s13428-017-0985-4

Neyer, F. J., & Asendorpf, J. B. (2001). Personality-relationship transaction in young adulthood. *Journal of Personality and Social Psychology*, *81*, 1190–1204. https://doi.org/10.1037/0022-3514.81.6.1190

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term-memory for unfamiliar voices. *Journal of the Acoustical Society of America*, *85*, 913–925.

Parsons, C. E., Young, K. S., Craske, M. G., Stein, A. L., & Kringelbach, M. L. (2014). Introducing the Oxford Vocal (OxVoc) Sounds database: A validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Frontiers in Psychology*, *5*, 562. https://doi.org/10.3389/fpsyg.2014.00562

Pemberton, C., McCormack, P., & Russell, A. (1998). Have women's voices lowered across time? A cross sectional study of Australian women's voices. *Journal of Voice*, *12*, 208–213. https://doi.org/10.1016/s0892-1997(98)80040-4

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. (2011). Human voice recognition depends on language ability. *Science*, *333*, 595–595. https://doi.org/10.1126/science.1207327

Perrachione, T. K., & Wong, P. C. M. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. *Neuropsychologia*, *45*, 1899–1910.

Raine, J., Pisanski, K., & Reby, D. (2017). Tennis grunts communicate acoustic cues to sex and contest outcome. *Animal Behaviour*, *130*, 47–55. https://doi.org/10.1016/j.anbehav.2017.06.022

Rakic, T., Steffens, M. C., & Mummendey, A. (2011). When it matters how you pronounce it: The influence of regional accents on job interview outcome. *British Journal of Psychology*, *102*, 868–883. https://doi.org/10.1111/j.2044-8295.2011.02051.x

Ramig, L. O., Gray, S., Baker, K., Corbin-Lewis, K., Buder, E., Luschei, E., . . . Smith, M. (2001). The aging voice: A review, treatment data and familial and genetic perspectives. *Folia Phoniatrica et Logopaedica, 53,* 252–265. https://doi.org/10.1159/000052680

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: 10 Item Big Five Inventory (BFI-10). *Methoden, Daten, Analysen*, *7*, 233–249.

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F-0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, *52*, 638–651. https://doi.org/10.1016/j.specom.2010.02.012

Roswandowitz, C., Schelinski, S., & von Kriegstein, K. (2017). Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. *NeuroImage*, *155*, 97–112. https://doi.org/10.1016/j.neuroimage.2017.02.064

Ryan, W. J., & Burk, K. W. (1974). Perceptual and acoustic correlates of aging in speech of males. *Journal of Communication Disorders*, *7*, 181–192.

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology*, *25*, 3079–3085. https://doi.org/10.1016/j.cub.2015.10.009

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*, 110–114. https://doi.org/10.2307/3002019

Schelinski, S., Roswandowitz, C., & Von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research*, *10*, 155–168.

Schulz, C., Kaufmann, J. M., Kurt, A., & Schweinberger, S. R. (2012). Faces forming traces: Neurophysiological correlates of learning naturally distinctive and caricatured faces. *NeuroImage*, *63*, 491–500.

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., . . . Zäske, R. (2008). Auditory adaptation in voice perception. *Current Biology, 18,* 684–688. https://doi.org/10.1016/j.cub.2008.04.015

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech Language and Hearing Research*, *40*, 453–463.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 15–25.

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

Shipp, T., & Hollien, H. (1969). Perception of aging male voice. *Journal of Speech and Hearing Research*, *12*, 703–710.

Shipp, T., Qi, Y. Y., Huntley, R., & Hollien, H. (1992). Acoustic and temporal correlates of perceived age. *Journal of Voice*, *6*, 211–216.

Skuk, V. G., Dammann, L. M., & Schweinberger, S. R. (2015). Role of timbre and fundamental frequency in voice gender adaptation. *Journal of the Acoustical Society of America*, *138*, 1180–1193. https://doi.org/10.1121/1.4927696

Skuk, V. G., Palermo, R., Broemer, L., & Schweinberger, S. R. (2019). Autistic traits are linked to individual differences in familiar voice identification. *Journal of Autism and Developmental Disorders*, *49*, 2747–2767. https://doi.org/10.1007/s10803-017-3039-y

Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research*, *295*, 131–140.

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech Language and Hearing Research*, *57*, 285–296. https://doi.org/10.1044/1092-4388

Starr, R. (2015). Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society*, *44*, 1–34. https://doi.org/10.1017/S0047404514000724.

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*, 1011–1021. https://doi.org/10.1044/1092-4388

Todorov, A., & Porter, J. M. (2014). Misleading First Impressions: Different for Different Facial Images of the Same Person. *Psychological Science, 25*(7), 1404-1417. https://doi.org/10.1177/0956797614532474

Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, *42*, 324–333.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology*, *43*, 161–204.

von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*, 948–955.

Wallis, J., Lipp, O. V., & Vanman, E. J. (2012). Face age and sex modulate the other-race effect in face recognition. *Attention, Perception, & Psychophysics*, *74*, 1712–1721. https://doi.org/10.3758/s13414-012-0359-z

Wickham, L. H. V., & Morris, P. E. (2003). Attractiveness, distinctiveness, and recognition of faces: Attractive faces can be typical or distinctive but are not better recognized. *American Journal of Psychology*, *116*, 455–468. https://doi.org/10.2307/1423503

Wiese, H. (2012). The role of age and ethnic group in face recognition memory: ERP evidence from a combined own-age and own-race bias study. *Biological Psychology*, *89*, 137–147.

Wiese, H., Altmann, C. S., & Schweinberger, S. R. (2014). Effects of attractiveness on face memory separated from distinctiveness: Evidence from event-related brain potentials. *Neuropsychologia*, *56*, 26–36. https://doi.org/10.1016/j.neuropsychologia.2013.12.023

Woo, P., Casper, J., Colton, R., & Brewer, D. (1992). Dysphonia in the aging—Physiology versus disease. *Laryngoscope*, *102*, 139–144.

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17*, 263–271.

Zajonc, R. B. (1980). Feeling and thinking—Preferences need no inferences. *American Psychologist*, *35*, 151–175. https://doi.org/10.1037/0003-066x.35.2.151

Zarate, J. M., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*, *5*, 11475. https://doi.org/10.1038/srep11475

Zäske, R., Limbach, K., Schneider, D., Skuk, V. G., Dobel, C., Guntinas-Lichius, O., & Schweinberger, S. R. (2018a). Electrophysiological correlates of voice memory for young and old speakers in young and old listeners. *Neuropsychologia*, *116*, 215–227. https://doi.org/10.1016/j.neuropsychologia.2017.08.011

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

Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, *268*, 38–45.

Zäske, R., Schweinberger, S. R., & Skuk, V. G. (2018b). *Attractiveness and distinctiveness in voices and faces of young adults*. PsyArXiv preprint. Retrieved from https://psyarxiv.com/2avu3/

Zäske, R., Skuk, V. G., Kaufmann, J. M., & Schweinberger, S. R. (2013). Perceiving vocal age and gender: An adaptation approach. *Acta Psychologica*, *144*, 583–593.

Zäske, R., Volberg, G., Kovacs, G., & Schweinberger, S. R. (2014). Electrophysiological correlates of voice learning and recognition. *Journal of Neuroscience*, *34*, 10821–10831. https://doi.org/10.1523/jneurosci.0581-14.2014

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.