JASA ARTICLE

Talker change detection by listeners varying in age and hearing loss

Virginia Best,[1,a)] Jayne B. Ahlstrom,[2] Christine R. Mason,[1] Tyler K. Perrachione,[1] Gerald Kidd, Jr.,[1,2] and Judy R. Dubno[2]

[1]*Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts 02215, USA*

[2]*Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, Charleston, South Carolina 29425, USA*

ABSTRACT:

Despite a vast literature on how speech intelligibility is affected by hearing loss and advanced age, remarkably little is known about the perception of talker-related information in these populations. Here, we assessed the ability of listeners to detect whether a change in talker occurred while listening to and identifying sentence-length sequences of words. Participants were recruited in four groups that differed in their age (younger/older) and hearing status (normal/impaired). The task was conducted in quiet or in a background of same-sex two-talker speech babble. We found that age and hearing loss had detrimental effects on talker change detection, in addition to their expected effects on word recognition. We also found subtle differences in the effects of age and hearing loss for trials in which the talker changed vs trials in which the talker did not change. These findings suggest that part of the difficulty encountered by older listeners, and by listeners with hearing loss, when communicating in group situations, may be due to a reduced ability to identify and discriminate between the participants in the conversation.

 https://doi.org/10.1121/10.0025539

(Received 21 November 2023; revised 6 March 2024; accepted 19 March 2024; published online 8 April 2024)

[Editor: Pamela Souza] Pages: 2482–2491

I. INTRODUCTION

Speech contains a rich array of acoustic cues that allow a listener to identify who is talking (Kreiman *et al.*, 2005; Mathias and von Kriegstein, 2014; Scott and McGettigan, 2016). These cues include features related to the anatomy and physiology of the vocal source and vocal tract (Schweinberger and Zaske, 2018) and also talkers' idiosyncratic patterns of articulation, which convey both individual and sociocultural identity (Perrachione *et al.*, 2010; Perrachione *et al.*, 2019). Because the acoustic characteristics of a given talker are so numerous, and so highly variable during natural speech dynamics, what listeners perceive as a talker's voice is ultimately an auditory gestalt that cannot be easily decomposed into a linear combination of acoustic primitives (Kreiman and Sidtis, 2011). Here, and throughout, we use the term "voice" to refer to a listeners' holistic auditory perception of a talker's identity. Recent work has suggested that voice recognition involves two complementary processes. First, listeners must be able to distinguish between different voices, which requires sensitivity to acoustic *differences* between one voice and another. Second, listeners must also be able to maintain a consistent perception of a given individual's voice across different utterances, which requires sensitivity to acoustic *similarities* in the face of substantial situational variability. These two processes are sometimes termed "telling people apart" and "telling people together," respectively (Lavan *et al.*, 2019). The acoustic cues supporting these two processes are likely to be different and highly dependent on the voice or set of voices in question (Lee *et al.*, 2019; Lee and Kreiman, 2022). Moreover, the role of learning may be quite different for the two processes. Telling people apart is possible based on very brief samples of each voice, such as a vowel (Baumann and Belin, 2010). Telling people together, on the other hand, would seem to require a more extended exposure as listeners must learn which features are reliably associated with a voice and acquire knowledge about the variance of the distributions of these features (Kanber *et al.*, 2022).

While the detrimental effects of advanced age and hearing loss on word recognition and speech comprehension are well known and have been extensively characterized (Humes and Dubno, 2010), surprisingly little is known about how these factors affect access to and use of talker-related information in speech. Best *et al.* (2018) examined the ability to identify talkers based on spoken sentences, in 32 listeners varying in age and hearing loss. Both age and hearing loss affected talker identification in quiet, with hearing loss further impairing talker identification in background noise. However, the identification task could not distinguish difficulties with voice recognition *per se* from difficulties with learning and remembering the previously unfamiliar voices. The authors speculated that hearing loss may disrupt the fine acoustic distinctions required to discriminate between similar-sounding voices, while age-related cognitive declines may disrupt the ability to learn and remember voices. Subsequent work using simpler tasks, with minimal memory requirements, have not fully clarified the unique

[a)]Email: ginbest@bu.edu

effects of aging vs hearing loss on processing voice information. Xu *et al.* (2021) examined talker discrimination for word pairs and found that older adults performed more poorly than younger adults. While the older group contained listeners with and without hearing loss, all listeners in this group performed extremely poorly on this task, obfuscating any unique effect of hearing loss. Zaltz and Kishon-Rabin (2022) tested older and younger listeners on a voice discrimination task in which the available cues were limited to $F0$, formant cues, or both. They also found poorer discrimination by older listeners, but could not rule out a contribution from their poorer hearing thresholds.

In the current study, we again approached the question of how age and hearing loss affect sensitivity to talker-related information in speech. We used a talker change detection task (inspired by Sharma *et al.*, 2019; Sharma *et al.*, 2020), in which listeners heard a sequence of words and listened for a change in talker during the sequence. Like the tasks of Xu *et al.* (2021) and Zaltz and Kishon-Rabin (2022), talker change detection does not require explicit learning of voices but rather a comparison of voices presented sequentially. In a modification of the task, we also asked listeners to identify the words in the sequence. This dual-task structure was adopted for two reasons. First, it reflects everyday communication, in which a listener often must process who is talking in addition to what is being said. Second, it is convenient in that effects of age and hearing loss on talker change detection and on word recognition can be measured simultaneously for the same stimuli. To provide a more complete picture of these abilities, we included both an ideal condition with no background noise and a more challenging condition with speech babble in the background. Finally, to enable us to tease apart effects of age and hearing loss, we used a four-group design (see Sec. II A) in which these two variables were not tightly coupled.

II. METHODS

A. Participants

We recruited 35 participants in total, all of whom considered English to be their primary language. As in the study by Best *et al.* (2018), we recruited participants into four groups that differed in age and/or hearing status. The groups will be referred to as younger with normal hearing (YNH), younger with hearing impairment (YHI), older with normal hearing (ONH), and older with hearing impairment (OHI). The younger participants were recruited and tested at Boston University (BU) and the older participants were recruited and tested at the Medical University of South Carolina (MUSC). Criteria for normal audiometric thresholds in the younger group were 20 dB hearing level (HL) or better at octave frequencies from 0.25–8 kHz in both ears. For the older individuals, normal hearing was defined as thresholds of 30 dB HL or better bilaterally from 0.25–4 kHz. The participants with hearing impairment had bilateral, symmetric, sensorineural losses. Symmetry was defined as a between-ear difference in the low-frequency pure-tone average (0.25–1 kHz) of no more than 10 dB and a between-ear difference in the high-frequency pure-tone average (2–8 kHz) of no more than 15 dB. Figure 1 shows individual audiograms for listeners in each group, and summary characteristics are presented in Table I. Note that we chose to use the all-frequency average hearing loss (0.25–8 kHz) to quantify hearing loss because losses in the YHI group were quite heterogeneous in their configuration, and no one frequency region captured the variation in severity across our population. Because of our four-group design, age and the all-frequency hearing loss were only weakly correlated in this population ($r = 0.34$; $p = 0.02$). The correlation was not significant when using the more common four-frequency average hearing loss (0.5–4 kHz; $r = 0.18$; $p = 0.15$).

B. Equipment

The BU and MUSC laboratories ran the same custom MATLAB code (MathWorks, Natick, MA) to control signal calibration, presentation of stimuli, instructions, and feedback, and collection of responses. At MUSC, the output of a LynxTWO sound card (Lynx Studio Technology, Inc., Costa Mesa, CA) was passed through a Tucker-Davis Technologies (Alachua, FL) anti-aliasing filter (FT5), signal mixer (SM3A), and headphone driver (HB7) before being delivered to the Sennheiser 280 Pro headphones (Wedemark, Germany). At BU, the output of an RME Digiface sound card (Haimhausen, Germany) was delivered to the same model headphones. At both sites, the experiment was conducted in a sound-attenuating booth. At BU, a monitor displayed instructions and the response interface, and responses were given using a standard computer mouse. At MUSC, listeners entered their responses using a touch screen.

C. Speech materials

The speech materials were taken from a multi-talker corpus that has been described previously and used for many experiments in the BU laboratory (Kidd *et al.*, 2008). The matrix-style corpus contains eight word choices in each of five syntactic categories (name, verb, number, adjective, object) for a total of 40 words. The entire corpus was recorded by multiple talkers with American-accented English, and for the current study, we used ten male and ten female talkers. The average $F0$, estimated for each talker using STRAIGHT (Kawahara *et al.*, 2005), ranged from 81–138 Hz for the male talkers and from 183–222 Hz for the female talkers.

Perceptual dissimilarity ratings of these talkers were obtained from eight additional YNH listeners who did not participate in the main experiment. These listeners were randomly assigned to rate either the male or female talker pairs. On each trial, the listener heard two talkers each say three random words from the corpus and were asked to rate their perceived dissimilarity on a continuous scale from 0 to 1 (with 0 being the most similar and 1 being the most dissimilar; Perrachione *et al.*, 2019). These listeners rated all possible pairings of the ten talkers (including same-voice pairs).

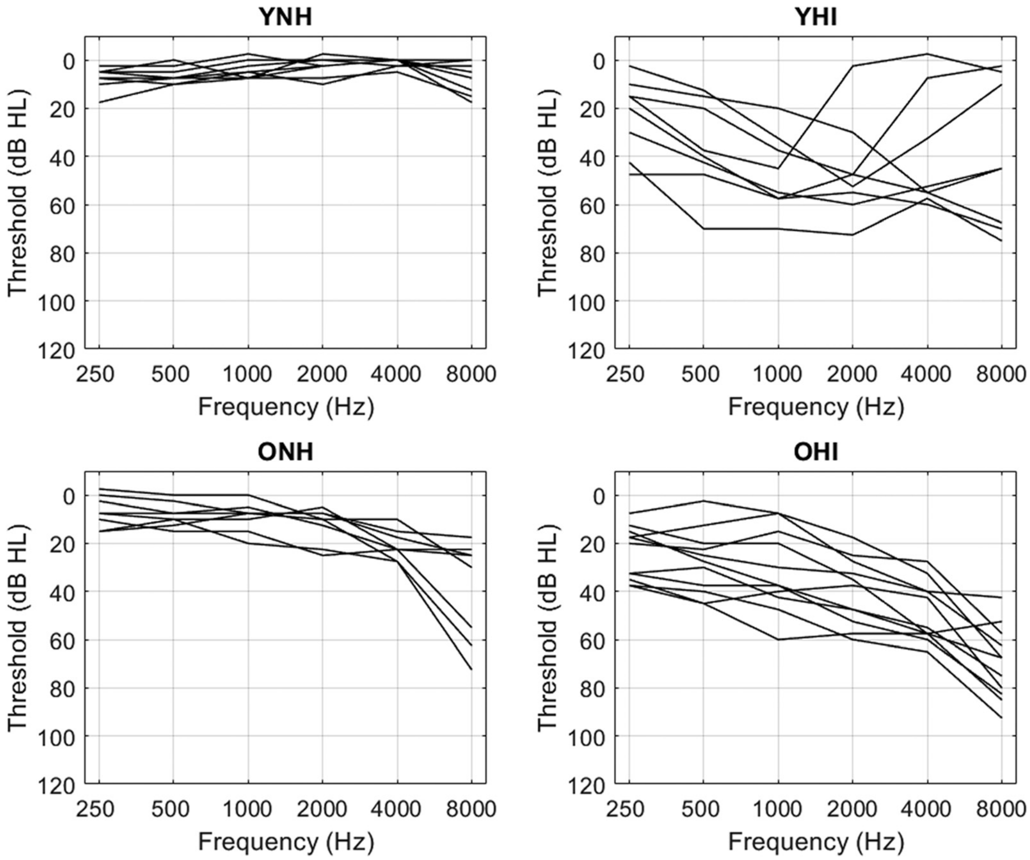

FIG. 1. Audiograms (averaged across left and right ears) for each listener in each of the four groups.

Dissimilarity ratings were highly consistent across listeners: Intraclass correlation coefficients were 0.91 for the female talkers and 0.83 for the male talkers. The mean dissimilarity rating for each pair of talkers was compared to listeners' talker change detection performance for that pair in the main experiment.

D. Stimuli and task

On every trial, the target was a five-word sequence that was constructed by choosing one word at random from each of the syntactic categories in order. On a given trial, the five target words were all spoken either by the same talker, or by two different talkers of the same sex. When there was a change in talker, it happened unpredictably at any word boundary in the sequence, with the constraint that each word boundary was sampled equally within a block. Figures 2(A) and 2(B) show examples of one-talker and two-talker trials, respectively. The target was presented either in quiet, or in the presence of same-sex two-talker babble.

TABLE I. Summary characteristics for the four participant groups.

Group	YNH	YHI	ONH	OHI
N	8	8	8	11
Age in years (range, mean ± standard deviation)	18–27 22 ± 3.2	18–32 24.5 ± 4.7	64–84 71.9 ± 6.1	62–88 73.6 ± 7.4
All-frequency average hearing loss in dB HL (mean ± standard deviation)	5.0 ± 2.3	39.5 ± 16.8	17.8 ± 7.4	42.6 ± 12.1

The babble was created by level-equalizing and concatenating (in random order) the 339 Bamford-Kowal-Bench sentences from the Hoosier Database of Native and Nonnative Speech for Children (speechperceptionlab.com). Concatenated sentence streams were made for two randomly chosen males and two randomly chosen females and then summed to produce male and female two-talker babble. Random starting points were chosen to begin each masker stream on each trial and end points were chosen to match the length of the chosen target sentence on any given trial.

The target was presented at 60 dB sound pressure level (SPL) and the masker, when present, was presented at 64 dB SPL, for a target-to-masker ratio of –4 dB. All participants with hearing loss (including some in the ONH group, when thresholds exceeded 25 dB HL) were given compensatory linear gain according to the National Acoustic Laboratories' revised formula with profound correction factor (NAL-RP; Byrne *et al.*, 1991). The gain was calculated from the average thresholds across ears for each individual and applied to the diotic speech signals before delivery to the headphones.

A dual-task structure was employed to assess the listener's ability to make a judgement about the talker change while trying to recognize the words in the

https://doi.org/10.1121/10.0025539

FIG. 2. (Color online) (A) Example one-talker sequence. (B) Example two-talker sequence. Different colors/fonts in (A) and (B) indicate different voices. (C) The response interface.

sequence. Figure 2(C) shows the response interface. Listeners were asked to indicate which keywords were spoken, and also to indicate whether the sequence was spoken by one or two talkers. Responses thus consisted of six mouse clicks (BU) or touch screen taps (MUSC) to indicate the five target words they heard and the number of target talkers (one or two). Participants were free to provide these six responses in any order they wished.

E. Procedures

Participants completed two listening sessions for a total of 2–3 h (including rest periods). In Session 1 all talkers were female and in Session 2 all talkers were male. All of the younger adults chose to complete both sessions during the same visit and all of the older adults chose to complete one session per visit. In each session, the participants received instructions, familiarization trials, two practice blocks, and eight test blocks.

Familiarization trials were self-guided. Participants were presented with an interface with buttons that played sample targets in quiet or in babble. Participants could select either condition and listen to as many examples as they wished. After familiarization, participants completed two practice blocks, each comprising eight trials. The first block was in quiet and the second block was in babble. After each stimulus, the participant responded using the test interface and the correct responses were displayed as feedback. After the practice blocks, participants completed eight test blocks (alternating between quiet and babble to get four of each). Each test block contained 24 trials. Half of the trials included words spoken by only one talker and half had a switch to a second target talker during the sequence.

III. RESULTS

A. Summary of talker change detection and word recognition

Figure 3(A) shows group mean scores for talker change detection in quiet and in babble (left and right panels). The four bars depict the four groups as shown along the x axis. Talker change detection scores were calculated based on 192 trials per participant and noise condition (24 trials $\times$ 4 blocks $\times$ male/female sessions). Figure 3(B) shows similar data for the word recognition task. Word recognition scores were based on a total of 960 keywords per participant and noise condition (24 trials $\times$ 5 keywords $\times$ 4 blocks $\times$ male/female sessions). This summary plot shows that scores were better in quiet than in babble, that word recognition scores were higher than talker change detection scores, and that there were differences in performance across the four groups (see Sec. III B).

To understand the relationship between word recognition and talker change detection in this dual-task paradigm, Fig. 4 shows individual data for the two tasks plotted against each other. In quiet (left panel), word recognition was at or near ceiling for nearly all listeners, whereas an enormous range of scores was observed for the talker change detection task (essentially from chance to ceiling performance). In babble (right panel), there was a range of scores in both tasks, and scores were significantly correlated across participants [$r(33) = 0.755$; $p < 0.0001$]. This confirms that, in babble, the better/worse performers in one task tended to be the better/worse performers in the other task. However, talker change detection scores were consistently poorer than word recognition scores (i.e., below the diagonal, or closer to chance; see also Fig. 3). Moreover, echoing what was seen in quiet, a good word recognition score was no guarantee of a good talker change detection score (note the substantial

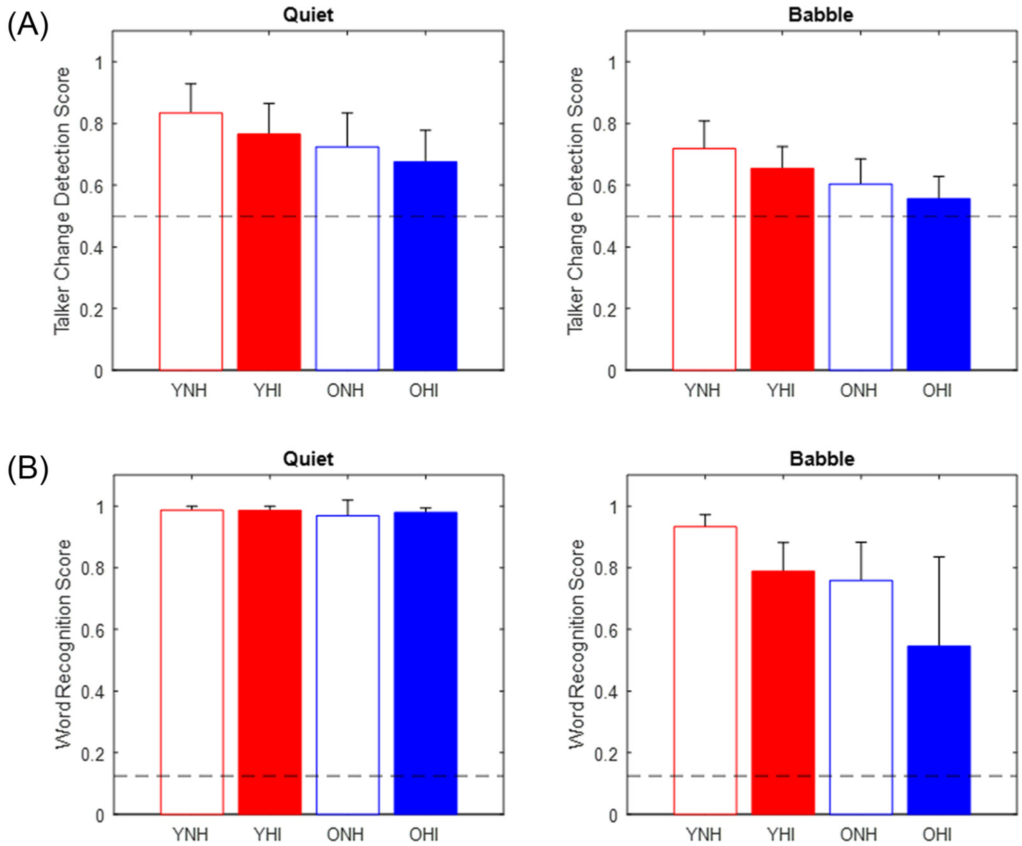

FIG. 3. (Color online) (A) Mean talker change detection scores in quiet (left) and in babble (right). (B) Mean word recognition scores in quiet (left) and in babble (right). The four bars in each panel show the four listener groups as labeled. Error bars show across-subject standard deviations. Dashed lines show chance performance for each task.

vertical spread in the data for high word recognition scores). The opposite was not true; all participants with poor word recognition also had poor talker change detection (and they were all OHI).

To better understand the kinds of errors driving the talker change detection scores, trials were divided into those with no change in talker (one-talker trials) and those with a change in talker (two-talker trials). Figure 5(A) shows scores for these two trial types, in a manner that parallels Fig. 3(A). This breakdown shows that performance for one-talker trials was slightly better than for two-talker trials on average, and also suggests subtle differences in how age and

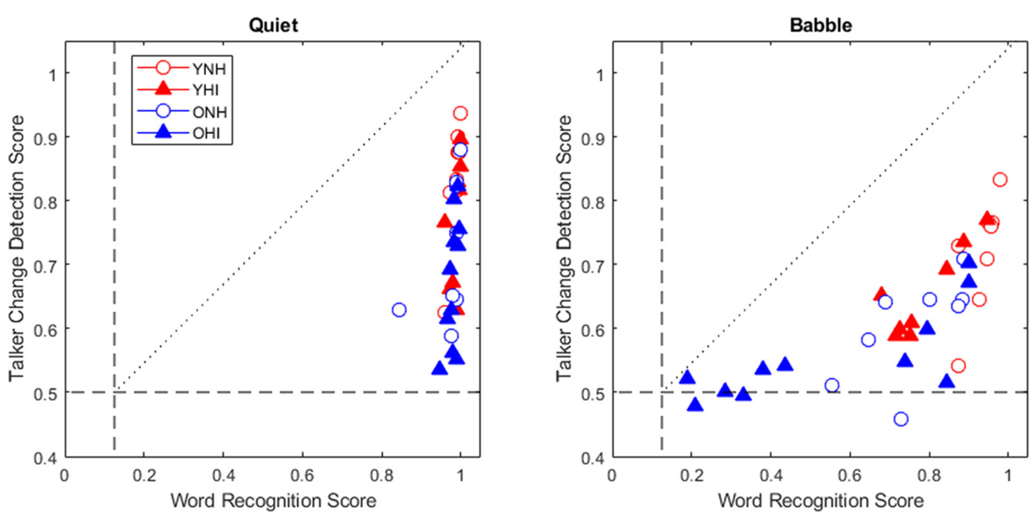

FIG. 4. (Color online) Individual talker change detection scores plotted as a function of word recognition scores in quiet (left) and in babble (right). Colors and symbols indicate the different groups as labeled. Dashed lines indicate chance performance for each task, and the dotted line indicates equivalent performance on the two tasks.

https://doi.org/10.1121/10.0025539

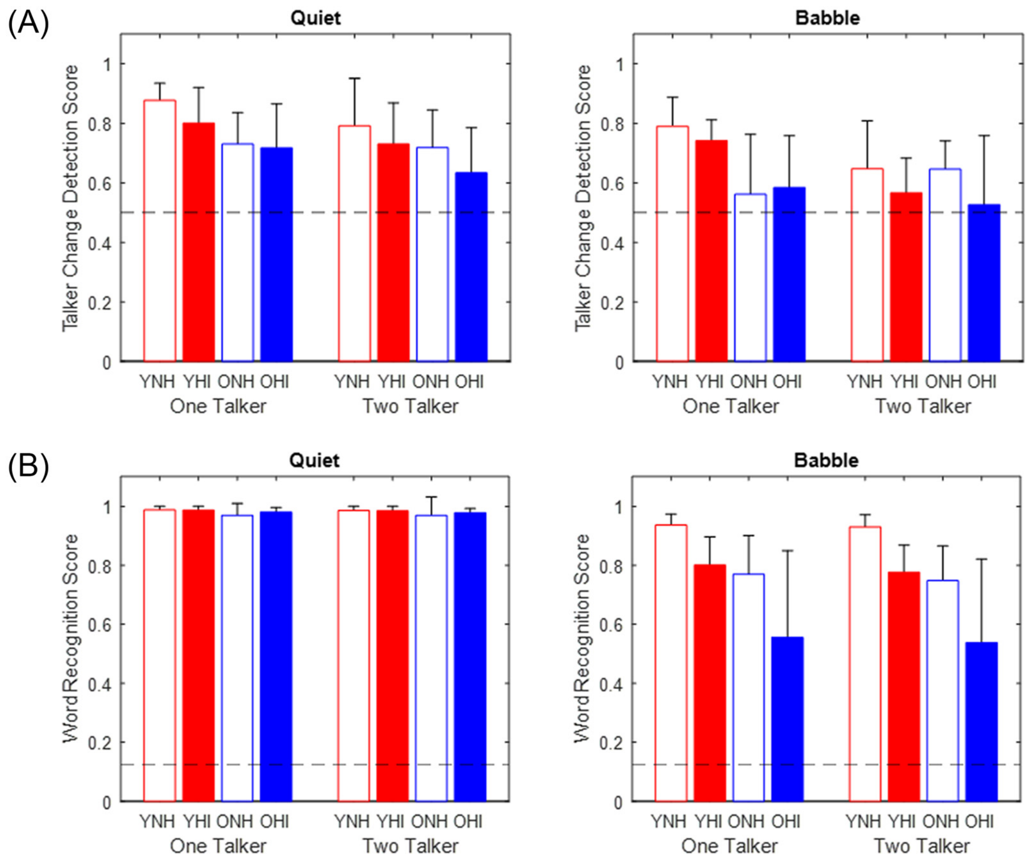

FIG. 5. (Color online) Mean talker change detection scores in quiet (left) and in babble (right) broken down by trial type. The four bars in each cluster show the four listener groups as labeled. Error bars show across-subject standard deviations. Dashed lines show chance performance.

hearing loss affected performance on each kind of trial. In babble, distinct patterns of scores are visible for the one-talker and two-talker trials. Specifically, for one-talker trials, the effects of age (red vs blue bars) are more prominent than the effects of hearing loss (open vs filled bars). Conversely, for two-talker trials, the effects of hearing loss are more prominent than the effects of age. In quiet, the effects of age and hearing loss appear to be more uniform across one-talker and two-talker trials, although a compressed effect of hearing loss for one-talker trials is again apparent in the older groups. Figure 5(B) shows the word recognition data [from Fig. 3(B)] broken down into one-talker and two-talker trials. For this task, the patterns of performance are essentially indistinguishable for one-talker and two-talker trials.

B. Statistical analysis of effects related to age and hearing loss

Talker change detection accuracy across all trials was analyzed in a generalized linear mixed-effects model for binomial data using the packages *lmerTest* and *lme4* in R. The model included fixed effects terms for the categorical factors *noise condition* (quiet vs babble) and *trial type* (one-talker vs two-talker), continuous factors *hearing loss* (all-frequency average) and *age*, and all interactions; the random effects terms included by-participant intercepts. By-participant random slopes were not included because we were testing *a priori* hypotheses about between-subjects factors that should affect performance on our tasks; namely, hearing loss and age. This decision was validated by observing that the same models, when fit including random slopes, were prone to singular fit and low variance on the random slope terms. Contrasts on categorical factors were sum coded. Values of hearing loss and age were standardized for this analysis (converted to *z*-scores). Results of the model fitting and contrasts on model terms are shown in Table II.

The model confirmed that talker change detection scores were significantly affected by noise condition, trial type, hearing loss, and age. Hearing loss interacted significantly with noise condition, revealing slightly larger effects of hearing loss in the quiet condition. Hearing loss also interacted with trial type, confirming our observation from Fig. 5(A) that there is a stronger effect of hearing loss for two-talker than one-talker trials. Age interacted significantly with trial type, confirming our observation from Fig. 5(A) that there is a stronger effect of age for one-talker than two-talker trials. A marginally significant interaction between hearing loss, age, and trial type confirmed the observation that the compressed effect of hearing loss in one-talker trials was most apparent for older listeners. Finally, a significant interaction between age, noise condition, and trial type, confirmed the observation that the compressed effect of age in two-talker trials was most apparent in babble.

An identically structured generalized linear mixed-effects model was applied to the word recognition data.

TABLE II. Generalized linear mixed-effects model for talker change detection scores. Significant *p*-values are bolded.

Model term	Estimate (β)	Standard error	z	p
Hearing loss	−0.226	0.067	−3.381	**<0.001**
Age	−0.233	0.066	−3.554	**<0.001**
Noise condition	−0.304	0.021	−14.581	**<0.001**
Trial type	0.165	0.021	7.929	**<0.001**
Hearing loss × Age	−0.083	0.064	−1.297	0.195
Hearing loss × Noise condition	0.066	0.021	3.100	**<0.002**
Age × Noise condition	0.000	0.021	−0.003	0.998
Hearing loss × Trial type	0.080	0.021	3.793	**<0.001**
Age × Trial type	−0.165	0.021	−7.748	**<0.001**
Noise condition × Trial type	−0.004	0.021	−0.213	0.832
Hearing loss × Age × Noise condition	0.036	0.020	1.784	0.074
Hearing loss × Age × Trial type	0.040	0.020	1.967	**0.049**
Hearing loss × Noise condition × Trial type	0.016	0.021	0.749	0.454
Age × Noise condition × Trial type	−0.097	0.021	−4.560	**<0.001**
Hearing loss × Age × Noise condition × Trial type	0.009	0.020	0.468	0.640

Results of the model fitting and contrasts on model terms are shown in Table III. The model found that word recognition scores were significantly affected by noise condition, trial type, hearing loss, and age. The significant effect of trial type reflected an extremely small drop in scores on two-talker trials [around 1%; see Fig. 5(B)]. Significant two-way interactions between hearing loss and noise condition, and age and noise condition, with stronger effects of hearing loss and age in the babble condition, are driven primarily by the fact that scores were limited by the ceiling in the quiet condition [see Figs. 3(B) and 5(B)]. A significant interaction between hearing loss, age, and noise condition reflects stronger effects of hearing loss in the older group that can only be observed in babble [again because of ceiling effects in quiet; see Figs. 3(B) and 5(B)].

C. Relationship to dissimilarity ratings

Responses on the talker change detection task were compared to subjective ratings of dissimilarity, collected on a different group of YNH listeners for these same stimuli (see Sec. II C). This comparison was done separately for male and female talker sets, as the respective dissimilarity ratings were collected in separate groups of participants. The tendency to report “two talkers” on the talker change detection task was strongly and positively correlated with mean dissimilarity ratings across all pairs of talkers (Fig. 6). The correlation was stronger in quiet (male: $r = 0.92$; female: $r = 0.96$; $p < 0.001$) than in babble (male: $r = 0.80$; female: $r = 0.82$; $p < 0.001$), perhaps because the dissimilarity ratings were obtained in quiet. Note that the talker change detection scores in Fig. 6 were averaged across groups. When the scores were separated out by group, the correlations remained significant in all cases ($p < 0.05$). It is also worth pointing out that these correlations with dissimilarity ratings (which are presumably based on a global collection of features) were substantially stronger than the correlations with any of the individual acoustic features of the voices we considered. For example, in quiet, correlations between talker change detection accuracy and talkers’ mean

TABLE III. Generalized linear mixed-effects model for word recognition scores. Significant *p*-values are bolded.

Model term	Estimate (β)	Standard error	z	p
Hearing loss	−0.491	0.125	−3.925	**<0.001**
Age	−0.388	0.122	−3.173	**<0.002**
Noise condition	−1.357	0.023	−57.824	**<0.001**
Trial type	0.065	0.023	2.775	**<0.006**
Hearing loss × Age	−0.181	0.119	−1.515	0.130
Hearing loss × Noise condition	−0.342	0.024	−14.438	**<0.001**
Age × Noise condition	−0.136	0.023	−5.822	**<0.001**
Hearing loss × Trial type	0.013	0.024	0.559	0.576
Age × Trial type	−0.025	0.023	−1.081	0.280
Noise condition × Trial type	0.011	0.023	0.468	0.640
Hearing loss × Age × Noise condition	−0.131	0.022	−5.893	**<0.001**
Hearing loss × Age × Trial type	−0.015	0.022	−0.698	0.485
Hearing loss × Noise condition × Trial type	−0.039	0.024	−1.671	0.095
Age × Noise condition × Trial type	0.019	0.023	0.801	0.423
Hearing loss × Age × Noise condition × Trial type	−0.007	0.022	−0.339	0.735

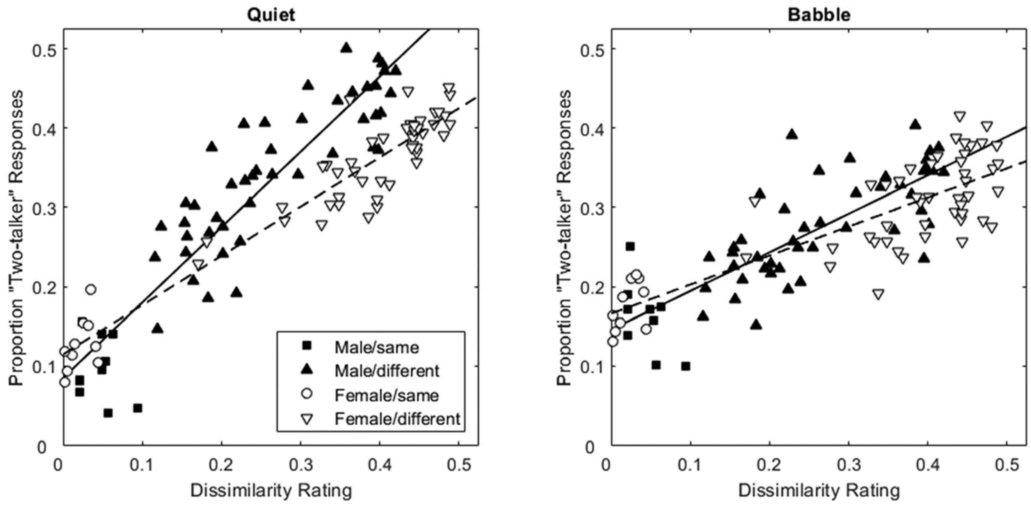

FIG. 6. Mean proportion of "two talker" responses in quiet (left) and babble (right) plotted as a function of mean dissimilarity ratings for each talker pair. Male and female talker sets are shown with filled/open symbols, respectively, and solid/dashed lines show lines of best fit.

$F0$ difference were relatively weak (male: $r = 0.26$; female: $r = 0.49$; $p < 0.05$). A similar observation was made by Sharma *et al.* (2019), who found that $F0$ differences did a very poor job of predicting reaction times in a talker change detection task, with better predictions for more complex representations of acoustic similarity.

IV. DISCUSSION

The goal of the current study was to determine how age and hearing loss affect sensitivity to talker-related information in speech. We used an experimental task that assessed a listener's ability to detect changes in the talker during a five-word sequence. This allowed us to use a large set of voices and words, capturing the variability that is present in real-world speech perception, but avoided the added cognitive requirements of learning and remembering specific individual voices. Talker change detection judgements were strongly correlated with dissimilarity ratings collected independently for these stimuli, suggesting that participants in our task based their decision on a broad and complex set of features that make these voices sound more or less alike, rather than on a change in a particular feature (such as $F0$). One perhaps unusual feature of our task was that talker changes occurred in the middle of a syntactically correct sentence (rather than between sentences like in the task of Sharma *et al.*, 2019; Sharma *et al.*, 2020). While this kind of change may break certain expectations about the relationship between speech structure and speaker identity (Narayan *et al.*, 2017; Warnke and de Ruiter, 2023), it can also be viewed as a realistic feature of lively conversations in which multiple people contribute, interrupt, etc. Our experimental task also allowed us to impose the realistic demands of understanding the speech while discerning unexpected changes in the talkers, and doing both tasks in the presence of irrelevant competing speech. This dual-task structure may have diluted attention to the talker change detection task through a kind of "change deafness" (Vitevitch, 2003), which could explain the wide variation in scores on the change detection task we observed across listeners. In future studies, it would be interesting to examine the extent to which the observed pattern of results related to age and hearing loss holds for different sentence structures, different task structures, and different background environments.

The results suggest that age and hearing loss have independent effects on talker change detection, a finding that reinforces that of our previous study using a talker identification paradigm (Best *et al.*, 2018). We deliberately chose a talker change detection task in the current study to focus on the immediate perception of voices, avoiding the requirement for listeners to learn and remember specific voices that limited the strength of the conclusions about age in the previous study. On the other hand, we also introduced a dual-task component, which may have been generally more taxing for older adults and contributed to the detrimental effect of age. Overall, however, we can now say that the effects of age seem to be consistently observed across a variety of talker perception tasks with different cognitive demands (Best *et al.*, 2018; Xu *et al.*, 2021; Zaltz and Kishon-Rabin, 2022).

Interestingly, there were indications that hearing loss and age may have different effects on two complementary processes that are involved in talker change detection. First, we assume that on two-talker trials, a correct response relies on listeners distinguishing between the two presented voices ("telling people apart"). Second, we assume that on one-talker trials, a correct response relies on listeners maintaining a consistent perception of a voice across the presented words ("telling people together"). We found that effects of hearing loss were stronger for two-talker than one-talker trials, which provides support for the speculation made by Best *et al.* (2018) that hearing loss may disrupt the fine acoustic distinctions required for discriminating between voices. Conversely, we found that effects of age were stronger for one-talker than two-talker trials, which supports the idea that advanced age may affect the ability to associate variable speech samples with the same talker as is required to learn new voices (Yonan and Sommers, 2000; Best *et al.*, 2018). In future work, it would be interesting to examine whether the ability to tell people apart/together improves over time with extended testing. If telling people apart relies primarily on fine acoustic distinctions, and the limits imposed by hearing loss and age are peripheral in nature, performance may be relatively stable across time. On the other hand, if telling people together relies on building up a detailed picture of unique voices, then we might expect improvements in performance over time, although the learning profile may differ for younger and older listeners. Investigations of telling people together in these populations could also make use of highly familiar vs unfamiliar voices (Johnsrude *et al.*, 2013; Stevenage *et al.*, 2023).

By combining our talker change detection task with a word recognition task, within the same stimulus, we were able to compare performance on these two tasks directly. This comparison indicated that effects of age and hearing loss on talker change detection are not simple extensions of the well-known effects of age and hearing loss on speech intelligibility. For example, in quiet, where word recognition was close to ceiling performance for all groups, talker change detection scores varied dramatically across listeners with some performing close to chance. In the more difficult babble conditions, where both tasks were away from ceiling, scores were correlated across tasks, but there was again a striking range of scores for talker change detection, even when word recognition accuracy was high. This dissociation may reflect differences in the acoustic features that are important for speech intelligibility vs talker perception. For example, while spectral characteristics (including $F0$) are highly informative regarding the gender and identity of different talkers, temporal modulations are generally thought to be the most essential acoustic features for speech recognition (Shannon *et al.*, 1995; Villard and Kidd, 2021). The dissociation may also reflect the amount of compensatory support that is available to complete each task. For example, speech intelligibility can be maintained in the face of severe degradation by "filling in" missing information based on linguistic knowledge, context, and expectations (e.g., Samuel, 1981; Sivonen *et al.*, 2006). If equally robust support mechanisms are not available for talker perception, then the fidelity of the peripheral representation may be more critical.

Reduced sensitivity to talker cues likely has several consequences for social communication in real-world contexts. First, the ability to detect changes in the talker is critical for following the natural flow of conversations. Second, the ability to focus on a talker of interest, based on their voice, is important for dealing with competition between simultaneous voices in noisy situations. Third, accurate encoding of talker identity may be an important factor allowing us to remember "who said what." Disruptions to some or all of these processes may well contribute to the communication challenges that accompany older age and hearing loss, compounding the effects of reduced audibility and poor speech intelligibility.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health–National Institute on Deafness and Other Communication Disorders (NIH-NIDCD) Grant Nos. R01 DC04545 (to G.K.), R01 DC000184 (to J.R.D.), R01 DC013286 (to G.K. and V.B.), and the National Center for Advancing Translational Sciences of the National Institutes of Health (Grant No. UL1 TR001450 to MUSC). The research at MUSC was conducted in a facility constructed with support from Research Facilities Improvement Program (Grant No. C06 RR 014516) from the National Institutes of Health/National Center for Research Resources. The authors thank Erin Lynch and Silvana Tellerico for assistance with data collection at MUSC and Chinazo Otiono, Jessica Tin, and Sung-Joo Lim for collecting the perceptual dissimilarity ratings at BU.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

All participants provided informed consent and all procedures were approved by the Boston University Institutional Review Board and the Institutional Review Board of the Medical University of South Carolina.

DATA AVAILABILITY

The data that support the findings of this study are available from the authors upon reasonable request.

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

Best, V., Ahlstrom, J. B., Mason, C. R., Roverud, E., Perrachione, T. K., Kidd, G., Jr., and Dubno, J. R. (**2018**). "Talker identification: Effects of masking, hearing loss, and age," J. Acoust. Soc. Am. **143**, 1085–1092.

Byrne, D. J., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe-profound hearing losses," in *The Vanderbilt Hearing Aid Report II*, edited by G. A. Studebaker, F. H. Bess, and L. B. Beck (York Press, Parkton, MD), pp. 295–300.

Humes, L. E., and Dubno, J. R. (**2010**). "Factors affecting speech understanding in older adults," in *The Aging Auditory System*, edited by S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay (Springer, New York), pp. 211–257.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., and Carlyon, R. P. (**2013**). "Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice," Psychol. Sci. **24**, 1995–2004.

Kanber, E., Lavan, N., and McGettigan, C. (**2022**). "Highly accurate and robust identity perception from personally familiar voices," J. Exp. Psych. Gen. **151**, 897–911.

Kawahara, H., de Cheveigné, A., Banno, H., Takahashi, T., and Irino, T. (**2005**). "Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT," in *Proceedings Interspeech 2005*, Lisbon, Portugal, pp. 537–540.

Kidd, G., Jr., Best, V., and Mason, C. R. (**2008**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.

Kreiman, J., and Sidtis, D. (**2011**). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception* (Wiley-Blackwell, West Sussex, UK).
Kreiman, J., Vanlancker-Sidtis, D., and Gerratt, B. R. (**2005**). "Perception of voice quality," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell Publishing Ltd, Malden, MA), pp. 338–362.
Lavan, N., Burton, A. M., Scott, S. K., and McGettigan, C. (**2019**). "Flexible voices: Identity perception from variable vocal signals," Psychon. Bull. Rev. **26**, 90–102.
Lee, Y., Keating, P., and Kreiman, J. (**2019**). "Acoustic voice variation within and between speakers," J. Acoust. Soc. Am. **146**, 1568–1579.
Lee, Y., and Kreiman, J. (**2022**). "Acoustic voice variation in spontaneous speech," J. Acoust. Soc. Am. **151**, 3462–3472.
Mathias, S. R., and von Kriegstein, K. (**2014**). "How do we recognise who is speaking," Front. Biosci. **6**, 92–109.
Narayan, C. R., Mak, L., and Bialystok, E. (**2017**). "Words get in the way: Linguistic effects on talker discrimination," Cog. Sci. **41**, 1361–1376.
Perrachione, T. K., Chiao, J. Y., and Wong, P. C. (**2010**). "Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices," Cognition **114**, 42–55.
Perrachione, T. K., Furbeck, K. T., and Thurston, E. J. (**2019**). "Acoustic and linguistic factors affecting perceptual dissimilarity judgments of voices," J. Acoust. Soc. Am. **146**, 3384–3399.
Samuel, A. G. (**1981**). "Phonemic restoration: Insights from a new methodology," J. Exp. Psych. Gen. **110**, 474–494.
Schweinberger, S. R., and Zaske, R. (**2018**). "Perceiving speaker identity from the voice," in *The Oxford Handbook of Voice Perception*, edited by S. Fruhholz and P. Belin (Oxford University Press, Oxford), pp. 539–560.
Scott, S., and McGettigan, C. (**2016**). "The voice: From identity to interactions," in *APA Handbook of Nonverbal Communication*, edited by D. Matsumoto, H. C. Hwang, and M. G. Frank (American Psychological Association, Washington, DC), pp. 289–305.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Sharma, N. K., Ganesh, S., Ganapathy, S., and Holt, L. L. (**2019**). "Talker change detection: A comparison of human and machine performance," J. Acoust. Soc. Am. **145**, 131–142.
Sharma, N. K., Krishnamohan, V., Ganapathy, S., Gangopadhayay, A., and Fink, L. J. (**2020**). "Acoustic and linguistic features influence talker change detection," J. Acoust. Soc. Am. **148**, EL414–EL419.
Sivonen, P., Maess, B., Lattner, S., and Friederici, A. D. (**2006**). "Phonemic restoration in a sentence context: Evidence from early and late ERP effects," Brain Res. **1121**, 177–189.
Stevenage, S. V., Singh, L., and Dixey, P. (**2023**). "The curious case of impersonators and singers: Telling voices apart and telling voices together under naturally challenging listening conditions," Brain Sci. **13**, 358.
Villard, S., and Kidd, G., Jr. (**2021**). "Speech intelligibility and talker gender classification with noise-vocoded and tone-vocoded speech," JASA Express Lett. **1**, 094401.
Vitevitch, M. S. (**2003**). "Change deafness: The inability to detect changes between two voices," J. Exp. Psych. Human Percept. Perform. **29**, 333–342.
Warnke, L., and de Ruiter, J. P. (**2023**). "Top-down effect of dialogue coherence on perceived speaker identity," Sci. Rep. **13**, 3458.
Xu, M., Shao, J., and Wang, L. (**2021**). "Effects of aging and age-related hearing loss on talker discrimination," in *Proceedings Interspeech 2021*, pp. 1728–1732.
Yonan, C. A., and Sommers, M. S. (**2000**). "The effects of talker familiarity on spoken word identification in younger and older listeners," Psychol. Aging **15**, 88–99.
Zaltz, Y., and Kishon-Rabin, L. (**2022**). "Difficulties experienced by older listeners in utilizing voice cues for speaker discrimination," Front. Psychol. **13**, 797422.