Auditory Learning and Generalization in Older Adults: Evidence from Voice Discrimination Training

Nuphar Singer[1] and Yael Zaltz[1,2]

Abstract

Auditory learning is essential for adapting to continuously changing acoustic environments. This adaptive capability, however, may be impacted by age-related declines in sensory and cognitive functions, potentially limiting learning efficiency and generalization in older adults. This study investigated auditory learning and generalization in 24 older (65–82 years) and 24 younger (18–34 years) adults through voice discrimination (VD) training. Participants were divided into training (12 older, 12 younger adults) and control groups (12 older, 12 younger adults). Trained participants completed five sessions: Two testing sessions assessing VD performance using a 2-down 1-up adaptive procedure with F0-only, formant-only, and combined F0 + formant cues, and three training sessions focusing exclusively on VD with F0 cues. Control groups participated only in the two testing sessions, with no intermediate training. Results revealed significant training-induced improvements in VD with F0 cues for both younger and older adults, with comparable learning efficiency and gains across groups. However, generalization to the formant-only cue was observed only in younger adults, suggesting limited learning transfer in older adults. Additionally, VD training did not improve performance in the combined F0 + formant condition beyond control group improvements, underscoring the specificity of perceptual learning. These findings provide novel insights into auditory learning in older adults, showing that while they retain the ability for significant auditory skill acquisition, age-related declines in perceptual flexibility may limit broader generalization. This study highlights the importance of designing targeted auditory interventions for older adults, considering their specific limitations in generalizing learning gains across different acoustic cues.

Introduction

Auditory-perceptual learning allows listeners to adjust to difficult listening conditions (e.g., to understand fast-rate speech, speech amid background noise, and competing talkers), using an interplay between top-down (cognitive) and bottom-up (sensory) processes (Bieber & Gordon-Salant, 2021; Davis & Johnsrude, 2007; Samuel, 2011; Samuel & Kraljic, 2009). With age, the sensory and cognitive processes mediating perceptual learning deteriorate (Anderson & Karawani, 2020; Harada et al., 2013; Jayakody et al., 2018; Pichora-Fuller & Singh, 2006; Schneider, 2011). Accordingly, the process of learning and the ability to generalize the learning gains to new conditions and tasks may be significantly impaired with increasing age. One way to explore the learning capacity of older adults is via auditory training. However, findings from the literature addressing the

efficacy of learning following auditory training for older adult listeners are limited and highly varied (Bieber & Gordon-Salant, 2021). The present study sought to compare the learning and generalization characteristics of older adults to younger adults, using auditory training with a basic

[1]Department of Communication Disorders, The Stanley Steyer School of Health Professions, Faculty of Medical and Health Science, Tel Aviv University, Tel Aviv, Israel
[2]Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

Corresponding author:
Yael Zaltz, Department of Communication Disorders, The Stanley Steyer school of Health Professions, Faculty of Medical and Health Science, Tel Aviv University, Tel Aviv, Israel.
Emails: yaelzaltz@gmail.com, yaelzalt@tauex.tau.ac.il

Data Availability Statement included at the end of the article

psychoacoustic task of voice discrimination (VD). Such a comparison may help infer how age could influence the neural processes undergoing reorganization after training (Ahissar et al., 2009; Censor, 2013; Karni, 1996; Karni & Sagi, 1993) and bear significance to the latent potential for enhancing functioning in advanced aging.

Perceptual skill learning is defined as a long-lasting improvement in the ability to respond to the environment and extract information from a given stimulus, due to experience (Gibson, 1963; Goldstone, 1998). One constraint on perceptual learning is that the training-induced gains may become more specific to the trained task and/or conditions as training progresses (Zaltz et al., 2020), suggesting that both the magnitude and the nature of changes in neural encoding of an auditory task evolve progressively throughout the learning process (Ahissar et al., 2009; Ahissar & Hochstein, 2004). Furthermore, the learning specificity was suggested to be affected by the similarity, or overlap in brain representation, between the trained and untrained stimuli or tasks, with better generalization of the learning-gains to untrained tasks or conditions that share the same neural circuits as the trained task (Amitay et al., 2014; Hesseg et al., 2016). Exploring the extent of generalization may therefore serve as a possible probe into the neural processes undergoing reorganization following training.

Understanding how perceptual learning interacts with age-related changes in peripheral and central auditory abilities is critical, as aging is associated with declines in auditory functions that extend beyond reduced hearing sensitivity (as measured by audiometry). These include deficits in the temporal and spectral processing of the auditory input (Goupell et al., 2017; Moore & Peters, 1992; Schvartz-Leyzac & Chatterjee, 2015) that may influence both the learning process and the extent of generalization. In terms of temporal processing, electrophysiological studies reveal significant age-related declines in temporal precision, particularly within the auditory brainstem (Anderson et al., 2011; Purcell et al., 2004). These deficits are thought to stem from diminished phase-locking (Anderson et al., 2021) and decreased auditory nerve synchrony (Anderson et al., 2012). Such temporal impairments may contribute to the poorer F0 discrimination observed in older adults, who require approximately twice the F0 difference to perform at levels comparable to younger listeners when tested with harmonic complex tones and synthetic vowels (Moore & Peters, 1992; Vongpaisal & Pichora-Fuller, 2007). In terms of spectral processing, broadened auditory filters—likely due to reduced outer hair cell function (Tun et al., 2012)—have been associated with the poorer speech-in-noise performance observed in older adults, even among those with normal audiometry (Nambi et al., 2016). Furthermore, age-related changes in the spiral ganglion and auditory nerve can further contribute to spectral resolution deficits, particularly in more advanced stages of age-related hearing loss (Schmiedt, 2010). Additionally, aging negatively affects top-down, cognitive abilities in the domains of working memory, attention, and inhibition (Craik & Salthouse, 2000; Schneider & Pichora-Fuller, 2000). Such declines may adversely impact the capacity for perceptual learning.

The literature regarding age effects on auditory perceptual learning is both limited and inconsistent (Bieber & Gordon-Salant, 2021). For example, listeners aged 65–87 showed rapid plateauing during implicit training on accented speech, with improvements ceasing after the first block (Adank & Janse, 2010). In contrast, substantial learning gains have been observed in older adults when engaging in explicit training paradigms that provided trial-by-trial feedback. Examples include gap detection training in individuals aged 60–80 (Kishon-Rabin et al., 2013), time-compressed speech (TCS) training in those aged 65–91 (Manheim et al., 2018), combined training on speech-in-noise, TCS, and competing speakers in individuals aged 60–72 (Karawani et al., 2015), and speech-in-noise training in participants aged 61–79 (Humes et al., 2014).

Additionally, evidence regarding the effect of age on the transfer of learning to new stimuli or conditions is varied, ranging from extensive (Fostick et al., 2020; Sabin et al., 2013) to restricted generalization (Manheim et al., 2018; Peelle & Wingfield, 2005). For example, older adults aged 56–82 displayed broad generalization of learning gains after undergoing spectral modulation detection training, exhibiting superior transfer to an untrained spectral modulation frequency compared to the young control group (Sabin et al., 2013). Conversely, following TCS training, older adults showed poorer generalization to untrained compression ratios (65–78 year olds, Peelle & Wingfield, 2005) or untrained sentences (65–91 years old, Manheim et al., 2018) compared to their younger counterparts. Manheim et al. (2018) proposed that older adults rely more on higher-level representations of training material, such as the semantic meaning of sentences, than lower-level representations like the acoustic structure. Consequently, they exhibit decreased generalization of the learning-gains to untrained material that solely shares acoustic properties with the trained material compared to younger adults. One way to explore whether this reduced generalization is a general characteristic of older adults or depends on task complexity is to employ a training protocol focused on a basic discrimination task, such as VD, rather than a speech perception task like TCS. This approach allows for the assessment of learning generalization in a task that emphasizes fundamental psychoacoustic processing.

VD holds significant importance in differentiating speakers within a multi-talker environment, a skill linked to improved speech perception amid background noise (Bronkhorst, 2015; Vestergaard et al., 2009). This task involves discriminating between auditory stimuli (e.g., words or sentences) based on variations in fundamental frequency (F0), influenced by vocal cord characteristics like length, mass, and vibration rate, and formant frequencies shaped by the vocal tract length of the speaker (Darwin et al., 2003; Mackersie et al., 2011; Schvartz-Leyzac & Chatterjee, 2015; Skuk & Schweinberger, 2013; Vestergaard et al., 2009, 2011).

These acoustic cues convey crucial information about the speaker, encompassing attributes such as age, gender, and individual characteristics (Başkent & Gaudrain, 2016; Shultz, 2015; Skuk & Schweinberger, 2013; Smith et al., 2007; Smith & Patterson, 2005). Research indicates that efficient F0 coding relies largely on processing the temporal envelope and/or temporal fine-structure cues of the signal, while formant coding predominantly involves the place-based coding of spectral energy peaks (Carlyon & Shackleton, 1994; Fant, 1960; Fu et al., 2004; Lieberman & Blumstein, 1988; Oxenham, 2008; Xu & Pfingst, 2008). Considering the temporal and spectral processing demands for effective F0 and formant perception, it is not surprising that a growing body of research indicates older adults exhibit poorer abilities in F0 discrimination (across various age groups, Anderson et al., 2021; Anderson & Karawani, 2020; Moore & Peters, 1992; Souza et al., 2011; Vongpaisal & Pichora-Fuller, 2007), vowel identification based on formant changes (60–81 years old, Chintanpalli et al., 2016, pp. 65–78 years old, Goupell et al., 2017, pp. 65–83 years, Vongpaisal & Pichora-Fuller, 2007), and VD based on either F0 or formant cues, compared to younger groups (65–78 years, Zaltz & Kishon-Rabin, 2022).

Although no study has yet attempted to enhance VD performance in older adults through training, findings from studies with young adults suggest that voice training can significantly improve performance across various tasks, such as explicit voice identification, implicit voice familiarity, and VD tasks (Kreitewolf et al., 2017; Nygaard et al., 1994; Yonan & Sommers, 2000; Zaltz, 2024). For example, two training sessions focused on VD using either F0 or formant cues resulted in significant improvements in performance (Zaltz, 2024). Similarly, two-session voice identification or familiarization training with sentence stimuli significantly enhanced both VD and voice recognition (Yonan & Sommers, 2000). Notably, even brief training protocols can yield measurable benefits in young adults. For instance, a 10-min voice training session improved speech intelligibility for speech produced by the trained voice (Holmes et al., 2021). However, in some cases, the effects of voice training have been reported to remain highly specific to the trained task and stimuli. Yonan and Sommers (2000) found that sentence-based voice training enhanced sentence recognition for the trained voice but did not improve the recognition of isolated words. Similarly, Biçer et al. (2023) demonstrated that while a 30-min audiobook listening protocol reduced pupil dilation when listening to the trained voices, indicating reduced listening effort, it did not improve VD with the same voices. Additionally, Zaltz (2024) recently reported that VD improvements did not generalize to the same trained VD task when using untrained voice cues. The limited generalization in young adults raises concerns about whether older adults would be able to generalize their learning gains following VD training, assuming they demonstrate improvement.

Table 1. Mean Pure-Tone Thresholds (in dB) Across Octave Frequencies for Training and Control Groups (Mean of two Ears), and Results of Multivariate Analysis.

Frequency (Hz)	Training Group M (SD)	Control Group M (SD)	p-value
250	22.71 (3.76)	24.58 (2.46)	.288
500	23.75 (4.46)	19.37 (4.15)	.021*
1000	22.33 (5.87)	18.75 (4.70)	.066
2000	31.87 (4.66)	23.33 (6.43)	.001**
4000	42.91 (9.58)	39.27 (7.02)	.115
8000	61.04 (11.89)	57.7 (14.78)	.549

Note. $* = p < .05$, $** = p < .005$.

The present study aims to advance our understanding of the constraints that influence auditory learning and generalization in older adulthood, using VD training that is focused on F0 information. To this end, two groups of participants, young adults and older adults underwent three-session VD training based on F0 cues. The extent of generalization of the learning gains was assessed to a different acoustic cue (formants) and to combined F0 + formant cues. Performance in these conditions was assessed by comparing it with control groups of young and older adults that did not undergo any training. The hypothesis suggested that while older adults would generally demonstrate lower VD compared to young adults (Zaltz & Kishon-Rabin, 2022), they would also benefit from training (Humes et al., 2014; Karawani et al., 2015; Kishon-Rabin et al., 2013; Manheim et al., 2018). However, poorer generalization of the learning gains to untrained conditions was expected, reflecting differences in the neural circuits engaged in learning at varying ages (Amitay et al., 2014; Hesseg et al., 2016).

Materials and Methods

Participants

Forty-eight participants took part in the study. Twenty-four young adults aged 18–34 years were divided into training (mean age of 28.58 years ± 4.10, $n = 12$) and control groups (mean age of 23.09 years ± 3.17, $n = 12$), and 24 older adults, aged 65–82 years were divided into training (mean age of 72.67 years ± 6.36, $n = 12$) and control groups (mean age of 73.18 years ± 5.18, $n = 12$). None of the participants had prior experience with similar psychoacoustic tasks, nor did they have any documented history of ear disease. Additionally, all participants had completed a minimum of 12 years of formal education. The young adults exhibited normal pure-tone air conduction thresholds in both ears, registering pure-tone air conduction thresholds of ≤20 dB HL across octave frequencies from 250 to 8,000 Hz (ANSI, 2018). The older adults displayed pure-tone air conduction thresholds that were equal to or superior to the 50th percentile for their age group (Engdahl et al., 2005). Table 1 presents the mean pure-

tone thresholds across the tested frequencies for the training and control groups of older participants, while Figure 1 provides detailed audiograms. Overall, hearing thresholds ranged from normal to mild hearing loss (up to 40 dB HL) at 2,000 Hz, minimal to moderately severe loss (25–70 dB HL) at 4,000 Hz, and mild to severe loss (35–90 dB HL) at 8000 Hz. Some variation was observed between the training and control groups, with statistically significant differences at 500 and 2000 Hz. None of the older adults had utilized a hearing aid. Each older adult demonstrated cognitive abilities within the normal range (Mini-Mental State Examination [MMSE] score ≥ 25, based on the English version [Folstein et al., 1975]), maintained independent living, and reported leading an active lifestyle. Informed consent was obtained from all participants. The study received approval from the Institutional Review Board of Tel Aviv University (approval number: 0007524-2).

VD Test

Stimuli. The specific VD test utilized in this study was previously described (Zaltz, 2023, 2024). In summary, the test involved 82 single-syllable CVC words from the Hebrew version of the Arthur Boothroyd (HAB) test (Kishon-Rabin et al., 2004). The words were recorded by a female native speaker in a soundproof room using an AT-892-TH microphone and Sound-Forge software (version 7.0). The recordings utilized stereo channels at a sampling rate of 44,100 Hz and a 16-bit quantization level. Ensuring consistent intensity levels across all stimuli, amplitudes were normalized to −16 dB RMS. The words were altered within a 14-point stimulus continuum, adjusting either F0 solely, formant frequencies solely, or both F0 and formants. This continuum followed an exponential progression in $\sqrt{2}$ steps, ranging from a change of −0.127 semitone to −8 semitones, mirroring techniques in previous papers (Levin & Zaltz, 2023; Zaltz, 2023; Zaltz et al., 2018, 2020; Zaltz & Kishon-Rabin, 2022). Specifically, the mean F0 was adjusted in increments of 0, −0.127, −0.18, −0.26, −0.36, −0.51, −0.72, −1.02, −1.44, −2.02, −2.86, −4.02, −5.67, and −8 semitones from the original stimulus's mean F0. This manipulation was carried out using the PSOLA algorithm (Moulines & Charpentier, 1990) for pitch extraction and modification. For instance, if a word had a mean F0 of 175.62 Hz, lowering the F0 transitioned the comparison word exponentially from 174.33 to 110.35 Hz in $\sqrt{2}$ steps. When manipulating formant frequencies, adjustments ranged exponentially from a ratio of 0.99 (the smallest change from the original frequencies) to 0.63 (the most significant change), necessitating resampling the stimulus to compress the frequency axis. This compression was achieved using factors akin to the F0 change, followed by the application of the PSOLA algorithm to restore the original pitch and duration.

VD Threshold Assessment. The study employed the three-interval three-alternative forced-choice (3I3AFC)

method to evaluate VD difference limens (DLs) based on F0, formants, or their combination. Each trial displayed two unprocessed reference stimuli alongside one manipulated comparison stimulus, timed with a 300-millisecond interstimulus interval. A corresponding square on the PC monitor signaled each stimulus presentation. Participants were directed to select the stimulus they perceived as "sounding different" by clicking the respective square using a computer mouse. The initial comparison stimulus underwent the most significant manipulation, with F0 lowered by eight semitones and formants adjusted by a ratio of 0.63. An adaptive tracking procedure, following a two-down, one-up pattern, was utilized to establish DLs corresponding to a 70.7% detection threshold on the psychometric function (Levitt, 1971). The difference between reference and comparison stimuli was successively halved until the first reversal, then adjusted using a $\sqrt{2}$ factor until the sixth reversal. Just noticeable differences were calculated as the average of the last four reversals, following a protocol documented in prior studies (Levin & Zaltz, 2023; Zaltz, 2023; Zaltz et al., 2018, 2020; Zaltz & Kishon-Rabin, 2022). Participants were not limited by time constraints in making their selections. Throughout the VD assessment, no feedback was provided regarding participants' responses. However, during training sessions, visual feedback was provided: The selected square on the PC monitor would illuminate in green for correct responses and in red for incorrect ones. Each VD block lasted around 3–4 min. Before formal testing, participants engaged in a short familiarization task with 5–10 trials featuring the greatest difference between the reference and comparison stimuli to confirm their understanding of the task.

Study Design

The trained participants completed a series of five sessions, two testing sessions and three training sessions, with 1–3 days between each consecutive session (Figure 2). The testing sessions were the first and last (fifth) sessions. They involved six VD blocks: two with F0-only cues, two with formant-only cues, and two combining F0 + formant cues. The order of VD cues was counterbalanced among participants. Additionally, during the first testing session, all participants underwent a bilateral air-conduction threshold hearing test across octave frequencies ranging from 250 to 8000 Hz (ANSI, 2018), and only the older adults underwent cognitive screening using the Hebrew version of the MMSE (Folstein et al., 1975). This assessment consists of 11 questions and covers various mental functions. A score of 25 points or more indicated cognitive abilities within the normal range. Overall, the first session lasted approximately 90 min, while the fifth session lasted about 45 min, with short breaks provided as needed.

Sessions 2–4 were dedicated to training, consisting of eight VD blocks using F0-only cues, with each session lasting approximately 60 min.
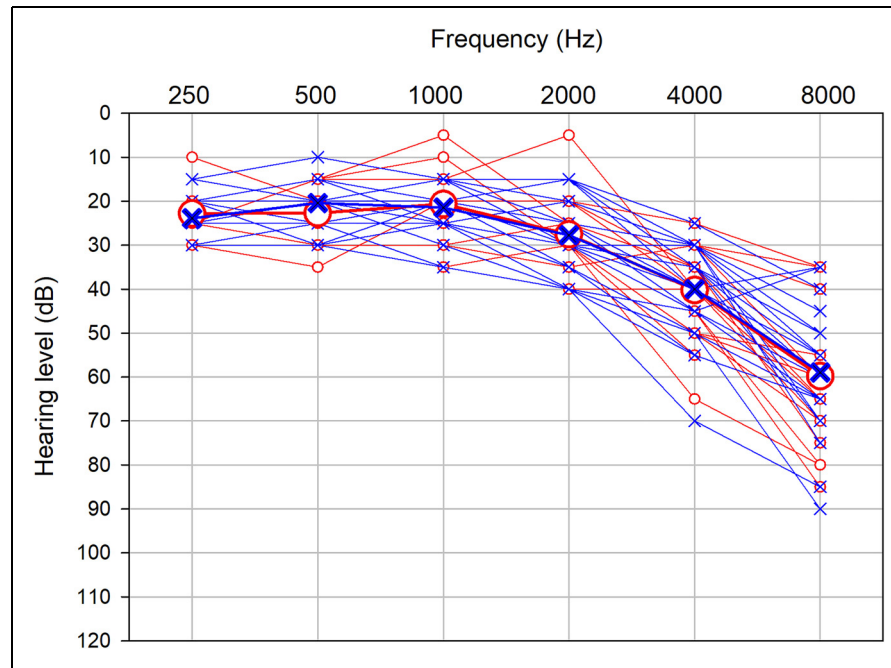
Figure 1. Mean (thick lines and symbols) and Individual (thin Lines and Symbols) Hearing Thresholds Across Octave Frequencies Ranging from 250 to 8000 Hz for the Right (red circle) and left (blue cross) Ears for the Older Adults ($n = 24$).

The control groups completed only two testing sessions, spaced 7–10 days apart. These sessions were identical to the first and fifth session of the training groups.

Apparatus and Data Analysis. The study was conducted in a sound-treated, single-walled room. Stimuli were delivered using the internal soundcard of a laptop personal computer through a GSI-61 audiometer to both ears via TDH-49 headphones. The individual mean pure-tone air conduction thresholds at 500, 1,000, 2,000, and 4,000 (PTA4) were calculated for each ear. These values served as a baseline to establish presentation levels for the VD task, and stimuli were presented at 35 dB sensation level (SL) above the individual PTA4, aiming to approximate a balanced SL across young and older participants. Pearson correlation coefficients showed no significant correlations between PTA4 and VD performance (mean of the two blocks) for either acoustic cue during the first testing session, in both age groups ($p > .05$).

Statistical analyses were carried out using SPSS-28 software (IBM Corp, Armonk, NY). All VD data underwent logarithmic transformation before being entered into analyses, following the approach used in previous studies (El Boghdady et al., 2019; Koelewijn et al., 2021, 2023; Zaltz, 2024). Two multilevel models (MLMs) for repeated measures were employed. The first MLM was used to assess learning, with a three-level structure: blocks (level 1) nested within sessions (level 2), and further within participants (level 3). Session and block were modeled as linear trends, and age

group as a between-participant factor. Power was determined using G*Power (Faul et al., 2007). Based on previous studies reporting medium to large effect sizes for similar interactions (Anderson et al., 2022; Karawani et al., 2015; Manheim et al., 2018), we adopted a conservative estimate of Cohen's $f = 0.25$ with $\alpha < 0.05$. The hypothesized age group * session * block interaction had a power of 0.99 in our sample.

The second MLM was utilized to assess generalization, with a five-level structure: Blocks (level 1) nested within sessions (level 2) and cue type (level 3), further nested within participants (level 4). Session, cue type and block were treated as within-person variables, and age group and group as between-participant factors. All variables were effect-coded so that lower-order terms (e.g., main effects) were not biased due to the inclusion of higher-order terms (e.g., interactions). Using the same power estimation approach, the hypothesized 2 * 2 * 2 * 3 interaction (group, age group, session, and cue type) had a power of 0.98 in our sample.

Results

Learning

Averaged F0-based VD thresholds for the training groups during the training sessions are detailed in Appendix 1, Table 1. A full model summary of the statistical analysis employed to assess learning is shown in Table 2. Results revealed a significant effect for age-group, signifying better thresholds among younger participants. In addition, there was a significant linear effect for session and block, indicating
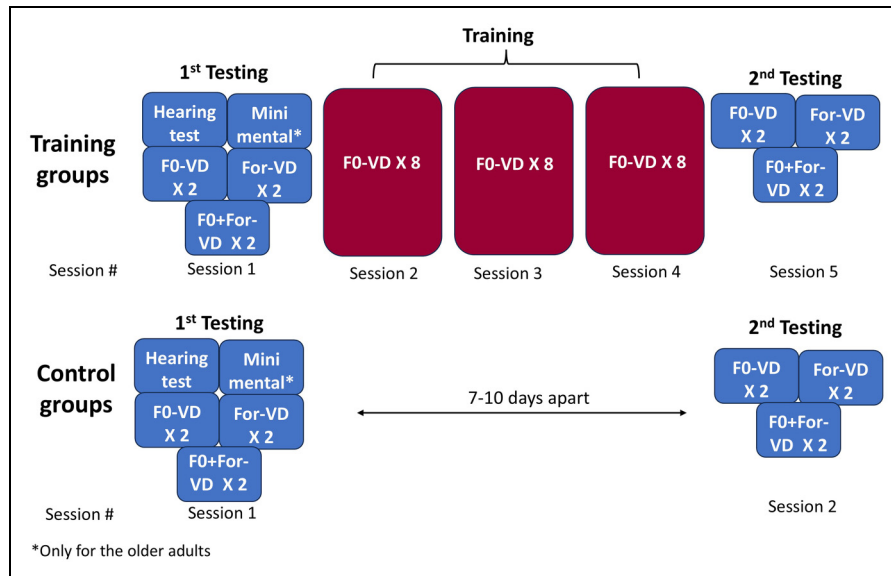
Figure 2. Study Design Outlining the Sessions Conducted by the Training and Control Groups, Each Including 12 Young Adults and 12 older Adults Training Groups Underwent Five Sessions (two testing and three training sessions), Separated by 1–3 days. Control groups Participated in two Testing Sessions only, with Sessions Spaced 7–10 days Apart. *Note.* VD = Voice Discrimination, F0 = Fundamental Frequency, For = Formants.

Table 2. A Model Summary of the Three-Level Multilevel Model (MLM) Employed to Assess Learning.

Parameter	B	SE	t	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Intercept	0.187	0.045	4.110	.000	0.093	0.281
AgeGroup	0.204	0.091	2.247	.035	0.016	0.392
Session	−0.057	0.016	−3.582	.002	−0.091	−0.024
Block	−0.014	0.004	−3.742	.001	−0.022	−0.007
*AgeGroup * Session*	−0.004	0.032	−.117	.908	−0.071	0.064
*AgeGroup * Block*	−0.005	0.008	−.641	.525	−0.020	0.010
*Session * Block*	0.002	0.005	.481	.633	−0.007	0.012
*AgeGroup * Session * Block*	0.059	0.150	.395	.693	−0.236	0.355

between-session (Figure 3A) and within-session learning (Figure 3B). No significant interactions were observed, suggesting comparable learning trajectories between the groups.

Generalization

Averaged VD thresholds based on either F0-only, formant-only or combined cues during the first and second testing sessions are outlined for the trained and control groups in Appendix 1, Table 2. A full model summary of the statistical analysis employed to assess generalization is shown in Table 3.

The results revealed a significant block effect, with no significant block interactions ($p > .05$), indicating rapid improvement between the first and second block across groups and cue types. A significant group * age-group * cue type * session interaction was observed. To further examine this effect,

we analyzed conditional interactions separately for each cue type. For the F0-only (Figure 4A), the analysis revealed a significant effect for age-group [$B = 0.12$, $SE = 0.03$, $p < .001$], with older adults performing worse. A significant effect was also found for session [$B = −.07$, $SE = 0.01$, $p < .001$], along with a significant session * group interaction [$B = 0.03$, $SE = 0.01$, $p = .007$], indicating significant improvement from the first to the second testing sessions only for the training groups [$B = −0.09$, $SE = 0.02$, $p < .001$]. No other interactions reached significance ($p > .05$), suggesting a similar magnitude of improvement in the F0-only condition for both young and older adults. Furthermore, visual inspection of the data suggested no significant difference between the F0 performance of the young training group in the first testing session and that of the older adults training group in the second testing session. To statistically assess this, a Bayesian analysis was conducted on the F0 performance scores from these two
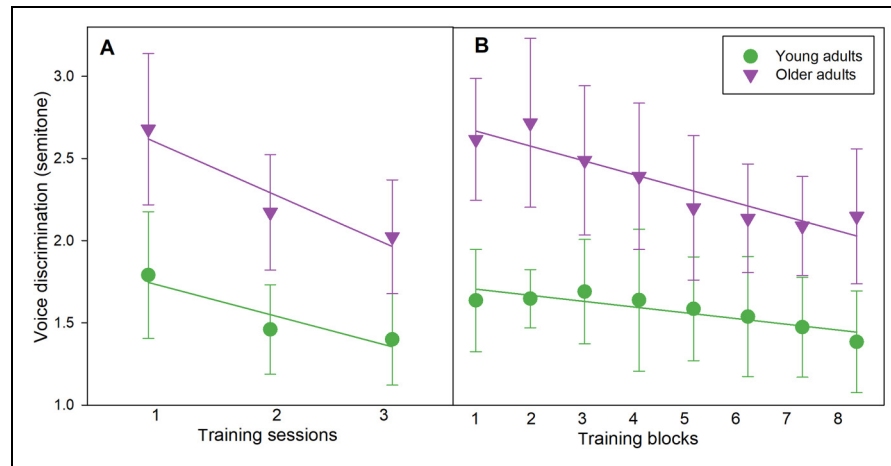
Figure 3. Mean Voice Discrimination Thresholds (± Standard error) Based on Fundamental Frequency (F0) Cues for the Young (n = 12) and Older (n = 12) Training Groups, Showing a Significant Effect for (A) Training Sessions (collapsed across training blocks), and (B) Training Blocks (collapsed across training sessions).

sessions. A Bayes Factor (BF01) of 3 or higher indicates that the null hypothesis (H0) is favored over the alternative hypothesis (H1), with BF01 = 3 corresponding to a 5% criterion in hypothesis testing. The analysis yielded a BF01 = 3.35, supporting the no-difference hypothesis. This suggests that three training sessions were sufficient for the older adults to reach the VD performance level of the young adults at their naïve baseline.

For the formant-only (Figure 4B), the analysis revealed significant effects for age-group [$B = 0.16$, $SE = 0.02$, $p < .001$] and session [$B = -0.04$, $SE = 0.02$, $p = .008$]. Significant interactions were found for group * session [$B = 0.03$, $SE = 0.03$, $p = .036$] age-group * session [$B = 0.04$, $SE = 0.02$, $p = .024$] and group * age-group * session [$B = -0.04$, $SE = 0.02$, $p = .009$]. We further broke down the tripled interaction using conditional interaction analysis. Results revealed significant improvement between the first and second testing sessions only for the young adults who received training [$B = -0.10$, $SE = 0.03$, $p < .001$].

For the F0 + formant (Figure 4C), the analysis revealed significant effects for group [$B = 0.08$, $SE = 0.03$, $p = .027$], with poorer thresholds in the control groups; for age-group [$B = 0.19$, $SE = 0.03$, $p < .001$], with worse performance among older adults; and for session [$B = -0.09$, $SE = 0.02$, $p < .001$], showing significant improvements between the first and second testing sessions. No interactions reached significance ($p > .05$), indicating a similar magnitude of improvement in the F0 + formant condition, across both the training and control groups and both age groups.

Discussion

The objective of the present study was to enhance our understanding of the constraints influencing auditory learning and generalization in the elderly population. This was achieved

by employing a basic psychoacoustic task of VD. Young and older adults underwent a three-session VD training regimen focused on F0 cues. The study assessed the efficiency of learning by examining the time course of learning both within and between sessions, as well as the extent to which the acquired learning generalized across different voice cues. The principal findings can be summarized as follows: (a) Although older adults exhibited poorer hearing thresholds and inferior VD performance compared to young adults, both groups demonstrated substantial improvements in VD based on F0 cues following training. Furthermore, no statistically significant differences were found in the learning trajectories or overall gains between the two age groups, indicating similar learning efficiency; (b) despite similar learning gains with the trained VD condition (F0-based), only the younger training group showed significant improvements in the formant-based VD condition following training, suggesting greater generalization of the learning gains in younger compared to older adults; (c) learning gains for VD based on combined F0 + formant cues did not differ between the training and control groups, indicating no added advantage of single-cue training for combined-cue VD. This result reveals limitations in the generalization for both young and older adults.

Our finding that, as a group, the older adults exhibited poorer VD thresholds based on F0-only, formant-only or combined F0 + formant cues compared to the young adults aligns with previous research demonstrating age-related impairments in F0 and formant perception (Anderson et al., 2021; Chintanpalli et al., 2016; Goupell et al., 2017; Souza et al., 2011; Vongpaisal & Pichora-Fuller, 2007) as well as inferior VD performance based on these cues in older populations (Zaltz & Kishon-Rabin, 2022). This outcome may be influenced by the potential confounding effect of hearing sensitivity in the older adult group. While their hearing thresholds generally aligned with typical age-related patterns, a notable

Table 3. A Full Model Summary of the Five-Level Multilevel Model (MLM) Employed to Assess Generalization.

Parameter	B	SE	t	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Intercept	0.30	0.08	3.90	<.001	0.15	0.46
Group	.12	0.05	−2.47	.017	−0.22	−0.02
AgeGroup	.32	0.05	6.42	<.001	−0.42	−0.22
Cue (0)	.25	0.02	12.05	<.001	−0.29	−0.21
Cue (1)	−.10	0.02	5.08	<.001	0.06	0.15
Session	−.13	0.02	7.86	<.001	0.10	0.17
Block	−.04	0.02	2.52	.012	0.01	0.08
*Group * AgeGroup*	−0.37	0.16	−2.38	.021	−0.68	−0.06
*Cue (1) * Group*	−0.29	0.13	−2.19	.033	−0.55	−0.02
*Session * Group*	0.00	0.10	−0.04	.968	−0.20	0.20
*Session * AgeGroup*	−0.04	0.10	−0.42	.676	−0.24	0.16
*Block * AgeGroup*	0.00	0.09	−0.04	.967	−0.18	0.17
*Cue (1) * Block*	−0.10	0.11	−0.98	.333	−0.31	0.11
*Session * Block*	0.05	0.08	0.58	.567	−0.12	0.21
*Session * Group * AgeGroup*	0.34	0.14	2.39	.021	0.05	0.62
*Block * Group * AgeGroup*	0.06	0.13	0.46	.645	−0.19	0.31
*Cue (0) * Session * Group*	0.03	0.12	0.28	.782	−0.20	0.27
*Cue (1) * Session * Group*	0.09	0.14	0.65	.522	−0.19	0.37
*Cue (0) * Block * Group*	−0.01	0.12	−0.10	.920	−0.25	0.23
*Cue (1) * Block * Group*	0.02	0.15	0.13	.900	−0.28	0.32
*Session * Block * Group*	−0.08	0.12	−0.65	.517	−0.31	0.16
*Cue (0) * Session * AgeGroup*	0.02	0.12	0.14	.888	−0.22	0.25
*Cue (1) * Session * AgeGroup*	0.16	0.14	1.13	.263	−0.12	0.43
*Cue (0) * Block * AgeGroup*	−0.06	0.12	−0.51	.611	−0.30	0.18
*Cue (1) * Block * AgeGroup*	0.19	0.15	1.28	.207	−0.11	0.49
*Session * Block * AgeGroup*	0.04	0.12	0.37	.714	−0.19	0.28
*Cue (0) * Session * Block*	−0.12	0.11	−1.08	.285	−0.34	0.10
*Cue (1) * Session * Block*	0.01	0.12	0.11	.917	−0.22	0.25
*Cue (0) * Session * Group * AgeGroup*	−0.22	0.17	−1.32	.193	−0.55	0.11
*Cue (1) * Session * Group * AgeGroup*	−0.42	0.19	−2.15	.036	−0.81	−0.03
*Cue (0) * Block * Group * AgeGroup*	0.04	0.17	0.22	.824	−0.30	0.37
*Cue (1) * Block * Group * AgeGroup*	−0.27	0.21	−1.31	.197	−0.70	0.15
*Session * Block * Group * AgeGroup*	−0.02	0.17	−0.09	.928	−0.35	0.32
*Cue (0) * Session * Block * Group*	0.12	0.15	0.81	.424	−0.19	0.43
*Cue (1) * Session * Block * Group*	0.17	0.17	1.04	.302	−0.16	0.50
*Cue (0) * Session * Block * AgeGroup*	0.07	0.15	0.48	.631	−0.24	0.38
*Cue (1) * Session * Block * AgeGroup*	−0.17	0.17	−1.02	.315	−0.50	0.16
*Cue (0) * Session * Block * Group * AgeGroup*	−0.05	0.22	−0.21	.837	−0.48	0.39
*Cue (1) * Session * Block * Group * AgeGroup*	0.13	0.233	0.569	.572	−0.34	0.60

Note. All variables were dummy coded. Group represents the effect of training vs. control (reference group); AgeGroup represents the effect of young adults vs. older adults (reference group); Cue 0 (F0 + For) and Cue 1 (F0-only) represent their effects compared to Cue 2 (formant-only; reference category); Session represents the effect of the first session vs. the second session (reference category); Block represents the effect of the first block vs. the second block (reference category).

prevalence of high-frequency hearing loss was observed, with some individuals exhibiting thresholds in the moderately severe to severe range. Such age-related hearing loss (presbycusis) is associated with impaired auditory processing, including reduced spectral resolution due to outer hair cell loss, broadened auditory filters, strial dysfunction, or cochlear synaptopathy, and/or decreased temporal resolution from reduced neural synchrony (Anderson & Karawani, 2020). Previous research suggests that effective F0 and formant frequency coding relies on efficient utilization of both temporal and spectral information (Fu et al., 2004; Oxenham, 2008; Xu & Pfingst,

2008). Therefore, deficits in spectro-temporal processing, such as impairments in periodicity and fine-structure perception (Souza et al., 2011), may account for the reduced VD performance observed in older adults. Alternatively, the poorer VD performance in older adults may partly reflect age-related declines in cognitive abilities, such as working memory and executive control (Harada et al., 2013; Mitchell et al., 2000) which may be critical for attending to the relevant acoustic cue for VD and retaining this information long enough to facilitate decision-making. However, all older adults in the current study passed the MMSE screening. Furthermore, if
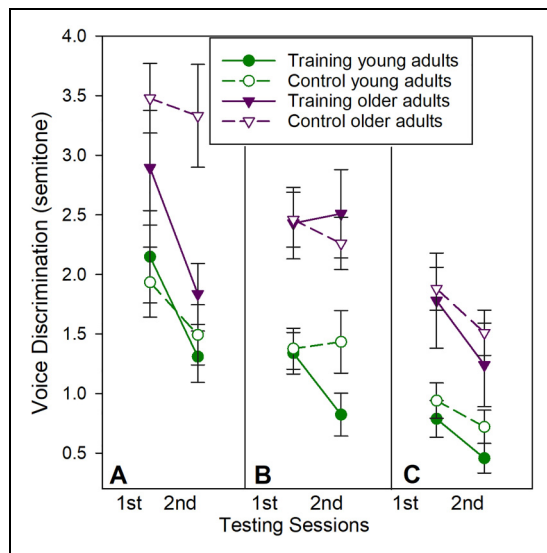
Figure 4. Mean Voice Discrimination Thresholds (± Standard error) Based on (A) Fundamental Frequency (F0) Cues, (B) Formant Cues, and (C) Combined F0 + formant Cues, Shown Separately for the Training (n = 24) and Control (n = 24) groups, as well as for Young (n = 12 per group) and Older Adults (n = 12 per group), Across the two Testing Sessions.

cognitive decline played a major role, it would also be expected to affect learning. Yet, in the present study, older and younger adults demonstrated comparable learning outcomes, making this explanation less likely.

Notably, although older adults exhibited poor hearing thresholds and overall VD performance, training with VD based on F0-only cues led to significant improvements, with both the time course and magnitude of their learning closely matching those of the young adults. The lack of significant differences in learning trajectories and overall gains between the two groups suggests that the fundamental mechanisms supporting auditory skill learning remain intact in older adults. This aligns with previous studies showing that older adults can benefit from auditory training despite initially lower performance relative to younger adults (Humes et al., 2014; Kishon-Rabin et al., 2013; Manheim et al., 2018). These findings support the notion that auditory perceptual learning is achievable even as sensory and cognitive capacities decline with age (Anderson & Karawani, 2020; Jayakody et al., 2018), provided that training is appropriately structured.

Moreover, the rapid improvements in VD performance observed across cues in older adults between the first and second blocks of the initial and final testing sessions, mirroring the young adults' rapid learning, suggest preserved adaptation mechanisms. Similar rapid learning effects have been documented in young adults during speech perception tasks involving degraded speech (e.g., noise-vocoded or accented speech [Banai et al., 2022; Borrie et al., 2012; Davis et al., 2005; Gordon-Salant et al., 2010]), as well as in a VD task

(Zaltz, 2024). While aging may negatively affect rapid learning for some types of complex speech, such as TCS (Manheim et al., 2018), studies have shown that older adults with hearing impairments exhibit comparable rapid learning effects to younger adults following repeated exposure to stimuli that are less acoustically demanding, such as accented speech (Gordon-Salant et al., 2010). Fast learning is thought to arise from top-down tuning and adaptation processes that facilitate effective task-solving strategies and reduce response bias (Hauptmann et al., 2005; Hauptmann & Karni, 2002), with recent evidence also suggesting involvement of early stages of perceptual learning (Banai et al., 2022). These mechanisms were suggested to contribute to enhanced speech understanding in difficult listening situations, allowing listeners to quickly adapt to changing acoustic environments (Banai et al., 2022). The current findings, indicating rapid VD improvements in older adults, suggest that brief, voice-targeted training has the potential to enhance the perception of basic acoustic features, which may, in turn, support better performance in challenging listening environments, such as noisy settings, for older adults.

Our finding that three VD training sessions with F0 cues were sufficient for older adults to "close the gap" and reach the naive performance level of the younger group further underscores the efficiency of auditory skill learning in this age group. Similarly, Kishon-Rabin et al. reported that four training sessions were enough for older adults to achieve gap detection performance comparable to the initial (naïve) performance of young adults (Kishon-Rabin et al., 2013). The fact that just a few sessions can bridge the performance gap across various psychoacoustic tasks suggests that the observed age-related differences may primarily result from declines in higher-order cognitive processes, such as auditory attention and decision-making, which are most impacted in the initial phases of learning (Censor, 2013; Karni, 1996). In line with this reasoning, the three-day training in the present study may have enhanced focused attention, enabling better processing of the most informative neural channels, reducing internal noise (Jones et al., 2013), and helping establish efficient task-specific routines (Karni, 1996; Ortiz & Wright, 2010; Wright & Zhang, 2009). However, it is also possible that, alongside these top-down "task learning" processes, lower-level neural changes may have taken place (Hawkey et al., 2004), improving fine-temporal coding deficits associated with aging and, consequently, enhancing F0 perception in older adults. Further research, incorporating a broader range of psychoacoustic tasks, is needed to determine if these age-related "gaps" can consistently be closed within a few training sessions, as suggested by our findings. Such evidence could have significant implications for developing clinical interventions and training protocols aimed at enhancing auditory skills in older populations.

While the study confirmed that both groups showed significant learning gains with the F0-only cues, significant improvement with formant-only cues was observed only in the

young training group, indicating a broader scope of learning transfer compared to older participants. A review of the literature reveals that only a limited number of studies have evaluated the generalization of auditory training effects to untrained stimuli in older versus younger adults (Bieber & Gordon-Salant, 2021). Among these, studies on TCS learning show that older adults exhibit less generalization compared to younger adults, with performance improvements confined to the specific compression ratio (Peelle & Wingfield, 2005) and sentence content (Manheim et al., 2018; Peelle & Wingfield, 2005) used during training. Other studies found that all listener groups showed limited but comparable generalization of learning for non-native speech (Bieber & Gordon-Salant, 2017), while no generalization of the learning-gains was observed following noise-vocoded speech training, across groups (Sheldon et al., 2008). Manheim et al. (2018) suggested that older adults may prioritize higher-level semantic information over lower-level acoustic details during auditory processing, which could limit their ability to generalize learning. However, the present findings indicate that even in a basic VD task that requires listeners to focus on acoustic features rather than linguistic content, older adults show a reduced generalization compared to younger adults. This suggests that the difficulty may not stem from a focus on semantic information but rather from age-related declines in perceptual flexibility and auditory processing. Older adults may struggle to adapt to new or altered acoustic patterns due to inefficient auditory processing or declines in cognitive functions such as auditory attention and working memory. These challenges likely hinder their ability to shift focus and adjust perceptual strategies when confronted with different acoustic cues, thereby limiting their capacity to generalize learning effectively. Future studies should test this theory by examining the role of auditory attention and flexibility in learning generalization across various acoustic tasks in older adults.

Interestingly, VD training using the F0-only cue did not significantly improve VD performance when combining F0 and formant cues for either young or older adults, beyond the gains observed in the control groups, which are likely due to rapid learning mechanisms (Banai et al., 2022; Gordon-Salant et al., 2010; Holmes et al., 2021). This result is consistent with a recent study that found no generalization to combined cues after two sessions of VD training with either F0-only or formant-only cues in young adults (Zaltz, 2024). A possible explanation for this outcome may be the distinct processing mechanisms involved for each acoustic cue in VD. Specifically, formant coding primarily relies on spectral processing, while F0 coding depends mainly on temporal processing (Fu et al., 2004; Oxenham, 2008; Xu & Pfingst, 2008). Therefore, in a VD task using both formant and F0 cues, an integration of these coding mechanisms is required. Although this task may benefit from cue redundancy, as indicated by the better performance with combined cues compared to individual ones across groups, it also demands

higher-level integration beyond basic sensory processing. The lack of generalization to the combined-cue condition suggests that the neural adaptations resulting from single-cue training may not extend to the complex processing required for integrating multiple acoustic cues (Ahissar & Hochstein, 2004; Irvine, 2018). This finding underscores the specificity of perceptual learning and implies that to facilitate generalization, training protocols may need to incorporate multiple cues or conditions simultaneously, irrespective of age.

Limitations and Future Directions

One limitation of this study is its focus on F0-based training without incorporating variability in the training materials. This protocol was chosen to minimize potential interference from interleaved or consecutive training involving additional conditions (Banai et al., 2010; Maidment et al., 2015), but it may have limited the extent of generalization achieved. Future studies could explore a more varied training approach, alternating between different voice cues (e.g., F0, formants) or introducing varied listening environments, such as background noise, to determine whether such strategies might promote broader learning generalization without compromising the learning timeline. Additionally, incorporating neurophysiological measures, such as auditory-evoked potentials, could offer valuable insights into the neural changes underlying both the successes and limitations in generalization. Finally, caution is warranted in interpreting these results due to the relatively small cohort used in the present study and the differences in hearing sensitivity between the two older adult groups. Future research should aim to replicate these findings while ensuring that hearing thresholds are matched between training and control groups. Additionally, employing a larger sample size and including a broader range of age groups and populations, such as older adults who use hearing aids, would further enhance the applicability and generalizability of the current study's implications.

Conclusions

The novel findings of the present study carry significant implications for the potential of auditory training in older adults with age-related hearing loss. Although older adults showed similar learning efficiency to young adults in the trained condition—VD based on F0 cues, their limited ability to generalize this learning to an untrained voice cue suggests constraints likely associated with age-related declines in processing flexibility (Harada et al., 2013; Schneider, 2011). The present study's results align with the previous suggestions that older adults' neural reorganization following training may be more constrained to the specific conditions experienced during the training process (Manheim et al., 2018; Peelle & Wingfield, 2005). However, as neural mechanisms were not directly assessed in this study, this interpretation should be approached with caution. Further research is needed to develop

training protocols that facilitate broader generalization, especially for older populations, to maximize the potential benefits of auditory training in enhancing speech perception abilities in complex listening environments. Additionally, given the varying degrees of hearing impairment among the older adults in this study, future research should explore whether auditory training outcomes differ based on hearing profile and whether individuals with more severe hearing loss may derive distinct benefits or require tailored intervention approaches.

Acknowledgments

Ethical Considerations

The study received approval from Tel Aviv University Review Board (approval number: 0007524–2).

Author Contributions

Conceptualization, methodology, formal analysis, supervision, writing—original draft, writing—review &editing: Y.Z. Investigation, data curing, formal analysis & writing—review &editing: N. S. All authors have read and agreed to the published version of the manuscript.

Funding

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data that support the findings of this study are publicly available on the Open Science Framework (OSF) (Zaltz, 2025).

Informed Consent

Informed written consent was obtained from all participants.

References

Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, *25*(3), 736–740. https://doi.org/10.1037/a0020054

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1515), 285–299. https://doi.org/10.1098/rstb.2008.0253

Amitay, S., Zhang, Y.-X., Jones, P. R., & Moore, D. R. (2014). Perceptual learning: Top to bottom. *Vision Research*, *99*, 69–77. https://doi.org/10.1016/j.visres.2013.11.006

Anderson, S., Bieber, R., & Schloss, A. (2021). Peripheral deficits and phase-locking declines in aging adults. *Hearing Research*, *403*, 108188. https://doi.org/10.1016/j.heares.2021.108188

Anderson, S., DeVries, L., Smith, E., Goupell, M. J., & Gordon-Salant, S. (2022). Rate discrimination training may partially restore temporal processing abilities from age-related deficits. *Journal of the Association for Research in Otolaryngology: JARO*, *23*(6), 771–786. https://doi.org/10.1007/s10162-022-00859-x

Anderson, S., & Karawani, H. (2020). Objective evidence of temporal processing deficits in older adults. *Hearing Research*, *397*, 108053. https://doi.org/10.1016/j.heares.2020.108053

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(41), 14156–14164. https://doi.org/10.1523/JNEUROSCI.2176-12.2012

Anderson, S., Parbery-Clark, A., Yi, H. G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear & Hearing*, *32*(6), 750–757. https://doi.org/10.1097/AUD.0b013e31822229d3

ANSI/ASA S3.6-2018 - Specification for Audiometers. (n.d.). Retrieved August 19, 2023, from https://webstore.ansi.org/standards/asa/ansiasas32018.

Banai, K., Karawani, H., Lavie, L., & Lavner, Y. (2022). Rapid but specific perceptual learning partially explains individual differences in the recognition of challenging speech. *Scientific Reports*, *12*(1), 10011. https://doi.org/10.1038/s41598-022-14189-8

Banai, K., Ortiz, J. A., Oppenheimer, J. D., & Wright, B. A. (2010). Learning two things at once: Differential constraints on the acquisition and consolidation of perceptual learning. *Neuroscience*, *165*(2), 436–444. https://doi.org/10.1016/j.neuroscience.2009.10.060

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*(3), EL51–EL56. https://doi.org/10.1121/1.4942628

Biçer, A., Koelewijn, T., & Başkent, D. (2023). Short implicit voice training affects listening effort during a voice cue sensitivity task with vocoder-degraded speech. *Ear & Hearing*, *44*, 900–916. https://doi.org/10.1097/AUD.0000000000001335

Bieber, R. E., & Gordon-Salant, S. (2017). Adaptation to novel foreign-accented speech and retention of benefit following training: Influence of aging and hearing loss. *The Journal of the Acoustical Society of America*, *141*(4), 2800–2811. https://doi.org/10.1121/1.4980063

Bieber, R. E., & Gordon-Salant, S. (2021). Improving older adults' understanding of challenging speech: Auditory training, rapid adaptation and perceptual learning. *Hearing Research*, *402*, 108054. https://doi.org/10.1016/j.heares.2020.108054

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies.

Journal of Speech, Language, and Hearing Research, 55(1), 290–305. https://doi.org/10.1044/1092-4388(2011/10-0349)

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. Attention, Perception, & Psychophysics, 77(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

Carlyon, R. P., & Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? The Journal of the Acoustical Society of America, 95(6), 3541–3554. https://doi.org/10.1121/1.409971

Censor, N. (2013). Generalization of perceptual and motor learning: A causal link with memory encoding and consolidation? Neuroscience, 250, 201–207. https://doi.org/10.1016/j.neuroscience.2013.06.062

Chintanpalli, A., Ahlstrom, J. B., & Dubno, J. R. (2016). Effects of age and hearing loss on concurrent vowel identification. The Journal of the Acoustical Society of America, 140(6), 4142–4153. https://doi.org/10.1121/1.4968781

Craik, F., & Salthouse, T. (2000). Handbook of aging and cognition (2nd ed.). Lawrence Erlbaum.

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. The Journal of the Acoustical Society of America, 114(5), 2913–2922. https://doi.org/10.1121/1.1616924

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. Hearing Research, 229(1-2), 132–147. https://doi.org/10.1016/j.heares.2007.01.014

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. Journal of Experimental Psychology: General, 134(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? The Journal of the Acoustical Society of America, 145(1), 417–439. https://doi.org/10.1121/1.5087693

Engdahl, B., Tambs, K., Borchgrevink, H. M., & Hoffman, H. J. (2005). Screened and unscreened hearing threshold levels for the adult population: Results from the nord-trøndelag hearing loss study. International Journal of Audiology, 44(4), 213–230. https://doi.org/10.1080/14992020500057731

Fant, G. (1960). Acoustic Theory of speech production.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/bf03193146

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fostick, L., Taitelbaum-Swead, R., Kreitler, S., Zokraut, S., & Billig, M. (2020). Auditory training to improve speech perception and self-efficacy in aging adults. Journal of Speech, Language, and Hearing Research, 63(4), 1270–1281. https://doi.org/10.1044/2019_JSLHR-19-00355

Fu, Q.-J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. Journal of the Association for Research in Otolaryngology, 5(3), 253–260. https://doi.org/10.1007/s10162-004-4046-1

Gibson, E. J. (1963). Perceptual learning. Annual Review of Psychology, 14, 29–56. https://doi.org/10.1146/annurev.ps.14.020163.000333

Goldstone, R. L. (1998). Perceptual learning. Annual Review of Psychology, 49, 585–612. https://doi.org/10.1146/annurev.psych.49.1.585

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Schurman, J. (2010). Short-term adaptation to accented English by younger and older adults. The Journal of the Acoustical Society of America, 128(4), EL200–4. https://doi.org/10.1121/1.3486199

Goupell, M. J., Gaskins, C. R., Shader, M. J., Walter, E. P., Anderson, S., & Gordon-Salant, S. (2017). Age-Related differences in the processing of temporal envelope and spectral cues in a speech segment. Ear & Hearing, 38(6), e335–e342. https://doi.org/10.1097/AUD.0000000000000447

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal cognitive aging. Clinics in Geriatric Medicine, 29(4), 737–752. https://doi.org/10.1016/j.cger.2013.07.002

Hauptmann, B, & Karni, A. (2002). From primed to learn: The saturation of repetition priming and the induction of long-term memory. Cognitive Brain Research, 13(3), 313–322. https://doi.org/10.1016/s0926-6410(01)00124-0

Hauptmann, B., Reinhart, E., Brandt, S. A., & Karni, A. (2005). The predictive value of the leveling off of within session performance for procedural memory consolidation. Cognitive Brain Research, 24(2), 181–189. https://doi.org/10.1016/j.cogbrainres.2005.01.012

Hawkey, D. J. C., Amitay, S., & Moore, D. R. (2004). Early and rapid perceptual learning. Nature Neuroscience, 7(10), 1055–1056. https://doi.org/10.1038/nn1315

Hesseg, R. M., Gal, C., & Karni, A. (2016). Not quite there: Skill consolidation in training by doing or observing. Learning & Memory, 23(5), 189–194. https://doi.org/10.1101/lm.041228.115

Holmes, E., To, G., & Johnsrude, I. S. (2021). How long does it take for a voice to become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training. Psychological Science, 32(6), 903–915. https://doi.org/10.1177/0956797621991137

Humes, L. E., Kinney, D. L., Brown, S. E., Kiener, A. L., & Quigley, T. M. (2014). The effects of dosage and duration of auditory training for older adults with hearing impairment. The Journal of the Acoustical Society of America, 136(3), EL224–EL230. https://doi.org/10.1121/1.4890663

Irvine, D. R. F. (2018). Auditory perceptual learning and changes in the conceptualization of auditory cortex. Hearing Research, 366, 3–16. https://doi.org/10.1016/j.heares.2018.03.011

Jayakody, D. M. P., Friedland, P. L., Martins, R. N., & Sohrabi, H. R. (2018). Impact of aging on the auditory system and related cognitive functions: A narrative review. *Frontiers in Neuroscience*, *12*, 125. https://doi.org/10.3389/fnins.2018.00125

Jones, P. R., Moore, D. R., Amitay, S., & Shub, D. E. (2013). Reduction of internal noise in auditory perceptual learning. *The Journal of the Acoustical Society of America*, *133*(2), 970–981. https://doi.org/10.1121/1.4773864

Karawani, H., Bitan, T., Attias, J., & Banai, K. (2015). Auditory perceptual learning in adults with and without age-related hearing loss. *Frontiers in Psychology*, *6*, 2066. https://doi.org/10.3389/fpsyg.2015.02066

Karni, A. (1996). The acquisition of perceptual and motor skills: A memory system in the adult human cortex. *Cognitive Brain Research*, *5*(1-2), 39–48. https://doi.org/10.1016/s0926-6410(96)00039-0

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*(6443), 250–252. https://doi.org/10.1038/365250a0

Kishon-Rabin, L., Avivi-Reich, M., & Ari-Even Roth, D. (2013). Improved gap detection thresholds following auditory training: Evidence of auditory plasticity in older adults. *American Journal of Audiology*, *22*(2), 343–346. https://doi.org/10.1044/1059-0889(2013/12-0084)

Kishon-Rabin, L., Patael, S., Menahemi, M., & Amir, N. (2004). Are the perceptual effects of spectral smearing influenced by speaker gender? *Journal of Basic and Clinical Physiology and Pharmacology*, *15*(1-2), 41–56. https://doi.org/10.1515/jbcpp.2004.15.1-2.41

Koelewijn, T., Gaudrain, E., Shehab, T., Treczoks, T., & Başkent, D. (2023). The role of word content, sentence information, and vocoding for voice cue perception. *Journal of Speech, Language, and Hearing Research*, *66*(9), 3665–3676. https://doi.org/10.1044/2023_JSLHR-22-00491

Koelewijn, T., Gaudrain, E., Tamati, T., & Başkent, D. (2021). The effects of lexical content, acoustic and linguistic variability, and vocoding on voice cue perception. *The Journal of the Acoustical Society of America*, *150*(3), 1620–1634. https://doi.org/10.1121/10.0005938

Kreitewolf, J., Mathias, S. R., & von Kriegstein, K. (2017). Implicit talker training improves comprehension of auditory speech in noise. *Frontiers in Psychology*, *8*, 1584. https://doi.org/10.3389/fpsyg.2017.01584

Levin, M., & Zaltz, Y. (2023). Voice discrimination in quiet and in background noise by simulated and real cochlear implant users. *Journal of Speech, Language, and Hearing Research*, *66*, 1–18. https://doi.org/10.1044/2023_JSLHR-23-00019

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), 467–477. https://doi.org/10.1121/1.1912375

Lieberman, P., & Blumstein, S. E. (1988). Source-filter theory of speech production. In *Speech physiology, speech perception, and acoustic phonetics (Cambridge studies in speech science and communication* (pp. 34–50). Cambridge University Press.

Mackersie, C. L., Dewey, J., & Guthrie, L. A. (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *The Journal of the Acoustical Society of America*, *130*(2), 1006–1019. https://doi.org/10.1121/1.3605548

Maidment, D. W., Kang, H., Gill, E. C., & Amitay, S. (2015). Acquisition versus consolidation of auditory perceptual learning using mixed-training regimens. *Plos One*, *10*(3), e0121953. https://doi.org/10.1371/journal.pone.0121953

Manheim, M., Lavie, L., & Banai, K. (2018). Age, hearing, and the perceptual learning of rapid speech. *Trends in Hearing*, *22*, 2331216518778651. https://doi.org/10.1177/2331216518778651

Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., & D'Esposito, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging*, *15*(3), 527–541. https://doi.org/10.1037//0882-7974.15.3.527

Moore, B. C., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America*, *91*(5), 2881–2893. https://doi.org/10.1121/1.402925

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5-6), 453–467. https://doi.org/10.1016/0167-6393(90)90021-Z

Nambi, P. M. A., Sangamanatha, A. V., Vikas, M. D., Bhat, J. S., & Shama, K. (2016). Perception of spectral ripples and speech perception in noise by older adults. *Ageing International*, *41*(3), 283–297. https://doi.org/10.1007/s12126-016-9248-4

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Ortiz, J. A., & Wright, B. A. (2010). Differential rates of consolidation of conceptual and stimulus learning following training on an auditory skill. *Experimental Brain Research*, *201*(3), 441–451. https://doi.org/10.1007/s00221-009-2053-5

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, *12*(4), 316–331. https://doi.org/10.1177/1084713808325881

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1315–1330. https://doi.org/10.1037/0096-1523.31.6.1315

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in Amplification*, *10*(1), 29–59. https://doi.org/10.1177/108471380601000103

Purcell, D. W., John, S. M., Schneider, B. A., & Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *The Journal of the Acoustical Society of America*, *116*(6), 3581–3593. https://doi.org/10.1121/1.1798354

Sabin, A. T., Clark, C. A., Eddins, D. A., & Wright, B. A. (2013). Different patterns of perceptual learning on spectral modulation detection between older hearing-impaired and younger normal-hearing adults. *Journal of the Association for Research in*

Otolaryngology, *14*(2), 283–294. https://doi.org/10.1007/s10162-012-0363-y

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, *62*, 49–72. https://doi.org/10.1146/annurev.psych.121208.131643

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Schmiedt, R. A. (2010). The physiology of cochlear presbycusis. In S. Gordon-Salant, R. D. Frisina, A. N. Popper, & R. R. Fay (Eds.), *The Aging Auditory System* (pp. 9–38). Springer Science & Business Media.

Schneider, B. A. (2011). How age affects auditory-cognitive interactions in speech comprehension. *Audiology Research*, *1*(1), e10. https://doi.org/10.4081/audiores.2011.e10

Schneider, B., & Pichora-Fuller, M. (2000). Implications of perceptual deterioration of cognitive aging research. In F. Craik, T. Salthouse, & N. Mahwah (Eds.), *The handbook of aging and cognition* (pp. 155–121). Lawrence Erlbaum Associates Inc.

Schvartz-Leyzac, K. C., & Chatterjee, M. (2015). Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing. *The Journal of the Acoustical Society of America*, *138*(3), 1687–1695. https://doi.org/10.1121/1.4929938

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, *123*(1), 489–499. https://doi.org/10.1121/1.2783762

Shultz, D. (2015). When your voice betrays you. *Science*, *347*(6221), 494–494. https://doi.org/10.1126/science.347.6221.494

Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research*, *296*, 131–140. https://doi.org/10.1016/j.heares.2012.11.004

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

Smith, D. R. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the Acoustical Society of America*, *122*(6), 3628–3639. https://doi.org/10.1121/1.2799507

Souza, P., Arehart, K., Miller, C. W., & Muralimanohar, R. K. (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear & Hearing*, *32*(1), 75–83. https://doi.org/10.1097/AUD.0b013e3181eccfe9

Tun, P. A., Williams, V. A., Small, B. J., & Hafter, E. R. (2012). The effects of aging on auditory processing and cognition. *American Journal of Audiology*, *21*(2), 344–350. https://doi.org/10.1044/1059-0889(2012/12-0030)

Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *The Journal of the Acoustical Society of America*, *125*(2), 1114–1124. https://doi.org/10.1121/1.3050321

Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *The Journal of the Acoustical Society of America*, *130*(1), 429–439. https://doi.org/10.1121/1.3596462

Vongpaisal, T., & Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identification. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1139–1156. https://doi.org/10.1044/1092-4388(2007/079)

Wright, B. A., & Zhang, Y. (2009). A review of the generalization of auditory learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1515), 301–311. https://doi.org/10.1098/rstb.2008.0262

Xu, L., & Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hearing Research*, *242*(1-2), 132–140. https://doi.org/10.1016/j.heares.2007.12.010

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, *15*(1), 88–99. https://doi.org/10.1037//0882-7974.15.1.88

Zaltz, Y. (2023). The effect of stimulus type and testing method on talker discrimination of school-age children. *The Journal of the Acoustical Society of America*, *153*(5), 2611. https://doi.org/10.1121/10.0017999

Zaltz, Y. (2024). The impact of trained conditions on the generalization of learning gains following voice discrimination training. *Trends in Hearing*, *28*, 23312165241275896. https://doi.org/10.1177/23312165241275895

Zaltz, Y. (2025). Auditory learning and generalization in older adults: Evidence from voice discrimination training. OSF. https://osf.io/5j9sn/?view_only=20826516538e42a0aacffd687d3c7c4f

Zaltz, Y., Goldsworthy, R. L., Eisenberg, L. S., & Kishon-Rabin, L. (2020). Children with normal hearing are efficient users of fundamental frequency and vocal tract length cues for voice discrimination. *Ear & Hearing*, *41*(1), 182–193. https://doi.org/10.1097/AUD.0000000000000743

Zaltz, Y., Goldsworthy, R. L., Kishon-Rabin, L., & Eisenberg, L. S. (2018). Voice discrimination by adults with cochlear implants: The benefits of early implantation for vocal-tract length perception. *Journal of the Association for Research in Otolaryngology*, *19*(2), 193–209. https://doi.org/10.1007/s10162-017-0653-5

Zaltz, Y., & Kishon-Rabin, L. (2022). Difficulties experienced by older listeners in utilizing voice cues for speaker discrimination. *Frontiers in Psychology*, *13*, 797422. https://doi.org/10.3389/fpsyg.2022.797422

Appendix

Table A1. Averaged F0-Based Voice Discrimination (VD) Thresholds for the Training Groups Across the Three Training Sessions, with Standard Errors Shown in Parentheses.

		Blocks							
Age-group	Sessions	1	2	3	4	5	6	7	8
Young adults	1	1.62	1.79	1.75	2.00	1.89	2.03	1.71	1.52
		(0.34)	(0.34)	(0.32)	(0.53)	(0.49)	(0.49)	(0.34)	(0.39)
	2	1.63	1.57	1.61	1.34	1.42	1.42	1.44	1.24
		(0.31)	(0.28)	(0.27)	(0.27)	(0.28)	(0.29)	(0.31)	(0.32)
	3	1.58	1.50	1.62	1.49	1.37	1.09	1.21	1.34
		(0.27)	(0.30)	(0.34)	(0.42)	(0.22)	(0.14)	(0.21)	(0.28)
Older adults	1	3.21	3.20	3.00	2.85	2.44	2.35	2.25	2.12
		(0.37)	(0.58)	(0.53)	(0.53)	(0.50)	(0.36)	(0.35)	(0.40)
	2	2.26	2.52	2.26	1.96	2.40	1.96	1.87	2.16
		(0.34)	(0.49)	(0.38)	(0.25)	(0.47)	(0.32)	(0.18)	(0.34)
	3	2.17	2.21	2.02	2.19	1.60	1.96	2.02	2.02
		(0.30)	(0.37)	(0.36)	(0.44)	(0.23)	(0.29)	(0.32)	(0.47)

Table A2. Averaged Voice Discrimination (VD) Thresholds Based on F0-Only, Formant-Only, and Combined F0 + Formant Cues for the Training and Control Groups Across the two Testing Sessions, with Standard Errors Shown in Parentheses.

			F0-only		Formant-only		F0 + Formants	
Age-group	Group	Blocks/sessions	1	2	1	2	1	2
Young Adults	Training	1	2.31	1.99	1.43	1.25	0.84	0.73
			(0.43)	(0.36)	(0.17)	(0.18)	(0.14)	(0.17)
		2	1.17	1.45	0.90	0.75	0.48	0.44
			(0.19)	(0.24)	(0.18)	(0.19)	(0.13)	(0.12)
	Control	1	2.22	1.88	1.62	1.20	0.96	0.91
			(0.33)	(0.29)	(0.18)	(0.16)	(0.17)	(0.13)
		2	2.02	1.35	1.49	1.44	0.67	0.78
			(0.38)	(0.18)	(0.22)	(0.30)	(0.12)	(0.16)
Older Adults	Training	1	3.33	2.46	2.55	2.32	1.82	1.74
			(0.56)	(0.37)	(0.36)	(0.24)	(0.42)	(0.39)
		2	1.81	1.87	2.51	2.52	1.20	1.28
			(0.32)	(0.19)	(0.27)	(0.46)	(0.28)	(0.42)
	Control	1	3.53	3.42	2.75	2.17	1.79	1.97
			(0.31)	(0.28)	(0.24)	(0.19)	(0.14)	(0.21)
		2	3.17	3.50	2.45	2.07	1.60	1.41
			(0.42)	(0.46)	(0.27)	(0.15)	(0.22)	(0.16)