



Difficulties Experienced by Older Listeners in Utilizing Voice Cues for Speaker Discrimination

Yael Zaltz* and Liat Kishon-Rabin

Department of Communication Disorders, The Stanley Steyer School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

OPEN ACCESS

Edited by:

Leah Fostick,
Ariel University, Israel

Reviewed by:

Antje Heinrich,
The University of Manchester,
United Kingdom
Sandra Gordon-Salant,
University of Maryland,
College Park, United States

*Correspondence:

Yael Zaltz
yaelzaltz@gmail.com

Specialty section:

This article was submitted to
Perception Science,
a section of the journal
Frontiers in Psychology

Received: 18 October 2021

Accepted: 24 January 2022

Published: 03 March 2022

Citation:

Zaltz Y and Kishon-Rabin L (2022)
Difficulties Experienced by Older
Listeners in Utilizing Voice Cues for
Speaker Discrimination.
Front. Psychol. 13:797422.
doi: 10.3389/fpsyg.2022.797422

Human listeners are assumed to apply different strategies to improve speech recognition in background noise. Young listeners with normal hearing (NH), e.g., have been shown to follow the voice of a particular speaker based on the fundamental (F0) and formant frequencies, which are both influenced by the gender, age, and size of the speaker. However, the auditory and cognitive processes that underlie the extraction and discrimination of these voice cues across speakers may be subject to age-related decline. The present study aimed to examine the utilization of F0 and formant cues for voice discrimination (VD) in older adults with hearing expected for their age. Difference limens (DLs) for VD were estimated in 15 healthy older adults (65–78 years old) and 35 young adults (18–35 years old) using only F0 cues, only formant frequency cues, and a combination of F0 + formant frequencies. A three-alternative forced-choice paradigm with an adaptive-tracking threshold-seeking procedure was used. Wechsler backward digit span test was used as a measure of auditory working memory. Trail Making Test (TMT) was used to provide cognitive information reflecting a combined effect of processing speed, mental flexibility, and executive control abilities. The results showed that (a) the mean VD thresholds of the older adults were poorer than those of the young adults for all voice cues, although larger variability was observed among the older listeners; (b) both age groups found the formant cues more beneficial for VD, compared to the F0 cues, and the combined (F0 + formant) cues resulted in better thresholds, compared to each cue separately; (c) significant associations were found for the older adults in the combined F0 + formant condition between VD and TMT scores, and between VD and hearing sensitivity, supporting the notion that a decline with age in both top-down and bottom-up mechanisms may hamper the ability of older adults to discriminate between voices. The present findings suggest that older listeners may have difficulty following the voice of a specific speaker and thus implementing doing so as a strategy for listening amid noise. This may contribute to understanding their reported difficulty listening in adverse conditions.

Keywords: speaker discrimination, older adults, fundamental frequency, formant frequencies, voice discrimination

INTRODUCTION

Older people often find it extremely difficult to understand speech and converse in noisy environments, particularly when the noise includes several speakers (e.g., Pichora-Fuller, 1997). Such difficulties may limit their ability to participate in social, occupational, and educational activities, isolating them from their families and friends (Cacciatore et al., 1999; Arlinger, 2003; Pronk et al., 2011; Gopinath et al., 2012). Many studies have attempted to assess the contribution of different factors to the difficulties older adults experience in speech perception amid noise and found that they explain the results only in part. These include elevated hearing thresholds (e.g., Jayakody et al., 2018), reduced frequency and temporal resolution (e.g., Anderson and Karawani, 2020), and declining cognitive functioning (e.g., Pichora-Fuller and Singh, 2006; Schneider et al., 2010). One approach to further our understanding of the difficulties experienced by older adults in noisy conditions is to assess their ability to apply listening strategies that are known to assist younger adults in listening amid noisy backgrounds. One such strategy includes identifying and following the acoustic voice cues of a target speaker, such as the fundamental frequency (F0) and formant frequencies of their voice (e.g., Bronkhorst, 2015). Young adults have been shown to efficiently implement this strategy to segregate the relevant and irrelevant speakers (e.g., Bronkhorst, 2015), relying on efficient spectral (formants) and temporal (F0) processing of the speech signal (e.g., Fant, 1960; Lieberman and Blumstein, 1988; Carlyon and Shackleton, 1994; Fu et al., 2004; Oxenham, 2008; Xu and Pfingst, 2008). Given that spectro-temporal processes are known to degrade with age (e.g., Vongpaisal and Pichora-Fuller, 2007; Souza et al., 2011; Schwartz-Leyzac and Chatterjee, 2015; Chintanpalli et al., 2016; Goupell et al., 2017; Anderson et al., 2021) it may be difficult for older adults to take advantage of differences in F0 and/or formant information for talker segregation. Age-related cognitive decline in executive functions, including attention, inhibition, and working memory (e.g., Mitchell et al., 2000; Salthouse, 2000; Harada et al., 2013), may add to the difficulty in utilizing these acoustic cues for voice discrimination. Thus, the goal of the present study was to examine the use of F0 and formant cues for voice discrimination and to assess the contribution of sensory and cognitive factors to this discrimination in older adults with normal hearing for their age (NHA), compared to young adults.

Multi-talker situations are particularly challenging for speech understanding because they force the listener to cope with both energetic (Brungart, 2001; Ezzatian et al., 2012) and informational masking (Durlach et al., 2003; Hoehn et al., 2007; Ezzatian et al., 2012). Energetic masking occurs when the energy of the frequencies of the competing voices overlap those of the target voice, activating similar areas along the basilar membrane. Informational masking occurs because the competing speech may invoke related linguistic activity and/or divert attention from the target speech, interfering with the processing of the speech signal at higher linguistic or cognitive levels and making it difficult for the listener to focus on the auditory stream of interest and ignore non-relevant sounds (also known

as the “cocktail party” effect, e.g., Kattner and Ellermeier, 2020). Therefore, to function well in multi-talker situations, the listener has to efficiently utilize both bottom-up (e.g., spectral separation) and top-down (e.g., focused attention) mechanisms. However, spectral and temporal processing have been shown to degrade with age, even in listeners with audiograms within normal hearing (Moore and Peters, 1992; Vongpaisal and Pichora-Fuller, 2007; Schwartz-Leyzac and Chatterjee, 2015; Chintanpalli et al., 2016; Goupell et al., 2017). A negative age effect has also been reported for higher cognitive abilities, such as executive functions, including memory, attention, and inhibition (e.g., Mitchell et al., 2000; Salthouse, 2000; Harada et al., 2013). The poor speech-in-noise understanding of older NH adults may, therefore, be the result of varying degrees of decline in their peripheral, linguistic, and/or central and cognitive processing (e.g., Working Group on Speech Understanding and Aging, 1988; Humes et al., 2012; Tremblay et al., 2021).

One listening strategy that is assumed to assist in segregating the target voice from competing non-relevant sounds is to identify and track the acoustic characteristics of the voice of the speaker of interest (e.g., Bronkhorst, 2015). These characteristics include fundamental frequency (F0), which is influenced by the length, mass, and rate of vibration of the vocal cords, and formant frequencies (i.e., resonant frequencies of the vocal tract), which are influenced by the vocal tract length (VTL; e.g., Darwin et al., 2003; Vestergaard et al., 2009, 2011; Mackersie et al., 2011; Schwartz-Leyzac and Chatterjee, 2015; Başkent and Gaudrain, 2016). Both of these cues provide robust information regarding the speaker, such as age and gender, as well as idiosyncratic characteristics that are unique to that speaker and his/her personality, nearly as unique as our fingerprints according to some researchers (Shultz, 2015). Studies in young adults and children have shown that these listeners rely heavily on both types of cues (F0 and formant frequencies) to discriminate among (Zaltz et al., 2020) and segregate talkers (e.g., Darwin et al., 2003) as well as to identify the gender of a specific speaker (e.g., Smith and Patterson, 2005; Smith et al., 2007; Skuk and Schweinberger, 2014; Başkent and Gaudrain, 2016). Moreover, listeners who have difficulty perceiving differences in F0 and formant frequencies, such as cochlear implant users, showed reduced voice discrimination, which may explain, at least in part, their poor performance when listening amid noise (e.g., Gaudrain and Başkent, 2018; Zaltz et al., 2018).

Studies suggest that F0 coding relies primarily on efficient processing of the temporal envelope and/or of the temporal fine-structure cues of the signal, whereas formant coding primarily involves place coding of spectral energy peaks (e.g., Fant, 1960; Lieberman and Blumstein, 1988; Carlyon and Shackleton, 1994; Fu et al., 2004; Chatterjee and Peng, 2008; Oxenham, 2008; Xu and Pfingst, 2008). Previous studies have shown the effect of age on F0 discrimination (e.g., Moore and Peters, 1992; Vongpaisal and Pichora-Fuller, 2007; Souza et al., 2011; Anderson et al., 2021). It has been shown, for example, that older adults require twice the difference between F0s to reach similar accuracy in concurrent vowel identification with harmonic complexes and synthetic vowels, compared to their younger peers (Moore and Peters, 1992; Vongpaisal and

Pichora-Fuller, 2007). Electrophysiological studies have demonstrated pronounced reductions in phase locking in older adults, suggesting reduced neural synchrony among this population (e.g., Anderson et al., 2021). These findings were interpreted to reflect impaired periodicity coding in older listeners (Schvartz-Leyzac and Chatterjee, 2015), which, in turn, may negatively influence the utilization of F0 cues for talker discrimination in this population. Other studies have showed the effect of age on the utilization of formant changes for vowel identification (Vongpaisal and Pichora-Fuller, 2007; Chintanpalli et al., 2016; Goupell et al., 2017). However, no study, to our knowledge, has specifically investigated the ability to use changes in F0, formants and their combination for speaker discrimination.

Efficient utilization of the relevant acoustic cues for talker discrimination may also require complex cognitive processing, such as, attending to F0 and formant information of the different talkers, and storing this information in memory for decision making and for future reference. Therefore, in older adults, the reported age-related deterioration in the ability to focus attention on the relevant features of the stimulus while inhibiting the processing of non-relevant features (McDowd and Shaw, 2000; Schneider et al., 2007; Harada et al., 2013), may negatively affect their ability to discriminate between speakers based on specific voice cues. Similarly, decline in working memory processes with age, including poor short-term maintenance and manipulation of information during encoding (e.g., Mitchell et al., 2000), and/or general slowing of cognitive processes (Salthouse, 2000), may add to difficulties in utilizing F0 and formant cues for talker identification and stream segregation. Support for this hypothesis can be found in a recent study where the authors argued that poor talker identification amid noise in older adults may have been related to their difficulty to learn and store in memory the voice information associated with a particular speaker (Best et al., 2018). It is possible that a simpler task that examines the perception of F0 and formant frequencies *via* discrimination rather than *via* identification may better assess the utilization of voice cues in older adults. Thus, the present study aimed to assess the use of F0 cues alone, formant cues alone, and the combination of F0 and formants in a VD task in older adults with NHA, and to compare their performance to that of young adults with NH. In addition, because even a simple discrimination task may require attention and working memory capabilities, our second aim was to assess the contribution of higher cognitive abilities, specifically, executive control abilities and working memory, to VD performance.

MATERIALS AND METHODS

Participants

A total of 50 participants were recruited for the present study: 15 older adults (65–78 years, mean = 68.93 ± 3.63 years; median = 68) and 35 young adults (18–35 years, mean = 22.29 ± 3.16 years; median = 22). The VD results of 15 participants from the young-adult group were previously reported

(Zaltz et al., 2020). For the current study, we tested an additional 20 young adults to obtain a larger dataset for comparison. As no significant difference in age or test results was observed between the two groups of young adults ($p > 0.05$), their data were combined for all further analyses. The young adults had hearing sensitivity within the normal range in both ears, with pure-tone air conduction thresholds <20 dB HL at octave frequencies of 500–4,000 Hz (Ansi, 2018). For the older adults, eight participants had thresholds less or equal to ≤25 dB HL, five participants had thresholds less or equal to ≤40 dB HL, one had thresholds less or equal to ≤55 dB HL, and one had thresholds less or equal to ≤70 dB HL at octave frequencies of 500–4,000 Hz (**Figure 1**). Overall, pure-tone air conduction thresholds for the older adults were within the normal range for their age (Engdahl et al., 2005), with a pure-tone average across ears and four frequencies (PTA4) of less than 33 dB HL. None of the participants had any previous psychoacoustic experience in similar tasks, they had no known history of ear disease, and they had completed at least 12 years of formal education. All participants were fluent in Hebrew. Six of the older participants were native Hebrew speakers. The other nine older adults immigrated to Israel at a mean age of 24 (±11) years (range: 3–43 years), and thus were exposed to Hebrew for an average of 45 (±11) years (range of exposure: 29–72 years). Their mother tongues were French ($n = 4$), Arabic ($n = 3$), English ($n = 1$), and Romanian ($n = 1$). All older adults had cognitive ability levels within the normal range (Mini Mental State Examination score ≥ 27; based on the English version; Folstein et al., 1975), lived independently, and led an active life based on self-report. Informed consent was obtained from all participants. The study was approved by the Institutional Review Board of Ethics at Tel Aviv University.

Stimuli

The VD test included three shortened (3-word) sentences from the Hebrew version of the Matrix sentence test. All the sentences had a simple grammatical structure (noun, verb, adjective) from a vocabulary that is appropriate for 5 year olds, and were recorded by a native Hebrew female speaker (Bugannim et al., 2019), similar to Zaltz et al. (2020). Sentences were manipulated using a 13-point stimulus continuum, exponentially ranging in $\sqrt{2}$ steps from a change of -0.18 semitone to a change of -8 semitones. This manipulation was conducted in three separate dimensions: (1) F0, (2) formant frequencies, and (3) combined F0+formants (Zaltz et al., 2020). For a detailed explanation of the stimuli for the VD test, see **Appendix A**.

Voice Discrimination Test

A three-interval, three-alternative forced-choice procedure was used to estimate VD based on difference limen (DL) for F0 cues, formant cues, and combined F0+formant cues. A two-down one-up adaptive tracking procedure yielded DLs corresponding to a 70.7% detection threshold on the psychometric function (Levitt, 1971). Each trial consisted of two reference sentences and a comparison sentence, specified at a random interval. Inter-stimulus interval was 300 milliseconds. When a sentence

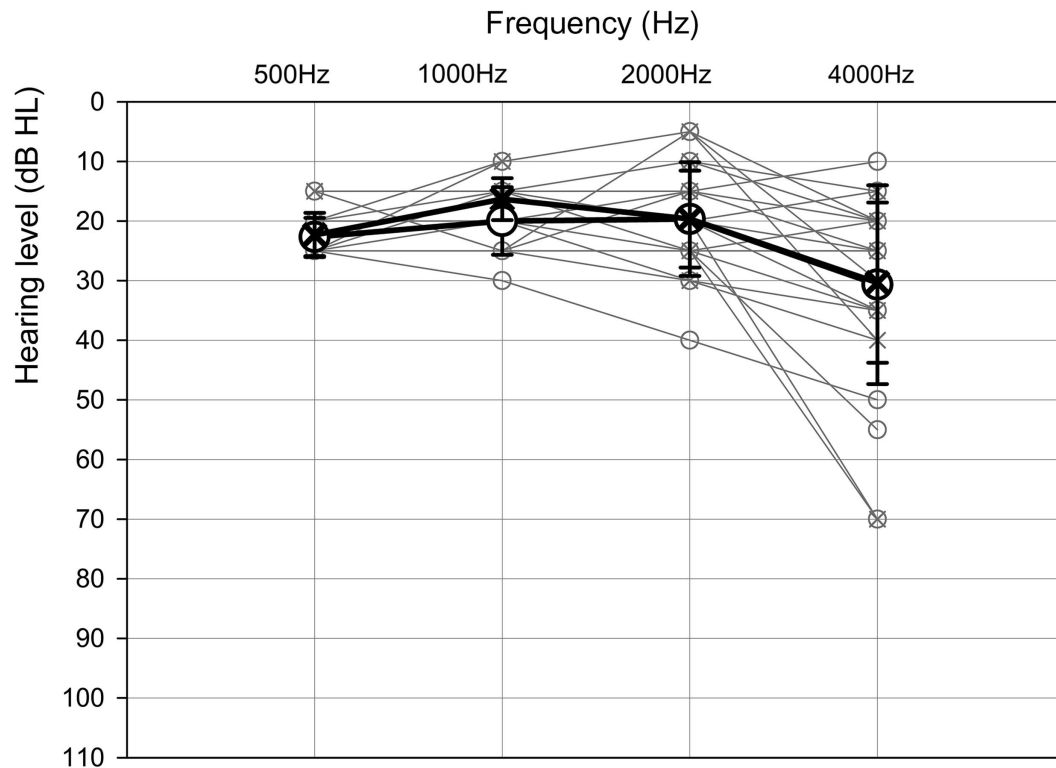


FIGURE 1 | Mean (thick black lines and symbols; ± 1 standard deviation) and individual (thin gray lines and symbols) hearing thresholds at 500Hz, 1,000Hz, 2000Hz, and 4,000Hz for the right (circle) and left (cross) ears for the older adults ($n = 15$).

was presented, one of the three squares on the PC monitor was highlighted to correspond to that sentence. The participant was instructed to select the sentence that “sounded different” by using the mouse to click on the corresponding square or telling the tester which sentence (1, 2, or 3) “sounded different.” The same sentence was used for each VD threshold, resulting in a change only in the tested acoustic cue (F0, formants, or F0+formants). No feedback was provided, and there was no time limit for responding. The difference between the stimuli was reduced by a factor of two until the first reversal and then reduced or increased by $\sqrt{2}$ until the sixth reversal. DLs were calculated as the arithmetic mean of the last four reversals.

Cognitive Tests

Cognitive capability was assessed only for older adults using the Hebrew version of the Mini Mental state examination test (based on the English version; Folstein et al., 1975). This test examined several mental functions, including orientation, memory, attention, naming, understanding of oral and written instructions, drawing, and writing. Overall, the test included 11 questions and lasted approximately 10 min. A score of 27 points or higher indicated cognitive ability within the normal range (Folstein et al., 1975).

Auditory working memory was assessed for all participants using a recorded version of the backward digit span subtest of the Wechsler Intelligence Scale (Wechsler, 1991). In the digit span test, the participants heard sequences of numbers

(e.g., 2, 6, 4, and 3) and were asked to repeat them verbally in the reverse order. The passing criterion to proceed to the next longer sequence was one successful repetition of a sequence of a specific length. The score represented the number of correctly repeated sequences.

A combined effect of executive control abilities, mental flexibility, and the perceptual speed of processing was assessed for all participants using the Trail Making test (TMT) part A and part B (e.g., Bowie and Harvey, 2006; Sánchez-Cubillo et al., 2009). In part A of the test, the participants were instructed to manually connect consecutive numbers from 1–24 by drawing a line as quickly and accurately as possible. In part B of the test, the participants were instructed to manually connect a set of 24 consecutive numbers and letters (e.g., 1A, 2B, 3C...) in sequential order as quickly as possible while maintaining accuracy. If a participant made an error, the tester corrected the response before moving on to the next dot. The scores for the TMT parts A and B represent the time taken for the participant to complete the test accurately (in seconds).

Study Design

All participants took part in a single testing session. At the beginning of the session, each participant from the older adult group was tested using the Mini Mental test. All participants performed three VD thresholds, one with each voice cue (F0, formants, F0+formants), with a different sentence presented for each cue. Voice cues and sentences

were randomized across participants. Before formal testing, each participant performed a short familiarization task with each of the voice cues to ensure that the task was understood. After the completion of VD testing, cognitive tests were conducted. Overall, the testing lasted approximately 45–50 min, including two short breaks of 5–8 min. The participants were not compensated for their time.

Apparatus

The stimuli were delivered using a laptop personal computer through an external sound card and a GSI-61 audiometer to both ears *via* THD-50 headphones at approximately 35–40 dB SL above individual PTA4. The testing session took place in a sound-treated single-walled room.

Data Analysis

All of the VD data were log-transformed for ANOVA to normalize the distribution of the residuals (Kolmogorov–Smirnov test: $p > 0.05$) and allow for parametric statistics. *Post hoc* analyses were conducted using Bonferroni corrections for multiple comparisons. Pearson's coefficient correlations were conducted on the raw data. Corrections for multiple testing (ANOVA and correlations) were applied using the False Discovery Rate method. Statistical analyses were conducted using the SPSS-20 software.

RESULTS

Box whisker plots of the VD thresholds based on F0, formant, and F0+formant cues for the young and older adult participants are shown in **Figure 2**. The older adults' thresholds were higher (i.e., worse) than those of the young adults for all voice cues. Two-way repeated measures ANOVA with Age as a between-subject variable and Cue (F0, formants, F0+formants) as a within-subject variable revealed a significant effect of Age [$F(1,48) = 27.060, p < 0.001, \eta^2 = 0.361$], with the young adults showing better VD thresholds ($M = 0.69 \pm 0.38$), compared to the older adults ($M = 1.40 \pm 0.85$). There was a significant effect of Cue [$F(2,48) = 46.056, p < 0.001, \eta^2 = 0.490$] with no significant Age*Cue interaction [$F(2,48) = 0.439, p = 0.646, \eta^2 = 0.009$]. A pairwise comparison revealed better thresholds across groups with the formant cues ($M = 0.90 \pm 0.63$), compared to the F0 cues ($M = 1.18 \pm 0.78; p = 0.011$), with the best thresholds achieved with the combined F0+formant cues ($M = 0.62 \pm 0.35; F0 > F0+formants, p < 0.001; Formants > F0+formants, p = 0.001$). Contrast analysis showed that both groups benefited similarly from the combined cues, compared to a single cue, with no significant Age*Cue interactions for F0, compared to F0+formants [$F(1,48) = 0.656, p = 0.422$], or for formants, compared to F0+formants [$F(1,48) = 0.01, p = 0.978$].

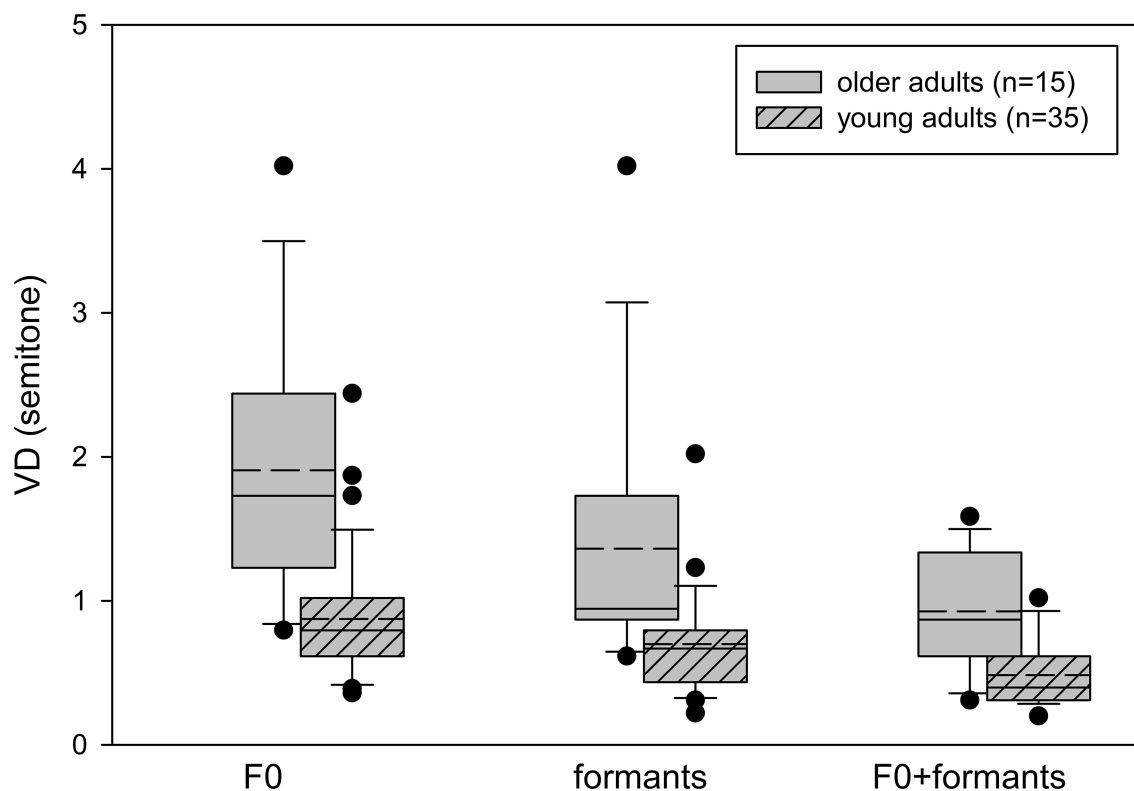


FIGURE 2 | Box plots for voice discrimination thresholds with the F0, formant, and combined F0+formant cues for the young adults ($n = 35$) and the older adults ($n = 15$). Box limits include the 25th to 75th percentile data, the solid line within the box represents the median, and the dashed line represents the mean. Bars extend to the 10th and 90th percentiles. Black dots represent outliers.

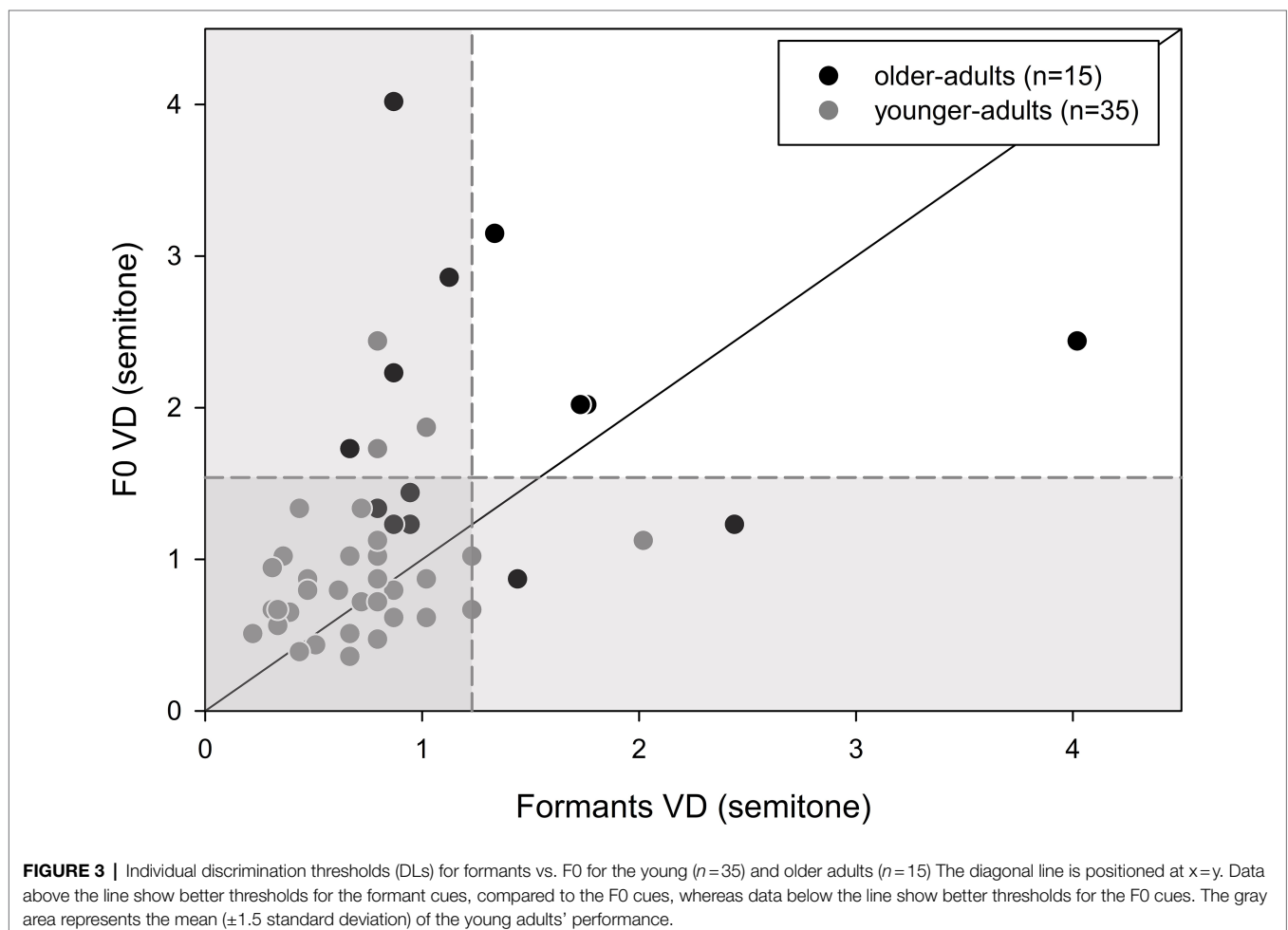
Mauchly's test of sphericity revealed larger between-subject variability in the VD results for older adults than for young adults ($p=0.019$). Thus, further analysis of the results was conducted at the individual level. Individual VD thresholds for F0 in relation to the formant cues for both young and older adults are shown in **Figure 3**. Also shown is the mean $VD \pm 1.5$ standard deviation of the young adults (gray areas). The majority of the participants demonstrated better discrimination for formant cues, compared to F0 cues, with a higher proportion of this preference in older adults ($n=12$, 80%) compared to younger adults ($n=20$, 57%). However, this difference was not significant (Fisher's exact test: $p=0.199$). Furthermore, although group analysis showed significantly worse thresholds for the older adults, individual analysis revealed that nine (60%), seven (47%), and five (33%) of the 15 older adults performed within the range of the young adults for the formant, F0, and formant+F0 conditions, respectively.

A comparison between the combined F0 + formant and single F0 or formant cues (**Figures 4A,B**) revealed that the majority of the participants benefited from the integration of two cues, compared to a single cue for VD, as is shown in **Figure 4**, by more data points below the diagonal line than above it. Specifically, for the older adults, 13 (87%) and 11 (73%)

participants benefited from the combined cues over F0 only or formants only, respectively. For the young adults, 31 (89%) and 24 (69%) participants benefited from the combined cues over F0 only and formants only, respectively. There was no significant difference in proportions between older and younger participants (Fisher's exact test: $p>0.05$). However, only six (40%) older adults performed within the range of the young adults with combined cues (**Figures 4A,B**).

Cognitive Abilities, Hearing Sensitivity, and Voice Discrimination

All the older adults exhibited cognitive performance in the normal range on each cognitive measure. The mean results of the cognitive tests for the young and older adults are shown in **Table 1**, along with the results of a one-way ANOVA comparing the two groups. Significantly better scores were achieved by the younger group in all cognitive tests. Pearson coefficient correlations were conducted separately for each group to test associations between the cognitive scores and age with the VD performance. For the older adults an additional Pearson coefficient correlation was conducted between hearing sensitivity (PTA4 averaged across both ears) and the VD performance.



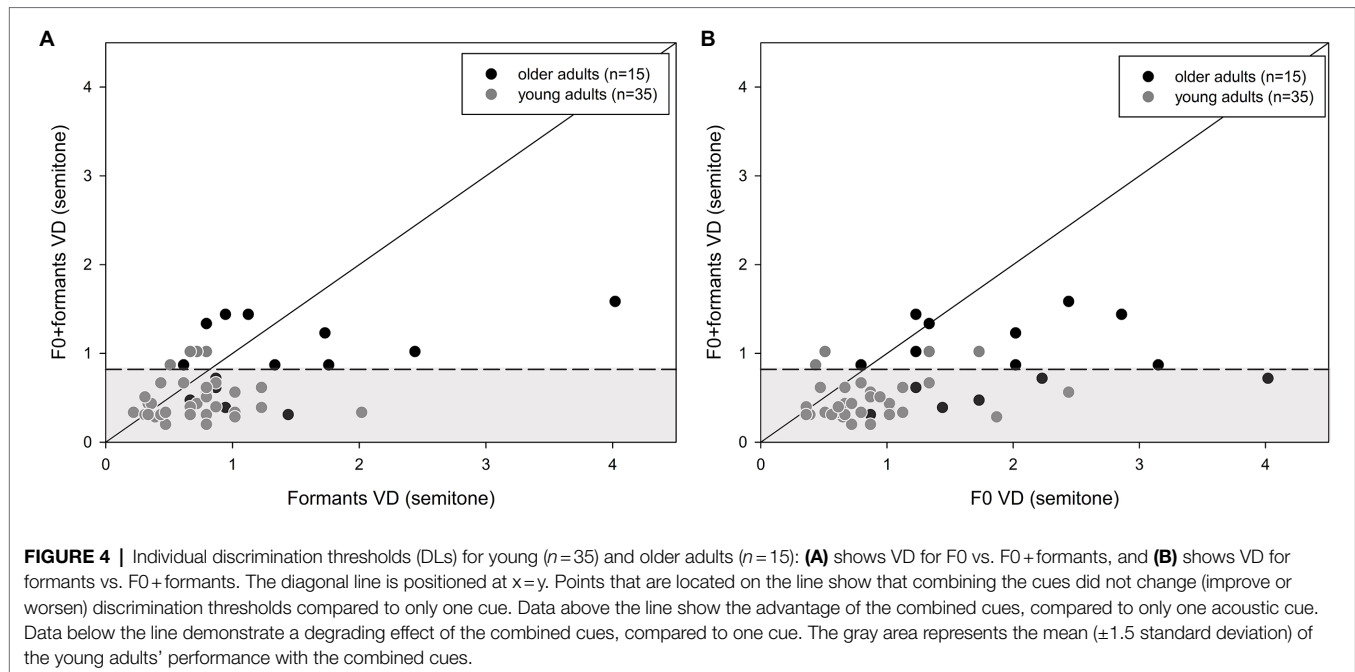


TABLE 1 | Mean performance for each of the cognitive tests for the young adults ($n=35$) and older adults ($n=15$).

	Young adults	Older adults	One-way ANOVA
Wechsler digit-span (#)	5.56 (± 2.02)	4.4 (± 1.12)	$F(1,48)=4.32$ $p=0.043$
TMT A (seconds)	22.46 (± 7.87)	38.8 (± 15.38)	$F(1,49)=23.92$ $p<0.01$
TMT B (seconds)	44.16 (12.10)	70.2 (± 24.38)	$F(1,48)=25.22$ $p<0.01$

In parentheses are \pm one standard deviation. Also shown are the results of the statistical analysis comparing the two groups of participants.

The full correlation results for these tests are shown in **Table 2**. For older adults, significant correlations were found in the combined F0 + formant testing condition between VD thresholds and TMT B scores [$r(13)=0.592$, $r^2=0.35$, $p=0.02$] (**Figure 5A**), and between VD thresholds and hearing sensitivity [$r(13)=0.594$, $r^2=0.35$, $p=0.02$] (**Figure 5B**). That is, shorter (i.e., better) TMT B times and better hearing sensitivity were associated with lower (i.e., better) VD thresholds when using F0 + formant cues. No significant association was found between TMT B scores and PTA4 ($p>0.05$) for this group. No significant associations were found between cognitive abilities and VD results for the young adults ($p>0.05$). The magnitudes of associations between F0 + formant VD and TMT B in the two groups (TMT_B X group interaction) was found statistically significant ($p=0.026$), adding $R^2=0.063$ to the proportion of explained variance in F0 + formant VD.

DISCUSSION

In the present study, we examined the utilization of F0 and formant cues for VD in 15 older adults (65–78 years old) as compared to 35 young adults (18–35 years old). The results

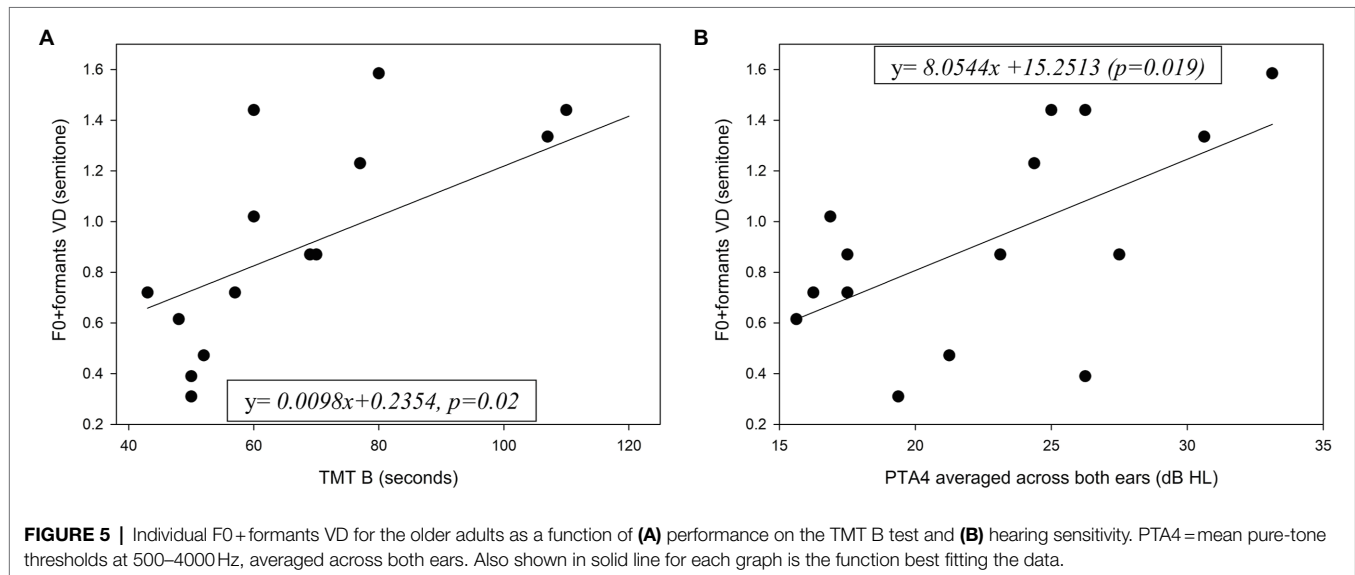
support the following findings: (a) Despite being generally high functioning, including hearing sensitivity expected for their age and normal cognitive function, the older adults as a group showed reduced ability to utilize voice cues (F0, formant frequencies, and the combined F0 + formants) for VD, compared to young adults; (b) However, individual analysis revealed large between-subject variability for the older adults, with 47–60% of them reaching VD performance that was within the range of young adults (mean ± 1.5 SD) with at least one acoustic cue; (c) An advantage was observed for formant cues, compared to F0 cues, for both young and older adults, in keeping with previous findings for children and young adults (Zaltz et al., 2020) and confirming that formants remain the reliable cue for VD throughout life; (d) Both the young and older participants benefitted more from the provision of the combined F0 + formant cues than from the provision of a single cue, supporting the hypothesis that older adults are capable of integrating spectral and temporal cues; (e) For the older adults, a combined effect of executive control abilities and speed of processing (as assume to be reflected by TMT B), and hearing sensitivity (PTA4) contributed significantly to the variance of VD in the combined F0 + formant condition, emphasizing the importance of basic requirements of bottom-up and top-down capabilities to allow for more advanced processing such as integration of acoustic cues.

Our major finding that as a group, older adults achieved larger (i.e., poorer) VD thresholds using F0 and formant frequency cues compared to young adults can be partly explained by age-related decline in cognitive abilities including a combined effect of processing speed, mental flexibility and executive control abilities. Normal aging is expected to include neurocognitive changes in working memory, attention, inhibition, and the speed of processing (e.g., Mitchell et al., 2000; Salthouse, 2000; Harada et al., 2013). These may be critical for focusing attention on the relevant acoustic

TABLE 2 | Pearson coefficient correlations between the cognitive scores, age, and hearing sensitivity (PTA4 averaged across both ears, available only for the older adults) and VD performance, separately for the young adults ($n=35$), and older adults ($n=15$).

	Young adults			Older adults		
	F0	Formants	F0 + formants	F0	Formants	F0 + formants
TMTA	-0.176 (0.311)	0.235 (0.174)	-0.137 (0.432)	0.197 (0.481)	-0.016 (0.955)	0.327 (0.235)
TMTB	-0.315 (0.070)	-0.043 (0.811)	-0.069 (0.700)	-0.092 (0.745)	0.170 (0.546)	0.592* (0.020)
Wechsler digit-span	0.193 (0.274)	0.240 (0.172)	0.224 (0.202)	0.095 (0.736)	0.129 (0.647)	0.054 (0.848)
Age	0.060 (0.731)	0.165 (0.345)	-0.085 (0.626)	-0.359 (0.189)	-0.017 (0.953)	0.081 (0.775)
PTA4	-	-	-	0.149 (0.596)	0.413 (0.126)	0.594* (0.020)

Significance level is shown in parentheses. *Correlation is significant at the 0.05 level. TMT, Trail Making Test.



cues for VD and storing this information in memory long enough for decision making. The significant positive association found for the older adults between the F0+formant VD and TMT B results may suggest that attention focusing, inhibition and perceptual speed of processing played an important role in VD.

Another explanation for the poorer VD performance of the older adults may stem from the significant association that was found between hearing sensitivity (PTA4) and DL for the F0 + formant condition in the older group. This association suggests that regardless of our attempts to compensate for loss in audibility by presenting the stimuli at 35–40 dB above the average individual hearing thresholds (PTA4), older adults with poorer audiograms showed inferior VD performance. This finding may support the notion that audibility is necessary but not sufficient for good auditory processing and that resolving capabilities in the spectral and temporal domains are needed (e.g., Schneider et al., 2010). Indeed, spectro-temporal processing has been suggested to decline with age as a result of numerous deficits, including a subclinical

loss of outer hair cells, broadened auditory filters, strial dysfunction, cochlear synaptopathy, and loss of neural synchrony (for a review, see Anderson and Karawani, 2020). Although temporal and spectral processing capabilities were not directly assessed in the current study, previous data suggest that F0 and formant frequency coding relies on efficient utilization of both temporal and spectral information (e.g., Fant, 1960; Lieberman and Blumstein, 1988; Carlyon and Shackleton, 1994; Fu et al., 2004; Chatterjee and Peng, 2008; Oxenham, 2008; Xu and Pfingst, 2008). Hence, deficits in the spectro-temporal processing, such as impairments in periodicity and fine-structure perception (Souza et al., 2011) may explain the poor performance in VD of older adults with greater loss of hearing sensitivity despite attempts to compensate for this loss of audibility.

One can argue that our finding of poorer VD for the older participants may have also been related to the fact that the VD test was conducted in Hebrew, which was the mother tongue for six of the 15 older participants, but for all the young

participants. This hypothesis is based on the notion that testing in a second language (L2) may have increased working memory demands, and in our case, influencing VD performance for those older participants. We believe, however, that this explanation is less likely, because the stimuli for the VD task in the current study included only three sentences, one for each VD assessment. Moreover, for each threshold estimation, the same sentence was used. The listeners were, therefore, required to identify the odd sentence based only on psychoacoustic perception, with no linguistic decisions to be made. Furthermore, all three sentences were taken from the Hebrew version of the Matrix sentence test which comprises words that are suitable for 5-year-old Hebrew speakers (Buganim et al., 2019), and included three words with a simple grammatical structure (noun, verb, adjective). Given that the nine older adults in our study whose mother tongue was not Hebrew were all fluent Hebrew speakers who were exposed to Hebrew for at least 29 years, with an average of 45 years, linguistic knowledge was probably not a contributing factor to the working memory demands of the VD task.

A second outcome of the present study is that the majority of older adults showed better discrimination thresholds with formant cues than F0 cues. This is similar to what was found in our young adults and in line with a recent study of school-age children (Zaltz et al., 2020). The advantage of formant cues for VD (in comparison to F0 cues) may be related to the amount of variation that each acoustic cue has in natural speech. While F0 varies significantly in terms of time for conveying prosodic information (standard deviation is approximately 3.7 semitones), formant frequencies are relatively constant over the duration of the vowel (standard deviation of approximately one semitone; Kania et al., 2006; Chuenwattanapranithi et al., 2008). The finding that formant frequencies remain a reliable cue for VD across the lifespan may also suggest faster degradation of temporal (F0) processing compared to spectral (formants) processing with age (Vongpaisal and Pichora-Fuller, 2007; Chintanpalli et al., 2016; Goupell et al., 2017).

Our finding that older adults benefit from combined cues to a similar degree as young adults supports the notion that older adults are able to integrate information from separately coded processes and use it to their advantage. Thus, despite the fact that formant frequencies were more easily accessed by older listeners, compared to F0 cues, this latter information was not ignored and was used to support the dominant channel of information for VD as long as their combined executive control and perceptual processing abilities were efficient (supported by the significant association found with TMT B; Bowie and Harvey, 2006; Sánchez-Cubillo et al., 2009).

Finally, our study showed a large between-subject variability in VD performance in the older group, with approximately half of the older adults showing a performance comparable to that of the young adults with at least one acoustic cue and 40% in the combined-cue condition. This variability was likely related to the considerable inter-individual variability reported for older adults in a range of auditory processing abilities and cognitive processes (e.g., Tun et al., 2012), as reflected by the significant associations found between VD performance and cognitive (TMT B) and sensory (PTA4) abilities. Given that

the efficient discrimination and integration of F0 and formants are essential for speech segregation (e.g., Darwin et al., 2003; Vestergaard et al., 2009), age-related declines in cognitive and sensory processing that negatively affect these abilities may lead to impaired speech perception in noise.

LIMITATIONS OF THE STUDY AND SUGGESTIONS FOR FUTURE STUDIES

Although we tested a relatively homogeneous group of older adults (all with normal audiograms for their age, normal mental capabilities, relatively high functioning, and similar audibility of the stimuli) and assessed several cognitive abilities in addition to VD performance, our findings provided only a partial explanation for the poor VD performance found in older adults. Psychoacoustic tests and more sensitive working memory tests may be included in future studies to further assess the relationship between temporal and spectral abilities and VD in older adults, and better reflect the correlations with age-related declines in cognitive abilities. Also, although the differences in audibility between the young and older adults were compensated by presenting the stimuli at approximately 35–40 dB SL above the individual PTA4, one cannot rule out the possibility that the inferior VD performance of the older adults was related to their poor hearing sensitivity. We believe that this issue did not have a major effect on the VD results because the acoustic cues for F0 and formants are mainly in the low-mid frequency range. However, a different or additional compensation method, such as frequency-shaped amplification on the stimuli (e.g., Souza et al., 2011) may have better accommodated the high-frequency threshold elevations of some of the older listeners. Finally, future studies may want to control the mother tongue of the participants in order to assure that there is no effect of the mono versus multilingualism status of the participants on VD performance.

CONCLUSION

The present study is the first to test the contribution of F0, formant, and the combination of F0 and formant cues to VD in older adults with no hearing loss other than loss of hearing sensitivity as expected to decline with age. The findings indicate that many of the older adults found the VD task more difficult than NH young adults, presenting poorer VD thresholds across conditions. The VD thresholds in the F0+formant condition for the older adults were associated with their audiograms, likely reflecting the importance of good bottom-up input and processing for efficient utilization of the acoustic voice cues of the talker. VD thresholds were also associated with TMT B scores, which are assumed to provide a general measure of executive control abilities, and perceptual processing, possibly reflecting their need to resort to cognitive resources in the presence of inefficient spectro-temporal processing (Schneider, 2011). These findings may explain the difficulties that older adults have in segregating talkers. The findings may also provide

a possible explanation for the difficulty older adults face when listening to speech in a multi-talker environment. However, these assumptions need to be confirmed in future studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board of Ethics at Tel Aviv University. The patients/participants provided their written informed consent to participate in this study.

REFERENCES

- Anderson, S., Bieber, R., and Schloss, A. (2021). Peripheral deficits and phase-locking declines in aging adults. *Hear. Res.* 403:108188. doi: 10.1016/j.heares.2021.108188
- Anderson, S., and Karawani, H. (2020). Objective evidence of temporal processing deficits in older adults. *Hear. Res.* 397:108053. doi: 10.1016/j.heares.2020.108053
- Ansi, S. (2018). American national standard specification for audiometers. New York: ANSI S3.6.
- Arlinger, S. (2003). Negative consequences of uncorrected hearing loss - A review. *Int. J. Audiol.* 42, 2S17–2S20. doi: 10.3109/14992020309074639
- Başkent, D., and Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *J. Acoust. Soc. Am.* 139, EL51–EL56. doi: 10.1121/1.4942628
- Best, V., Ahlstrom, J. B., Mason, C. R., Roverud, E., Perrachione, T. K., Kidd, G. Jr., et al. (2018). Talker identification: effects of masking, hearing loss, and age. *J. Acoust. Soc. Am.* 143, 1085–1092. doi: 10.1121/1.5024333
- Bowie, C. R., and Harvey, P. D. (2006). Administration and interpretation of the trail making test. *Nat. Protoc.* 1, 2277–2281. doi: 10.1038/nprot.2006.390
- Bronkhorst, A. W. (2015). The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Atten. Percept. Psychophys.* 77, 1465–1487. doi: 10.3758/s13414-015-0882-9
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696
- Bugannim, Y., Roth, D. A., Zechoval, D., and Kishon-Rabin, L. (2019). Training of speech perception in noise in pre-lingual hearing impaired adults with cochlear implants compared with normal hearing adults. *Otol. Neurotol.* 40, e316–e325. doi: 10.1097/MAO.0000000000002128
- Cacciatore, F., Napoli, C., Abete, P., Marciano, E., Triassi, M., and Rengo, F. (1999). Quality of life determinants and hearing function in an elderly population: Osservatorio Geriatrico Campano study group. *Gerontology* 45, 323–328. doi: 10.1159/000022113
- Carlyon, R. P., and Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J. Acoust. Soc. Am.* 95, 3541–3554. doi: 10.1121/1.409971
- Chatterjee, M., and Peng, S. C. (2008). Processing F0 with cochlear implants: modulation frequency discrimination and speech intonation recognition. *Hear. Res.* 235, 143–156. doi: 10.1016/j.heares.2007.11.004
- Chintanpalli, A., Ahlstrom, J. B., and Dubno, J. R. (2016). Effects of age and hearing loss on concurrent vowel identification. *J. Acoust. Soc. Am.* 140, 4142–4153. doi: 10.1121/1.4968781
- Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., and Maneewongvatana, S. (2008). Encoding emotions in speech with the size code. *Phonetica* 65, 210–230. doi: 10.1159/000192793

AUTHOR CONTRIBUTIONS

All authors contributed to this work to a significant extent, have read the article and agreed to submit it for publication after discussing the results and implications and commented on the article at all stages, and are, therefore, responsible for the reported research and have approved the final article as submitted.

ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of the following undergraduate students from the Department of Communication Disorders at Tel Aviv University for assisting in data collection: Elisheva Tibi, Sara Aviv, Alona Kandelevi Shamor, Ester Reuveni, Daniel Lex Rabinovitch, and Orpaz Shariki.

- Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114, 2913–2922. doi: 10.1121/1.1616924
- Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G. (2003). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J. Acoust. Soc. Am.* 114, 368–379. doi: 10.1121/1.1577562
- Engdahl, B., Tamba, K., Borchgrevink, H. M., and Hoffman, H. J. (2005). Screened and unscreened hearing threshold levels for the adult population: results from the Nord-Trøndelag hearing loss study. *Int. J. Audiol.* 44, 213–230. doi: 10.1080/14992020500057731
- Ezzatian, P., Li, L., Pichora-Fuller, M. K., and Schneider, B. A. (2012). The effect of energetic and informational masking on the time-course of stream segregation: evidence that streaming depends on vocal fine structure cues. *Lang. Cogn. Process.* 27, 1056–1088. doi: 10.1080/01690965.2011.591934
- Fant, G. (1960). *Acoustic Theory of Speech Perception*. The Hague: Mouton.
- Fitch, W. T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522.
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- Fu, Q. J., Chinchilla, S., and Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J. Assoc. Res. Otolaryngol.* 5, 253–260. doi: 10.1007/s10162-004-4046-1
- Gaudrain, E., and Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in Cochlear implant users. *Ear Hear.* 39, 226–237. doi: 10.1097/AUD.0000000000000480
- Gopinath, B., Schneider, J., McMahon, C. M., Teber, E., Leeder, S. R., and Mitchell, P. (2012). Severity of age-related hearing loss is associated with impaired activities of daily living. *Age Ageing* 41, 195–200. doi: 10.1093/ageing/afr155
- Goupell, M. J., Gaskins, C. R., Shader, M. J., Walter, E. P., Anderson, S., and Gordon-Salant, S. (2017). Age-related differences in the processing of temporal envelope and spectral cues in a speech segment. *Ear Hear.* 38, e335–e342. doi: 10.1097/AUD.0000000000000447
- Harada, C. N., Love, M. C. N., and Triebel, K. L. (2013). Normal cognitive aging. *Clin. Geriatr. Med.* 29, 737–752. doi: 10.1016/j.cger.2013.07.002
- Hoen, M., Meunier, F., Grataloup, C. L., Pellegrino, F., Grimault, N., Perrin, F., et al. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Comm.* 49, 905–916. doi: 10.1016/j.specom.2007.05.008
- Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., et al. (2012). Central presbycusis: a review and evaluation of the evidence. *J. Am. Acad. Audiol.* 23, 635–666. doi: 10.3766/jaaa.23.8.5

- Jayakody, D., Friedland, P. L., Martins, R. N., and Sohrabi, H. R. (2018). Impact of aging on the auditory system and related cognitive functions: a narrative review. *Front. Neurosci.* 12:125. doi: 10.3389/fnins.2018.00125
- Kania, R. E., Hartl, D. M., Hans, S., Maeda, S., Vaissiere, J., and Brasnu, D. F. (2006). Fundamental frequency histograms measured by electroglottography during speech: a pilot study for standardization. *J. Voice* 20, 18–24. doi: 10.1016/j.jvoice.2005.01.004
- Kattner, F., and Ellermeier, W. (2020). Distraction at the cocktail party: attenuation of the irrelevant speech effect after a training of auditory selective attention. *J. Exp. Psychol. Hum. Percept. Perform.* 46, 10–20. doi: 10.1037/xhp0000695
- Lammert, A. C., and Narayanan, S. S. (2015). On short-time estimation of vocal tract length from formant frequencies. *PLoS One* 10:e0132193.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375
- Lieberman, P., and Blumstein, S. E. (eds.) (1988). “Source-filter theory of speech production,” in *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge: Cambridge University Press), 34–50.
- Mackersie, C. L., Dewey, J., and Guthrie, L. A. (2011). Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *J. Acoust. Soc. Am.* 130, 1006–1019. doi: 10.1121/1.3605548
- McDowd, J., and Shaw, R. (2000). “Attention and aging: a functional perspective,” in *The Handbook of Aging and Cognition. 2nd Edn.* eds. F. Craik and T. Salthouse (NJ: Erlbaum), 221–292.
- Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., and D’Esposito, M. (2000). Aging and reflective processes of working memory: binding and test load deficits. *Psychol. Aging* 15, 527–541. doi: 10.1037//0882-7974.15.3.527
- Moore, B. C. J., and Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *J. Acoust. Soc. Am.* 91, 2881–2893. doi: 10.1121/1.402925
- Moulines, E., and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Commun.* 9, 453–467.
- Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends Amplif.* 12, 316–331. doi: 10.1177/1084713808325881
- Pichora-Fuller, M. K. (1997). Language comprehension in older adults. *J. Speech. Lang. Path. Audiol.* 2, 125–142.
- Pichora-Fuller, M. K., and Singh, G. (2006). Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiological rehabilitation. *Trends Amplif.* 10, 29–59. doi: 10.1177/108471380601000103
- Pronk, M., Deeg, D. J., Smits, C., Van Tilburg, T. G., Kuik, D. J., Festen, J. M., et al. (2011). Prospective effects of hearing status on loneliness and depression in older persons: identification of subgroups. *Int. J. Audiol.* 50, 887–896. doi: 10.3109/14992027.2011.599871
- Salthouse, T. A. (2000). Aging and measures of processing speed. *Biol. Psychol.* 54, 35–54. doi: 10.1016/S0301-0511(00)00052-1
- Sánchez-Cubillo, I., Periañez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J., et al. (2009). Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J. Int. Neuropsychol. Soc.* 15, 438–450. doi: 10.1017/S1355617709090626
- Schneider, B. A. (2011). How age affects auditory-cognitive interactions in speech comprehension. *Audiol. Res.* 1:e10. doi: 10.4081/audiore.2011.e10
- Schneider, B. A., Li, L., and Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.* 18, 559–572. doi: 10.3766/jaaa.18.7.4
- Schneider, B. A., Pichora-Fuller, K., and Daneman, M. (2010). “Effects of senescent changes in audition and cognition on spoken language comprehension,” in *The Aging Auditory System: Springer Handbook of Auditory Research.* eds. S. Gordon-Salant, R. D. Frisina and A. N. Popper (New York, NY: Springer), 167–210.
- Schvartz-Leyzac, K. C., and Chatterjee, M. (2015). Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing. *J. Acoust. Soc. Am.* 138, 1687–1695. doi: 10.1121/1.4929938
- Shultz, D. (2015). The privacy arms race. When your voice betrays you. *Science* 347:494. doi: 10.1126/science.347.6221.494
- Skuk, V. G., and Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J. Speech Lang. Hear. Res.* 57, 285–296. doi: 10.1044/1092-4388(2013)12-0314
- Smith, D. R., and Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *J. Acoust. Soc. Am.* 118, 3177–3186. doi: 10.1121/1.2047107
- Smith, D. R., Walters, T. C., and Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *J. Acoust. Soc. Am.* 122, 3628–3639. doi: 10.1121/1.2799507
- Souza, P., Arehart, K., Miller, C. W., and Muralimanohar, R. K. (2011). Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear.* 32, 75–83. doi: 10.1097/AUD.0b013e3181ecce9
- Tremblay, P., Brisson, V., and Deschamps, I. (2021). Brain aging and speech perception: effects of background noise and talker variability. *NeuroImage* 227:117675. doi: 10.1016/j.neuroimage.2020.117675
- Tun, P. A., Williams, V. A., Small, B. J., and Hafter, E. R. (2012). The effects of aging on auditory processing and cognition. *Am. J. Audiol.* 21, 344–350. doi: 10.1044/1059-0889(2012)12-0030
- Vestergaard, M. D., Fyson, N. R., and Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *J. Acoust. Soc. Am.* 125, 1114–1124. doi: 10.1121/1.3050321
- Vestergaard, M. D., Fyson, N. R., and Patterson, R. D. (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *J. Acoust. Soc. Am.* 130, 429–439. doi: 10.1121/1.3596462
- Vongpaisal, T., and Pichora-Fuller, M. K. (2007). Effect of age on F0 difference limen and concurrent vowel identification. *J. Speech. Lang. Hear. Res.* 50, 1139–1156. doi: 10.1044/1092-4388(2007)079
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—III*. San Antonio, TX: The Psychological Corporation.
- Working Group on Speech Understanding and Aging (1988). Speech understanding and aging. *J. Acoust. Soc. Am.* 83, 859–895. doi: 10.1121/1.395965
- Xu, L., and Pfungst, B. E. (2008). Spectral and temporal cues for speech recognition: implications for auditory prostheses. *Hear. Res.* 242, 132–140. doi: 10.1016/j.heares.2007.12.010
- Zaltz, Y., Goldsworthy, R. L., Eisenberg, L. S., and Kishon-Rabin, L. (2020). Children with normal hearing are efficient users of fundamental frequency and vocal tract length cues for voice discrimination. *Ear Hear.* 41, 182–193. doi: 10.1097/AUD.0000000000000743
- Zaltz, Y., Goldsworthy, R. L., Kishon-Rabin, L., and Eisenberg, L. S. (2018). Voice discrimination by adults with cochlear implants: the benefits of early implantation for vocal-tract length perception. *J. Assoc. Res. Otolaryngol.* 19, 193–209. doi: 10.1007/s10162-017-0653-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zaltz and Kishon-Rabin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Appendix A

Three-word sentences including a subject, a predicate, and an object (mean duration = 104.67 ± 6.11 msec) were used for the VD test. The sentences were shortened from the original 5-word sentences from the Hebrew version of the Matrix test to three words in order to minimize working memory demands. Sentences were manipulated using a 13-point stimulus continuum, exponentially ranging in $\sqrt{2}$ steps from a change of -0.18 semitone to a change of -8 semitones. This manipulation resulted in the following three separate dimensions: (1) F0, (2) formant frequencies, in which all formants were shifted down by an equal ratio, according to the 13-point stimulus continuum, and (3) combined, F0+formants, in which both F0 and formants were shifted down in a similar manner. For example, the mean F0 for the F0-manipulated sentences varied by 0, -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 sentence mean F0. Consequently, for the first sentence, the mean F0 was 175.62 Hz and the comparison sentences changed exponentially in $\sqrt{2}$ steps from 174 to 110.35 Hz, using the PSOLA algorithm (Moulines and Charpentier, 1990) for pitch extraction and manipulation. The formant frequencies for this sentence were manipulated exponentially, ranging in $\sqrt{2}$ steps from 0.99 (the smallest ratio between the original formant frequencies and the manipulated formant frequencies) to 0.63 (the highest ratio). This manipulation required resampling the stimulus to compress the frequency axis by a range of factors similar in ratio to the F0 change. The PSOLA algorithm was then applied to regain the original pitch and duration. Given that the average VTL of an adult female is approximately 140 mm (Fitch and Giedd, 1999), and formant frequencies are inversely proportional to VTL (Lammert and Narayanan, 2015), the formants range corresponded to a change from 2 to 88 mm. Note that this is a wider VTL range than expected from humans and was originally chosen to avoid floor effects for CI participants who had difficulty in perceiving normal difference limen (DL) for VTL (Zaltz et al., 2018). All manipulations were implemented using the PRAAT software version 5.4.17 (copyright 1992–2015 by Boersma and Weenink). Spectrographic displays for one of the sentences that was manipulated in F0, formants and F0+formants for the VD test can be found in Zaltz et al. (2020).