Musical Experience Offsets Age-Related Decline in Understanding Speech-in-Noise: Type of Training Does Not Matter, Working Memory Is the Key

Lei Zhang,[1,2] Xueying Fu,[1] Dan Luo,[1] Lidongsheng Xing,[1,2] and Yi Du[1,2,3]

Objectives: Speech comprehension under "cocktail party" scenarios deteriorates with age even in the absence of measurable hearing loss. Musical training is suggested to counteract the age-related decline in speech-in-noise (SIN) perception, yet which aspect of musical plasticity contributes to this compensation remains unclear. This study aimed to investigate the effects of musical experience and aging on SIN perception ability. We hypothesized a key mediation role of auditory working memory in ameliorating deficient SIN perception in older adults by musical training.

Design: Forty-eight older musicians, 29 older nonmusicians, 48 young musicians, and 24 young nonmusicians all with (near) normal peripheral hearing were recruited. The SIN task was recognizing nonsense speech sentences either perceptually colocated or separated with a noise masker (energetic masking) or a two-talker speech masker (informational masking). Auditory working memory was measured by auditory digit span. Path analysis was used to examine the direct and indirect effects of musical expertise and age on SIN perception performance.

Results: Older musicians outperformed older nonmusicians in auditory working memory and all SIN conditions (noise separation, noise colocation, speech separation, speech colocation), but such musician advantages were absent in young adults. Path analysis showed that age and musical training had opposite effects on auditory working memory, which played a significant mediation role in SIN perception. In addition, the type of musical training did not differentiate SIN perception regardless of age.

Conclusions: These results provide evidence that musical training offsets age-related speech perception deficit at adverse listening conditions by preserving auditory working memory. Our findings highlight auditory working memory in supporting speech perception amid competing noise in older adults, and underline musical training as a means of "cognitive reserve" against declines in speech comprehension and cognition in aging populations.

Key words: Aging, Auditory working memory, Musical training, Speech-in-noise perception.

(Ear & Hearing 2021;42;258–270)

INTRODUCTION

Older adults' speech comprehension problem is one of the most prevalent and debilitating aspects of aging and is associated with myriad negative outcomes, including social isolation, depression, and dementia (Uhlmann et al., 1989). Moreover, understanding speech amid competing sounds (e.g., environmental noise, other people talking, music), which is important for everyday communication, represents a significant challenge for older listeners even with normal peripheral hearing (Helfer & Freyman, 2008; Du et al., 2016). Speech-in-noise (SIN) perception is a multifaceted process, supported by the fidelity of bottom-up sensory encoding of target speech (Du et al., 2011; Coffey et al., 2017a), compensatory sensorimotor integration (Du et al., 2014), and higher-level cognitive functions such as auditory working memory and selective attention (Anderson & Kraus, 2010; Kraus et al., 2012; Gordon-Salant & Cole, 2016; Escobar et al., 2020; Puschmann et al., 2019; Yeend et al., 2019). It is established that in addition to peripheral hearing loss, declined central auditory processing and cognitive abilities are critical predictors to the age-related deficit in understanding SIN (Anderson et al., 2013).

Interestingly, compared with people without musical training, musicians have been found to perform better in various SIN tasks on different timescales, from phonemes in white noise to sentences in multi-talker babble (see Coffey et al., 2017b for a review), although several studies conducted in young adults failed to find the musician advantage in SIN perception at the sentence level (Fuller et al., 2014; Ruggles et al., 2014; Boebinger et al., 2015; Escobar et al., 2020). Moreover, musical training exhibits a differential preservation pattern in mitigating the age-related decline in SIN perception (Zendel & Alain, 2012; Alain et al., 2014). That is, there is an interaction between age and the amount of training, that the rate of age-related decline in SIN performance is slower in musicians than nonmusicians. However, the mechanisms engendering the "cognitive reserve" that can delay the aging effects on SIN ability by musical training remain poorly understood. On the sensory-perceptual level, musically trained older adults showed strengthened central auditory processing, represented as better pitch discrimination (Dubinsky et al., 2019), less delay in neural timing of speech-evoked brainstem responses (Parbery-Clark et al., 2012), and more coordinated speech representations in the auditory brainstem and cortex (Bidelman & Alain, 2015). On the cognitive level, older adults with long- or short-term musical experience exhibited stronger SIN perception along with enhanced attention-related brain activity (Zendel & Alain, 2014; Zendel et al., 2019). The musician advantage in SIN perception was also correlated with better auditory working memory, the ability to temporarily hold sound information in memory for processing, in young adults (aged 18 to 35; Parbery-Clark et al., 2009; Yoo & Bidelman, 2019) and middle-aged adults (aged 45 to 65; Parbery-Clark et al., 2011). However, to date no study has directly investigated the relationships between those musical training-related perceptual-cognitive benefits and SIN performance in older adults.

In the current study, we tried to put the pieces of the puzzle together by first investigating whether musical training was associated with better speech sentence perception in speech-spectrum

[1]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China; [2]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China; and [3]CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site (www.ear-hearing.com).

0196/0202/2021/422-258/0 • Ear & Hearing • Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved • Printed in the U.S.A.

<zdoi; 10.1097/AUD.0000000000000921>

noise (i.e., energetic masking) and two-talker speech (i.e., informational masking) using a perceived spatial separation paradigm (Wu et al., 2005) in both older and young adults. Next, we conducted a path analysis (Pearl, 2012) which provides a way to infer the causal relationships of age and musical experience on SIN performance in a statistical sense, although it cannot confirm the causal link. Considering that working memory is critical for speech comprehension even in the absence of noise (Wingfield & Tun, 2007), and auditory working memory predicted SIN ability at the sentence level in both young, middle-aged and older adults (Parbery-Clark et al., 2009, 2011; Anderson et al., 2013; Yeend et al., 2019; Yoo & Bidelman, 2019; Escobar et al., 2020), we hypothesized an important mediation role of auditory working memory in offsetting the age-related decline of SIN perception by musical training experience. In addition, different types of musical training emphasize different parts of auditory, cognitive, and neural plasticity (Merrett et al., 2013; Slater et al., 2017), which may lead to distinct outcomes in SIN tasks. Recent studies found that young percussionists (skilled at rhythmic discrimination) but not young vocalists (skilled at melodic discrimination) outperformed young nonmusicians in sentence-in-noise perception (Slater & Kraus, 2016) and inhibitory control (Slate et al., 2017). In older adults, 10 weeks of choir singing improved sentence-in-noise perception (Dubinsky et al. 2019) but 6 months of extensive piano training failed to benefit sentence-in-noise perception (Fleming et al., 2019). Nonetheless, the effect of various kinds of musical training on SIN perception or auditory working memory has not been directly investigated in older adults yet. Here, although the groups were not perfectly matched, older musicians were divided into vocalists and instrumentalists (a combination of wind/string players and pianists), and young musicians were divided into wind/string players, pianists, and percussionists to ensure enough statistical power for investigating the musician type effect on auditory working memory and SIN performance (see Methods for details). We hypothesized that auditory working memory would be enhanced regardless of training type which would in turn contribute to undistinguishable SIN performance in older musicians but not necessarily in young musicians.

MATERIALS AND METHODS

Participants

Seventy-seven Chinese older adults (age 57 to 73 years) and 72 Chinese young adults (age 18 to 33 years) were recruited and signed the written consent approved by the Institute of Psychology, Chinese Academy of Sciences. Older participants included 29 nonmusicians (18 females), 24 instrumentalists (9 females), and 24 vocalists (12 females). Young participants included 24 nonmusicians (12 females), 16 wind/string players (9 females), 16 pianists (9 females), and 16 percussionists (9 females). Most musicians were recruited from conservatory of music, chorus, and orchestra. Sample size was estimated based on power analyses ($\alpha = 0.05$, power = 0.8) in G*Power 3.1 (Faul et al., 2009). Using a 2×2 two-way analysis of variance (ANOVA), 149 participants of 4 groups would be sufficient to detect a median effect size (Cohen's $f = 0.25$) with 86% power, and using a one-way ANOVA of three or four groups, 77 or 72 participants would be sufficient to detect a large effect size (Cohen's $f = 0.4$) with 88% or 80% power.

All participants were healthy, right-handed, native Chinese speakers. All young participants had normal hearing (average pure-tone threshold ≤ 20 dB HL for 250 to 8,000 Hz) at both ears; all older participants had normal hearing (average pure-tone threshold ≤ 25 dB HL for 250 to 3,000 Hz, frequency range most relevant for speech understanding, Turner & Cummings, 1999) at both ears with no more than 1 threshold between 250 and 3,000 Hz over 30 dB HL. Older participants showed typical age-related high-frequency hearing loss, the averaged hearing level for 4,000 to 8,000 Hz was 30.47 dB in older musicians and 28.71 dB in older nonmusicians. Figure 1 shows the averaged air conduction audiograms measured by the Bell Plus audiometer (Bell, Inventis, Italy) with a TDH-39 headphone for older nonmusicians, older musicians (combining instrumentalists and vocalists), young nonmusicians, and young musicians (combining pianists, wind/string players, and percussionists). A mixed ANOVA of age × group × frequency showed that older adults had higher pure-tone hearing levels than young adults ($F(1, 145) = 293.907$, $p < 0.001$), particularly at high frequencies (age × frequency interaction, $F(6, 870) = 88.824$, $p < 0.001$), while musicians and nonmusicians had equal pure-tone hearing ($F(1, 145) = 0.118$, $p = 0.732$). Further analyses

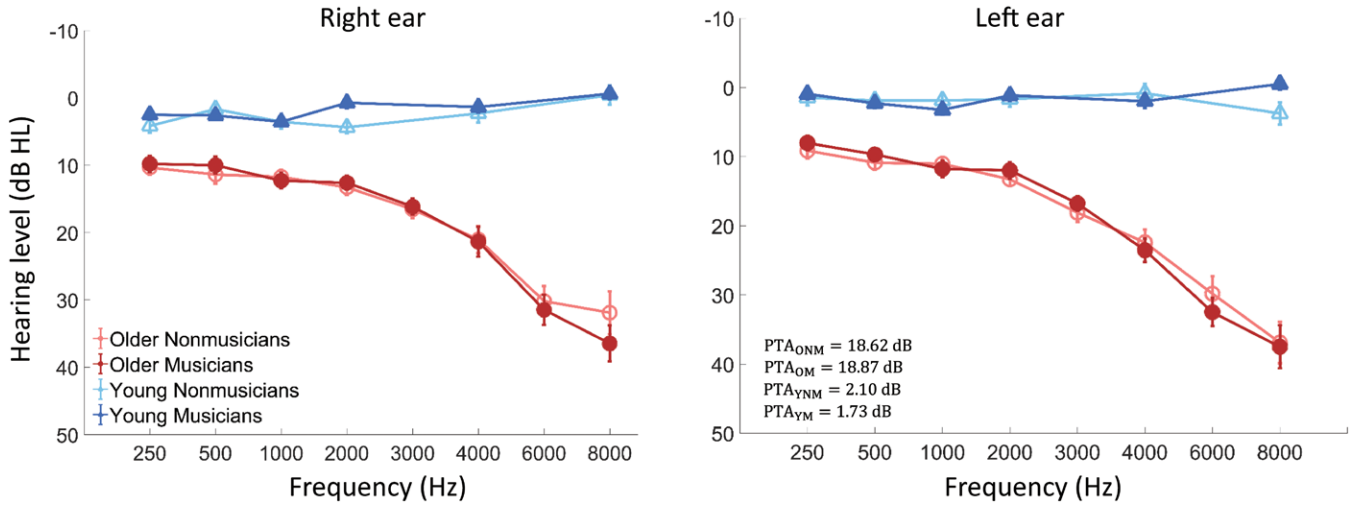

Fig. 1. Group mean pure-tone hearing thresholds at each frequency for young musicians, young nonmusicians, older musicians, and older nonmusicians. Error bars indicate standard error of the mean.

TABLE 1. The Group Mean (Standard Deviation) Values and Statistics of Age, Higher Education, Hearing Level at 250–8,000 Hz, MOCA Score, Nonverbal IQ, Stroop, Auditory Digit Span (Sum of Forward and Backward Digit Span), Age of Training Onset, and Years of Music Training in Each Group

Group	Age	Education	Hearing Level	MOCA	Nonverbal IQ	Stroop	Auditory Digit Span	Age of Onset	Years of Training
O Instrumentalists	64.92 (4.25)	5.21 (2.78)	19.14 (6.14)	28.17 (1.31)	NA	0.62 (0.33)	15.88 (2.11)	10.92 (3.05)	51.50 (9.48)
O Vocalists	64.67 (3.58)	4.97 (2.66)	18.59 (7.51)	27.58 (0.97)	NA	0.51 (0.25)	13.92 (1.69)	12.83 (5.52)	42.00 (12.54)
O Nonmusicians	65.66 (4.13)	5.03 (3.15)	18.62 (5.73)	27.97 (1.02)	NA	0.60 (0.42)	12.72 (2.37)	NA	NA
F/*t* (*p*)	0.44 (0.644)	0.04 (0.959)	0.06 (0.945)	1.74 (0.183)		0.74 (0.479)	14.89 (<0.001)	1.49 (0.143)	2.96 (0.005)
Y Wind/Strings	21.19 (2.07)	5.88 (1.41)	1.25 (4.13)	NA	29.63 (2.92)	NA	17.75 (2.57)	4.88 (1.54)	16.31 (2.18)
Y Pianists	24.19 (4.09)	7.38 (2.09)	2.78 (3.63)	NA	29.81 (5.34)	NA	17.87 (1.63)	5.69 (1.20)	15.56 (2.80)
Y Percussionists	21.19 (2.34)	6.06 (2.26)	1.16 (2.36)	NA	32.44 (4.08)	NA	18.56 (2.68)	5.00 (1.21)	13.38 (2.60)
Y Nonmusicians	23.21 (3.05)	6.88 (1.62)	2.10 (2.95)	NA	29.08 (3.49)	NA	17.63 (2.41)	NA	NA
F (*p*)	4.18 (0.009)	2.08 (0.078)	0.89 (0.452)		2.45 (0.071)		0.55 (0.652)	1.74 (0.187)	5.76 (0.006)

One-way ANOVAs or independent two-sample t tests (for age of onset and years of training) were used for examining the group differences.
ANOVA, analysis of variance; MOCA, Montreal Cognitive Assessment; NA, data were not collected; O, older; Y, younger.

showed that the three older groups (F(,74) = 0.06, p = 0.945) and four young groups (F(3,68) = 0.89, p = 0.452) did not differ in hearing level at 250 to 8,000 Hz. Moreover, no correlation was found between age and hearing level at 250 to 8,000 Hz in young or older adults (both $r < 0.20$, $p > 0.05$). Note that, since handness and language lateralization were not the research scope here, only right-handed participants were recruited.

Data about musical history of young musicians were collected via the Chinese version of Montreal Music History Questionnaire (MMHQ; Coffey et al., 2011). Older musicians completed a brief version including the primary and secondary instruments, age of training onset, years of total training, and practice hours per week in recent 3 years. Table S1 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686 shows the detailed training information of musicians. Older musicians had started training before 20 years old (except two vocalists started at age 21 and 23), had at least 20 years of training (except one instrumentalist with 16 years of training), and practiced consistently in recent 3 years (1.5 to 39 hours per week, mean = 9.77 ± 8.03 hours). Young musicians had started training before 8 years old, had at least 10 years of continuous training (2 to 50 hours per week, mean = 12.60 ± 11.29 hours). Nonmusicians reported less than 2 years of musical training experience. Older musician groups did not differ in the age of training onset (t(46) = −1.49, p = 0.143), but older instrumentalists had more years of training than older vocalists (t(46) = 2.96, p = 0.005, Cohen's d = 0.85). Young musician groups started training at similar age (F(2,45) = 1.74, p = 0.187), but young percussionists had slightly fewer years of training than young wind/string players (p = 0.006, 95% confidence interval [CI] = 0.76, 5.12) and pianists (p = 0.049, 95% CI = 0.01, 4.37). In addition, years of training did not correlated with hearing level at 250 to 8,000 Hz in either young or older musicians (both $|r| < 0.12$, $p > 0.05$).

All older subjects completed a questionnaire including self-reported health status (4-point scale: very good, good, fair, poor), self-rated life satisfaction (4-point scale: very satisfied, neutral, dissatisfied, very dissatisfied), living status (dichotomy: live alone, live with families), social activity (dichotomy: still working or often attend community activities, barely no social activity), and passed the Montreal Cognitive Assessment (MOCA) of Beijing version (≥ 26 scores) for screening out subjects with mild cognitive impairment (Yu et al., 2012). The three older groups were matched for age (F(2,74) = 0.44, p = 0.644), years of higher education (F(2,74) = 0.04, p = 0.959), MOCA score (F(2,74) = 1.74, p = 0.183), self-rated health status (χ^2(2,2) = 2.75, p = 0.601), self-rated life satisfaction (χ^2(1,2) = 0.60, p = 0.739), living status (χ^2(1,2) = 4.12, p = 0.127), and social activity (χ^2(1,2) = 5.26, p = 0.072). The four young groups were matched for years of higher education (F(3,68) = 2.38, p = 0.078) and nonverbal IQ (Cattell's culture fair intelligence test, form 3A, Cattell & Cattell, 1960) (F(3,68) = 2.45, p = 0.071). The nonverbal IQ was controlled in young adults since several studies have found that the SIN performance could be predicted by young participants' nonverbal IQ (Ruggles et al., 2014; Boebinger et al., 2015), but it was not measured in older adults because the test version is too difficult and time-consuming for older subjects. No age difference was found between young nonmusicians and musicians (t(70) = 1.28, p = 0.203). The descriptive analyses of participants were summarized in Table 1.

SPEECH-IN-NOISE PERCEPTION

Stimuli

Speech stimuli were Chinese nonsense sentences (Wu et al., 2005), which were translated from English nonsense sentences developed by Helfer (1997) and widely used in psychoacoustic studies (Freyman et al., 2001; Ruggles et al., 2014). Nonsense sentences are syntactically correct but semantically meaningless, that is, the sentence frame does not provide any contextual support for recognition of key words. For instance, the English translation of a Chinese nonsense sentence "一些条令已经翻译我的大衣" is "Some *rules* had *translated* my *coat*" (the three 2-character keywords are italic). Sentences with low or no context are widely used in SIN perception tests. For example, sentences in the QuickSIN, a widely used clinical measure, are syntactically correct yet contain low semantic or contextural cues, for example, "*The square peg will settle in the round hole.*" Wilson et al. (2007) found that the QuickSIN is more sensitive to performance difference between normal hearing and hearing impaired groups than the BKB-SIN and Hearing In Noise Test (HINT), which use meaningful sentences. In a systematic review by Coffey et al. (2017b) and some recent works (Dubinsky et al., 2019; Yoo & Bidelman, 2019; Escobar et al., 2020), nearly a half of studies that have examined the relationship between musical

training and SIN perception used sentences with low or no context. Moreover, it is found that working memory correlated with recognition performance of low-context sentences in noise but not necessarily correlated with high-context sentences in noise (Parbery-Clark et al. 2009, 2011; Wayne et al., 2016; Escobar et al., 2020). Thus, nonsense sentences with no context were used here to diminish the impact of top-down prediction and directly investigate the effects of musical training and aging on the perceptual (bottom-up) level of speech processing, and increase the load on working memory to test the hypothesis that working memory mediates the musical training-related alleviation of speech perception difficulty in older adults.

Target sentences were spoken by a young female talker (Talker A). There were two types of the masker: speech spectrum noise and two-talker speech. The speech masker consisted of two different Chinese nonsense sentences spoken by two young female talkers (Talkers B and C). Nonsense sentences in the speech masker were similar in linguistic structure to the target nonsense sentences but differed in their content. The spectrum of the noise masker was representative of the average spectrum of 500 Chinese sentences from Talker B and C (Wu et al., 2005).

The stimuli were presented binaurally through Sennheiser HD380 Pro headphones driven by a Dell desktop computer. The perceived spatial relationship between the target and the masker was achieved by manipulating the interaural time difference (ITD). The target speech was always presented at 0 ms ITD, thus perceived as coming from the center of the head. The masker sound was presented at three ITD conditions: −2 ms (left ear led right ear 2 ms), 0 ms, and 2 ms (right ear led left ear 2 ms). According to the precedence effect (Wallach et al., 1949), the masker was perceived as coming from the left ear, the center of the head, and the right ear, respectively. Therefore, there were two spatial relationships between the target and the masker at the perceptual level: colocation and separation, although both the target and the masker were played at both ears and had no physical separation. It has been confirmed that listeners could benefit from perceived target–masker spatial separation in recognition of target signals (Wu et al., 2005).

Stimulus levels were calibrated using a Larson-Davis sound level meter (Model 831, Depew, NY). The target stimuli were presented at 65 dB sound pressure level and the level of the maskers was adjusted to produce five different signal-to-noise ratios (SNR = −12, −8, −4, 0, and 4 dB).

Procedure

The SIN task contained three within-subject variables: (1) masking type (noise, speech); (2) perceived target-masker spatial relationship (colocation, separation); (3) SNR (−12, −8, −4, 0, and 4 dB). The four combinations of masker type and spatial relationship were separately presented in four blocks (noise colocation, noise separation, speech colocation, speech separation), which were partially counterbalanced across subjects in each group using a Latin square order. In the two spatial separation blocks, the maskers were perceived as coming from the left ear in half of trials and the right ear in other half of trials. Each block contained 40 trials, with eight trials per SNR randomly arranged in the block. Before the formal experiment, a practice session of 18 trials [2 masker types × 3 SNRs (−12, −4, and 4 dB) × 3 masker locations (left, middle, and right)] was presented to get participants familiar with the stimuli and the task.

Participants sat in a sound-attenuating chamber, which was 293 × 293 × 198 cm in size (length × width × height), to perform the task. Before each block, participants were informed of the masker type and the target-masker spatial relationship. In each trial, participants pressed the "Space" key to start the masker sound. About 1 s later, a single target sentence was presented and gated off with the masker. Participants were instructed to loudly repeat the whole target sentence as best as they could immediately after the sentence was completed.

Participants' responses were online scored by the experimenter sitting outside the sound-attenuating chamber and also recorded by a digital voice recorder (PHILIPS VTR6600, AMS, NL) for off-line examination. The keyword was scored only when both of the characters were repeated correctly. The number of correctly identified keywords was tallied later.

Data Analysis

A logistic psychometric function,

$$y = \frac{1}{1 + e^{-\sigma(x-\mu)}}$$

was employed in Matlab 2016b to fit each subject's data of four blocks separately, using the Levenberg–Marquardt method (Wolfram, 1992), where y is the probability of correct recognition of the keywords, σ determines the slope of the psychometric function, x is the SNR corresponding to y, and μ is the SNR corresponding to 50% correct identification (the threshold ratio in dB). The release amount of spatial unmasking was calculated as the difference in the threshold between separation and colocation, for the noise masker and speech masker separately. Larger spatial release amount indicates larger cognitive benefit from spatial attention and better binaural auditory processing to the target in perceiving SIN.

Auditory Working Memory

Auditory working memory was measured using the forward and backward Digit Span subtest of the Wechsler Adult Intelligence Scale of Chinese version (Gong, 1992). Participants were presented with a recording of a series of digits spoken by a young Chinese female. The number of digits increased from 3 to 12 in the forward part and from 2 to 10 in the backward part (two trials per length), participants were asked to repeat the digits in a normal and reverse order, respectively. The task stopped if both trials for the same length were incorrect. Auditory working memory score was defined as the sum of the longest numbers participants could repeat in the forward and backward parts. In comparison with forward digit span which includes primarily a memory component, backward digit span includes an additional executive function component. Therefore, besides the sum of forward and backward digit span, backward digit span was separately tested in group analysis (see Figure S1 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686), correlation analyses (see Table S3 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686), and path analyses (see Figure S2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686). Note that, raw scores instead of age-normed scores were used for analysis because the age difference in auditory working memory was the research interest in the current study.

Inhibition Control

The Stroop test (Stroop, 1935) was used to assess inhibition control ability of the elderly. Participants were asked to name the color of dots, Chinese words (e.g., "sun (阳)" in green), and Chinese color words (e.g., "red (红)" in green, the lexical meaning and color were always incongruent) in three cards. Performance was indexed as the following equation: (time of color words-time of words)/time of words. Larger values indicated poorer performance. Note that young participants did not perform the Stroop test because unfortunately the Stroop test was included after all young participants had completed the experiment.

Statistical Analyses

Depending on situations, paired *t* tests, independent 2-sample *t* tests, one-way and two-way ANOVA were conducted to explore the group differences on behavioral data using Matlab 2016b. Multiple comparison tests were conducted using Tukey's honestly significant difference procedure. Jarque–Bera test and Lilliefors test in Matlab 2016b were used for examining data normality. SIN perception threshold of young musicians/adults under noise colocation condition, SIN threshold of old nonmusicians under noise separation condition, total training hours of young musicians, and years of training of old musicians were not normally distributed. Other variables used in the analysis were normally distributed. Pearson partial correlations (Spearman partial correlations were used for analyses involving nonnormally distributed variables) were implemented to test the relationships between SIN thresholds and years of training (only in musicians), working memory or inhibitory control after controlling for hearing level at 250 to 8,000 Hz and age.

Path Analyses

Path analysis was performed to examine the direct and indirect effects of musical expertise and age on SIN performance for the four conditions separately, using AMOS software 22.0. Note that, path analysis was done for the whole sample, not on each group separately like the correlation between SIN performance and year of training did, all the variables passed the normality test. The bootstrapping method with 5,000 iterations was used to estimate a 95% CI. If zero was outside the 95% bias-corrected CI computed by the bootstrapping procedure, the direct/indirect effect would be considered significant. The indices of model fitting included Chi-square statistic ($\chi 2$), its degrees of freedom, and *p* value, root mean square error of approximation (RMSEA) and its associated confidence interval, root mean square residual (RMR), normed-fit index (NFI) and comparative fit index (CFI). *P* of $\chi 2 > 0.05$, RMSEA < 0.07, RMR < 0.08, NFI > 0.95, and CFI > 0.95 indicate an acceptable fit of the model (Hooper et al., 2008).

To test our hypothesis, auditory working memory (Fig. 6: sum of forward and backward digit span; see Figure S2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686: backward digit span) was set as a mediator in explaining the effects of musical training and age on SIN performance. Because no significant correlation between years of training and SIN performance was found in either older or young musicians (see Table S2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686) and the path model treating musical training as a continuous variable (years of training) failed in model fitting (*p* of $\chi 2 < 0.001$, RMR > 7.35, CFI > 0.749, NFI > 0.753, RMSEA > 0.482), musical training was coded as a dummy variable in the path analysis. In addition, individual's hearing level was not included in the path model, because hearing level at 250 to 8,000 Hz was highly correlated with age when young and older subjects were combined ($r = 0.86$, $p < 0.001$, no correlation was found in young or older adults alone), making it unsuitable as a covariate due to the multicollinearity problem. Last, the path model including both auditory working memory and hearing as mediators failed in model fitting (RMR > 0.45).

RESULTS

Group Differences in SIN Perception

Figures 2–4 show the group mean percent of correct as a function of SNR and the SIN perception threshold (in dB) computed by the psychometric function under four conditions. Comparisons were first conducted among four older groups (Fig. 2) and three young groups (Fig. 3), then musicians were combined regardless of training type and comparisons were implemented among older and young nonmusicians and musicians (Fig. 4).

For three older groups (Fig. 2), separate one-way ANOVA showed a significant main effect of group on SIN threshold under four conditions (noise separation: $F(2, 74) = 13.08$, $p < 0.001$, $\eta^2 = 0.26$; noise colocation: $F(2, 74) = 6.79$, $p = 0.002$, $\eta^2 = 0.16$; speech separation: $F(2, 74) = 6.40$, $p = 0.003$, $\eta^2 = 0.15$; speech colocation: $F(2, 74) = 7.07$, $p = 0.002$, $\eta^2 = 0.16$). Note that large effect sizes (Cohen's $f = 0.42$–0.59) were found by one-way ANOVAs, which validates the rationality of using a large effect size in sample size estimation. The multiple comparison tests showed that the thresholds of older instrumentalists (noise separation: $p < 0.001$, 95% CI = 0.65, 2.25; noise colocation: $p = 0.002$, 95% CI = 0.31, 1.67; speech separation: $p = 0.007$, 95% CI = 0.33, 2.37; speech colocation: $p = 0.002$, 95% CI = 0.58, 2.85) and older vocalists (noise separation: $p < 0.001$, 95% CI = 0.66, 2.25; noise colocation: $p = 0.025$, 95% CI = 0.08, 1.44; speech separation: $p = 0.012$, 95% CI = 0.23, 2.27; speech colocation: $p = 0.034$, 95% CI = 0.07, 2.35) were both significantly lower than the threshold of older nonmusicians. No significant difference was found between the threshold of older instrumentalists and that of older vocalists ($p > 0.05$).

For four young groups (Fig. 3), separate one-way ANOVA showed insignificant effect of group on SIN threshold under four conditions (noise separation: $F(3, 68) = 0.69$, $p = 0.563$, $\eta^2 = 0.03$; noise colocation: $F(3, 68) = 0.80$, $p = 0.496$, $\eta^2 = 0.03$; speech separation: $F(3, 68) = 0.66$, $p = 0.578$, $\eta^2 = 0.03$; speech colocation: $F(3, 68) = 0.67$, $p = 0.573$, $\eta^2 = 0.03$).

Since no significant effect of musician type was found in SIN perception, we combined different types of musicians into older musicians group and young musicians group in later analyses. As showed in Figure 4, a two-way ANOVA revealed significant interaction between age and musical training under four conditions (noise separation: $F(1, 145) = 15.81$, $p < 0.001$, $\eta^2 = 0.08$; noise colocation: $F(1, 145) = 5.86$, $p = 0.017$, $\eta^2 = 0.03$; speech separation: $F(1, 145) = 5.00$, $p = 0.027$, $\eta^2 = 0.03$; speech colocation: $F(1, 145) = 9.49$, $p = 0.003$, $\eta^2 = 0.04$), in addition to main effects of age (noise separation: $F(1, 145) = 15.27$, $p < 0.001$, $\eta^2 = 0.08$; noise colocation: $F(1, 145) = 23.56$, $p < 0.001$, $\eta^2 = 0.13$; speech separation: $F(1, 145) = 4.09$, $p = 0.045$, $\eta^2 = 0.03$; speech colocation: $F(1, 145) = 68.82$, $p < 0.001$, $\eta^2 = 0.31$) and music training (noise

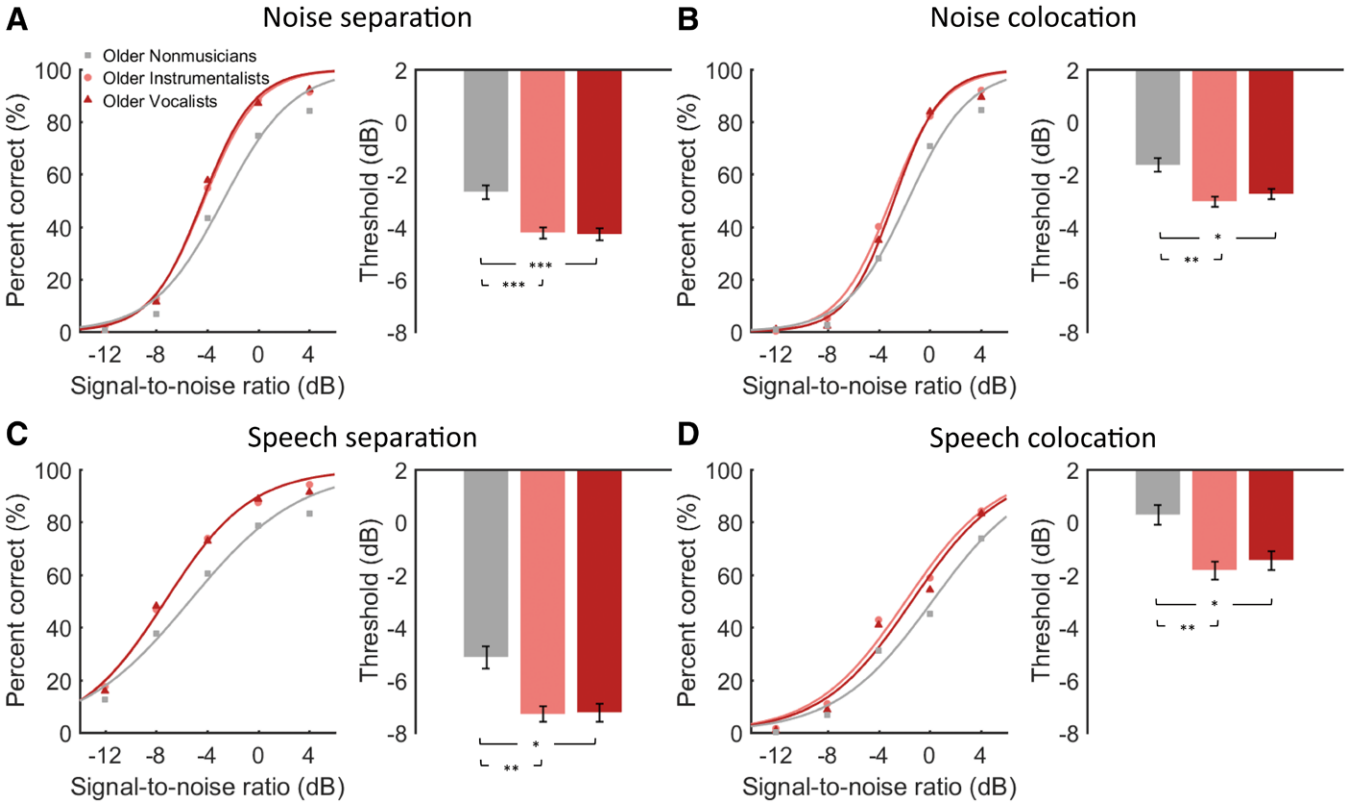

Fig. 2. Group mean percent of correct as a function of signal-to-noise ratio (left panel) and the mean speech-in-noise threshold (right panel) in older nonmusicians, older instrumentalists, and older vocalists under (A) noise separation, (B) noise colocation, (C) speech separation, and (D) speech colocation. Error bars indicate standard error of the mean. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, one-way ANOVA followed by Tukey's multiple comparison tests.

separation: $F(1, 145) = 16.67$, $p < 0.001$, $\eta^2 = 0.09$; noise colocation: $F(1, 145) = 9.2$, $p = 0.003$, $\eta^2 = 0.05$; speech separation: $F(1, 145) = 8.46$, $p = 0.004$, $\eta^2 = 0.05$; speech colocation: $F(1, 145) = 4.51$, $p = 0.035$, $\eta^2 = 0.02$). That is, young adults outperformed older adults and musicians outperformed nonmusicians, but the musician advantage in SIN was larger in older adults than in young adults. Note that small to large effect sizes (Cohen's $f = 0.14$–67) were found by two-way ANOVAs, which validates the rationality of using a median effect size in sample size estimation.

The multiple comparison tests showed that the threshold of older nonmusicians was significantly lower than that of older musicians (noise separation: $p < 0.001$, 95% CI = 0.82, 2.09; noise colocation: $p < 0.001$, 95% CI = 0.31, 1.44; speech separation: $p = 0.001$, 95% CI = 0.41, 2.19; speech colocation: $p = 0.001$, 95% CI = 0.47, 2.45), that of young nonmusicians (noise separation: $p < 0.001$, 95% CI = 0.67, 2.17; noise colocation: $p < 0.001$, 95% CI = 0.50, 1.83; speech separation: $p = 0.040$, 95% CI = 0.03, 2.12; speech colocation: $p < 0.001$, 95% CI = 2.04, 4.36) and that of young musicians (noise separation: $p < 0.001$, 95% CI = 0.81, 2.08; noise colocation: $p < 0.001$, 95% CI = 0.70, 1.83; speech separation: $p = 0.002$, 95% CI = 0.36, 2.13; speech colocation: $p < 0.001$, 95% CI = 1.94, 3.92) under all four conditions. Although older musicians had poorer peripheral hearing than young adults particularly at high frequencies, older musicians did equally well in SIN perception as young nonmusicians and young musicians in all conditions except the speech colocation condition. Under this condition, older musicians still performed worse than young adults (young nonmusicians: $p < 0.001$, 95% CI = 0.68, 2.79; young musicians: $p < 0.001$, 95% CI = 0.61, 2.33).

In addition, as shown in Table 2, after controlling for hearing and age, years of training did not significantly correlate with SIN threshold in older musicians or young musicians (all $|r| < 0.26$, $p > 0.077$), nor did total hours of training correlate with SIN performance in young musicians (all $r < -0.20$, $p > 0.184$). The three older groups (nonmusicians, instrumentalists, vocalists) performed equally well in inhibitory control ($F(2,45) = 0.74$, $p = 0.479$, Table 1) and inhibitory control did not predict SIN threshold in older adults after controlling for hearing and age (all $r < 0.18$, $p > 0.117$, Table 2). No correlation was found either between SIN threshold and nonverbal IQ in young adults (all $|r| < 0.07$, $p > 0.540$, Table 2).

Spatial Release From Masking

As shown in Table 3, larger spatial release was observed from speech masking than noise masking in spite of age and musical experience (all $t < -5.41$, $p < 0.001$, Cohen's $d > 1.52$, paired t tests). Older musicians achieved greater spatial release from noise masking than older nonmusicians (1.42 dB vs. 0.84 dB, $t(75) = -2.68$, $p = 0.009$, Cohen's $d = 0.63$, independent two-sample t test), but not so when the masker switched to speech ($t(75) = 0.57$, $p = 0.572$). Young musicians and young nonmusicians did not differ in spatial release amount either for noise masker ($t(70) = 0.40$, $p = 0.690$) or speech masker ($t(70) = -1.09$, $p = 0.281$).

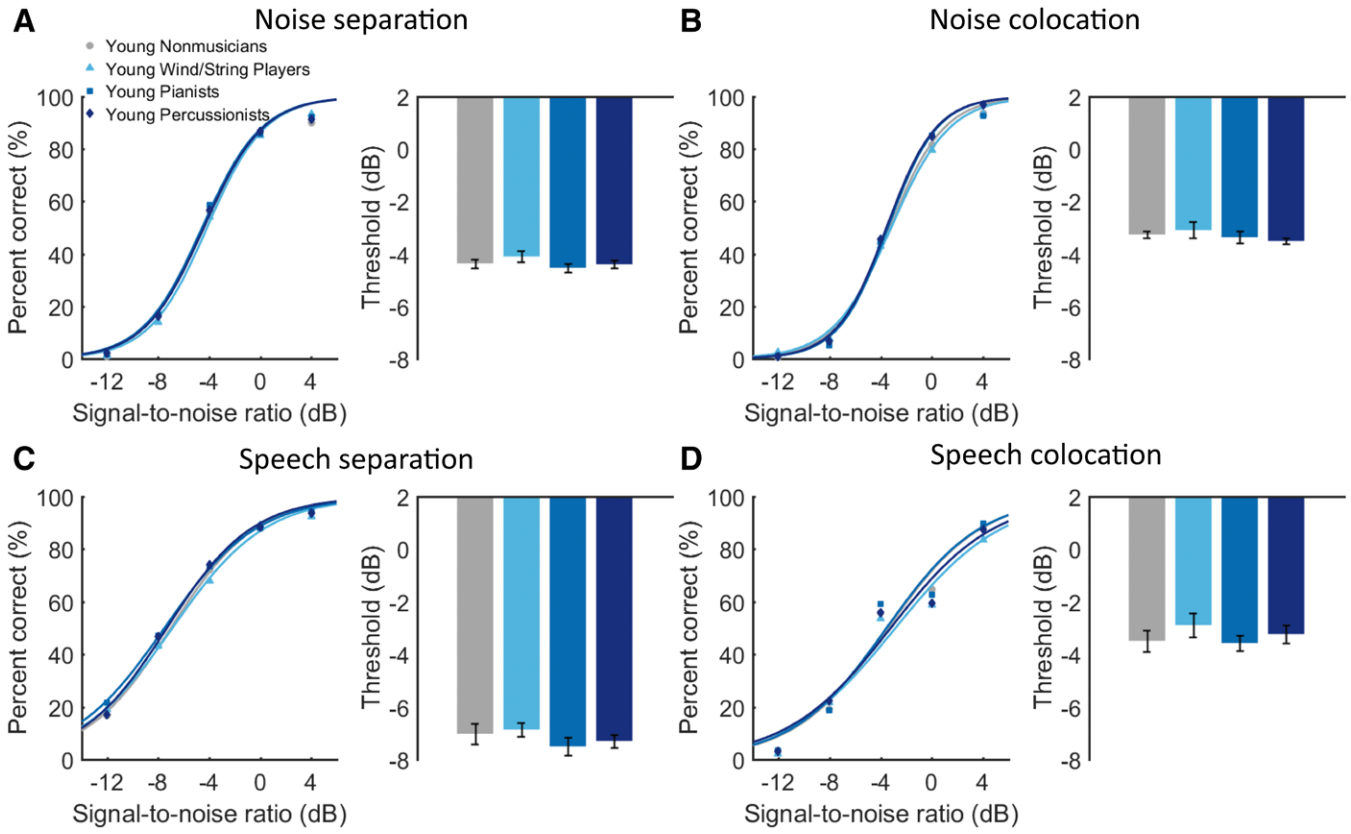

Fig. 3. Group mean percent of correct as a function of signal-to-noise ratio (left panel) and the mean speech-in-noise threshold (right panel) in young nonmusicians, young wind/string players, young pianists, and young percussionists under (A) noise separation, (B) noise colocation, (C) speech separation, and (D) speech colocation. Error bars indicate standard error of the mean. Note that no group difference was observed for young adults.

Auditory Working Memory and Its Correlation With SIN

A one-way ANOVA showed a significant group effect on auditory working memory represented as the sum of forward and backward digit span (Fig. 5A, $F(3, 145) = 40.56$, $p < 0.001$, $\eta^2 = 0.47$). Older nonmusicians exhibited significantly lower auditory working memory than older musicians ($p < 0.001$, 95% CI = 0.79, 3.55), young nonmusicians ($p < 0.001$, 95% CI = 3.28, 6.52) and young musicians ($p < 0.001$, 95% CI = 3.96, 6.72), while older musicians had lower auditory working memory than young nonmusicians ($p < 0.001$, 95% CI = 1.26, 4.20) and young musicians ($p < 0.001$, 95% CI = 1.97, 4.37). That is, musical training experience was associated with improved auditory working memory in older adults but not in young adults. Forward and backward digit span alone showed the same pattern of group difference as the sum of both (see Figure S1 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686). Moreover, after controlling for age, years of training was significantly correlated with auditory working memory in older musicians (Fig. 5B, $r = 0.33$, $p = 0.025$) but not in young musicians ($r = -.10$, $p = 0.503$; $r = 0.00$, $p = 0.980$ when years of training was replaced by total hours of training). After controlling for both age and hearing level, the relationship between years of training and auditory working memory became marginally significant in older musicians ($r = 0.25$, $p = 0.097$). In addition, as shown in Table 1, older instrumentalists had stronger auditory working memory ($t(46) = -3.54$, $p < 0.001$, Cohen's $d = 1.02$) and more years of training ($t(46) = -2.96$, $p = 0.005$, Cohen's $d = 0.85$) than older vocalists, whereas the three young musician groups did not differ in auditory working memory ($F(2,45) = 0.56$, $p = 0.576$) although young percussionists had fewer years of training than young wind/string players ($p = 0.006$, 95% CI = 0.76, 5.12) and pianists ($p = 0.049$, 95% CI = 0.01, 4.37). After matching older vocalists and older instrumentalists in training length by removing eight vocalists with no more than 34 years of training, older instrumentalists still performed better in auditory working memory ($t(38) = 2.59$, $p = 0.014$, Cohen's $d = 0.84$) but not in SIN perception (all $|t(38)| < 0.46$, $p > 0.649$) than older vocalists. This suggests that both the amount of training and the type of training impact auditory working memory in older adults.

More importantly, after controlling for hearing level and age, auditory working memory correlated with SIN performance under four conditions in older adults (noise separation: $r = -0.50$, $p < 0.001$; noise colocation: $r = -0.46$, $p < 0.001$; speech separation: $r = -0.43$, $p < 0.001$; speech colocation: $r = -0.53$, $p < 0.001$), but not in young adults (all $|r| < 0.22$, $p > 0.062$; Fig. 5C to F). Notably, this pattern was repeatedly found when forward and backward digit span were separately tested, although the sum of forward and backward digit span showed the strongest correlation in older adults (see Table S2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686).

Path Analyses

Figure 6 summarizes the results of the path analyses for four conditions when auditory working memory (sum of forward and backward digit span) was set as a mediator in explaining

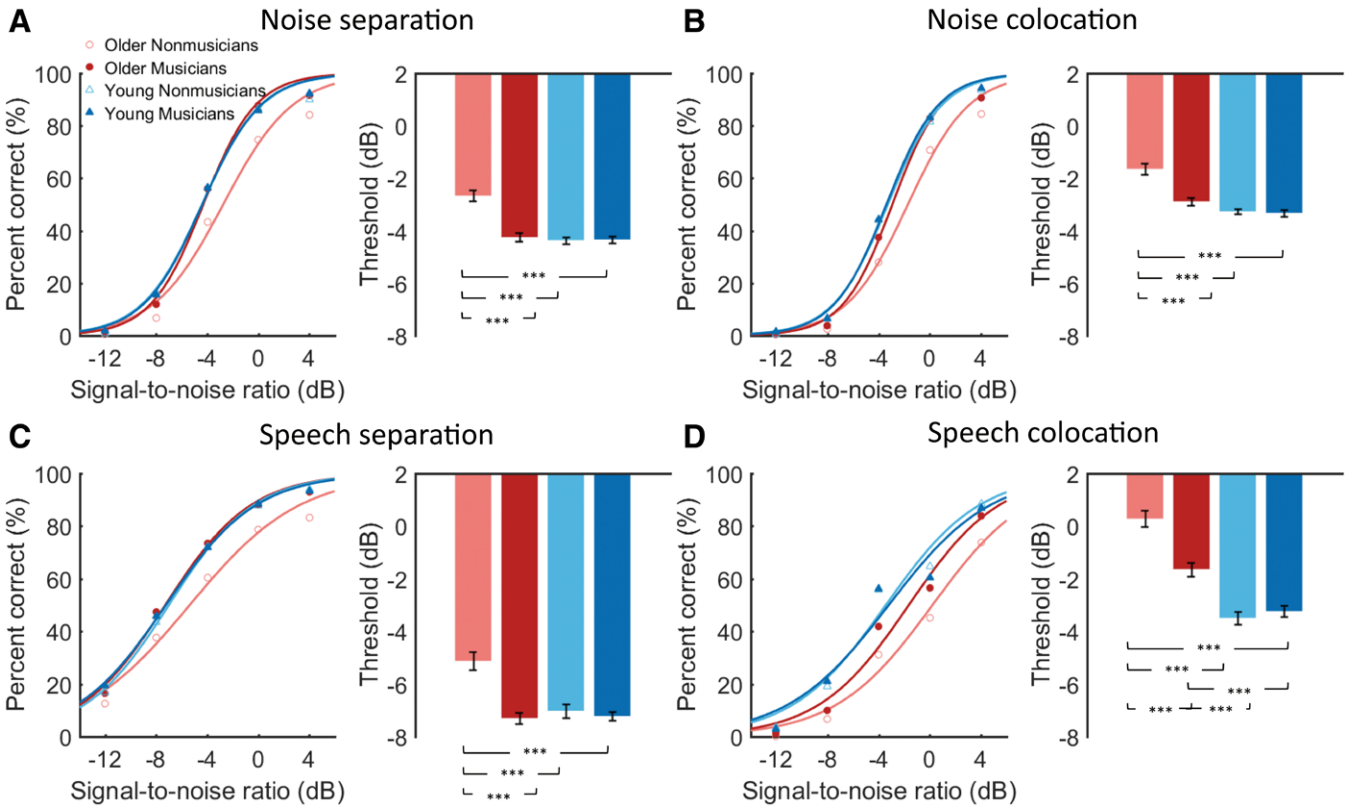

Fig. 4. Group mean percent of correct as a function of signal-to-noise ratio (left panel) and the mean speech-in-noise threshold (right panel) in older nonmusicians, older musicians, young nonmusicians, and young musicians under (A) noise separation, (B) noise colocation, (C) speech separation, and (D) speech colocation. Error bars indicate standard error of the mean. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, two-way ANOVA followed by Tukey's multiple comparison tests.

the effects of musical training and age on SIN performance. All four models fitted the data well: $\chi2(1) = 0.30$, $p = 0.582$; RMSEA = 0.00, 90% CI = 0.00, 0.18; RMR = 0.04; NFI = 1.00; CFI = 1.00. Four models could separately explain 21%, 24%, 10%, and 38% of variance in SIN thresholds under noise separation, noise colocation, speech separation, and speech colocation.

Musical expertise and age showed significant but opposite effects on auditory working memory (musician: $\beta = 0.21$, $p < 0.001$, 95% CI = 0.09, 0.33; age: $\beta = -0.62$, $p < 0.001$, 95% CI = −0.53, −0.70), and the two factors in total explained 43% of variance in auditory working memory. Auditory working memory in turn played a significant contribution to SIN performance under noise separation ($\beta = -0.33$, $p < 0.001$, 95% CI = −0.53, −0.12), noise colocation ($\beta = -0.37$, $p < 0.001$, 95% CI = −0.56, −0.15) and speech colocation ($\beta = -0.38$, $p < 0.001$, 95% CI = −0.55, −0.18) conditions, but only marginally predicted SIN threshold under speech separation ($\beta = -0.20$, $p = 0.056$, 95% CI = −0.43, 0.04).

The direct effect of musical expertise on SIN threshold was significant under noise separation ($\beta = -0.25$, $p = 0.001$, 95% CI = −0.37, −0.11), noise colocation ($\beta = -0.16$, $p = 0.033$, 95% CI = −0.29, −0.02) and speech separation ($\beta = -0.20$, $p = 0.027$, 95% CI = −0.37, −0.02) conditions, but not significant under speech colocation condition ($\beta = -0.08$, $p = 0.26$, 95% CI = −0.21, 0.06). The indirect effect of musical expertise on SIN threshold was significant under noise separation ($\beta = -0.07$, $p < 0.001$, 95% CI = −0.14, −0.03), noise colocation

TABLE 2. Pearson or Spearman (Labeled *) Partial Correlation Coefficients and Corresponding *p* Values (in Parenthesis) Between Speech-in-Noise Threshold and Variables Including Years of Training, Total Hours of Training Stroop Score, and Nonverbal IQ, After Controlling for Hearing and Age

		Speech-in-Noise Threshold			
		Noise Separation	Noise Colocation	Speech Separation	Speech Colocation
Years of training	Older musicians	0.00 (0.975)*	−0.11 (0.464)*	−0.17 (0.248)*	−0.26 (0.077)*
	Young musicians	0.04 (0.801)	0.20 (0.181)*	−0.02 (0.871)	−0.09 (0.574)
Hours of training	Young musicians	−0.20 (0.184)*	−0.13 (0.394)*	−0.05 (0.766)*	−0.09 (0.554)*
Stroop	Older adults	0.15 (0.194)	0.18 (0.117)	0.07 (0.553)	0.08 (0.476)
Nonverbal IQ	Young adults	0.01 (0.905)	−0.07 (0.540)*	0.00 (0.977)	0.07 (0.558)

TABLE 3. The Amount of Spatial Release From Masking in Decibels (Values in Parenthesis Are Standard Deviations) Under Noise and Speech Masker Conditions Across Subject Groups

Masker Type	Older Nonmusicians	Older Musicians	*t*(*p*)	Young Nonmusicians	Young Musicians	*t*(*p*)
Noise	0.84 (0.93)	1.42 (0.92)	−2.68 (0.009)	1.10 (0.78)	1.02 (0.79)	0.40 (0.690)
Speech	5.68 (1.34)	5.51 (1.17)	0.57 (0.572)	3.55 (2.14)	3.99 (1.28)	−1.09 (0.281)
t(*p*)	−17.30 (<0.001)	−20.09 (<0.001)		−5.41 (<0.001)	−14.26 (<0.001)	

Two-tailed paired t tests and independent two-sample t tests were used for examining the differences between maskers and groups, respectively.

($\beta = -0.08$, $p = 0.001$, 95% CI = −0.15, −0.02) and speech colocation ($\beta = -0.08$, $p < 0.001$, 95% CI = −0.15, −0.03) conditions, and marginally significant under speech separation ($\beta = -0.04$, $p = 0.094$, 95% CI = −0.12, 0.01).

The direct effect of age on SIN threshold was significant under speech colocation ($\beta = 0.29$, $p < 0.001$, 95% CI = 0.12, 0.46), but not under the other three conditions. The indirect effect of age on SIN threshold was significant under noise separation ($\beta = 0.21$, $p = 0.004$, 95% CI = 0.07, 0.34), noise colocation ($\beta = 0.23$, $p = 0.001$, 95% CI = 0.09, 0.36), and speech colocation ($\beta = 0.23$, $p < 0.001$, 95% CI = 0.11, 0.36) conditions, but not significant under speech separation ($\beta = 0.12$, $p = 0.093$, 95% CI = −0.02, 0.27).

Backward digit span alone was also put into the model as a mediator, which showed similar results as above (see Figure S2 in Supplemental Digital Content 1, http://links.lww.com/EANDH/A686).

DISCUSSION

The present study revealed that lifelong musical training of older adults was associated with strengthened perception of speech sentences under "cocktail party" scenarios (with either speech or noise maskers) almost to the same level as young listeners, but such a musician advantage was absent in young adults. Compared with older nonmusicians, older musicians also exhibited better auditory working memory indexed by auditory digit span which correlated with years of training and SIN performance in older but not young participants. Importantly,

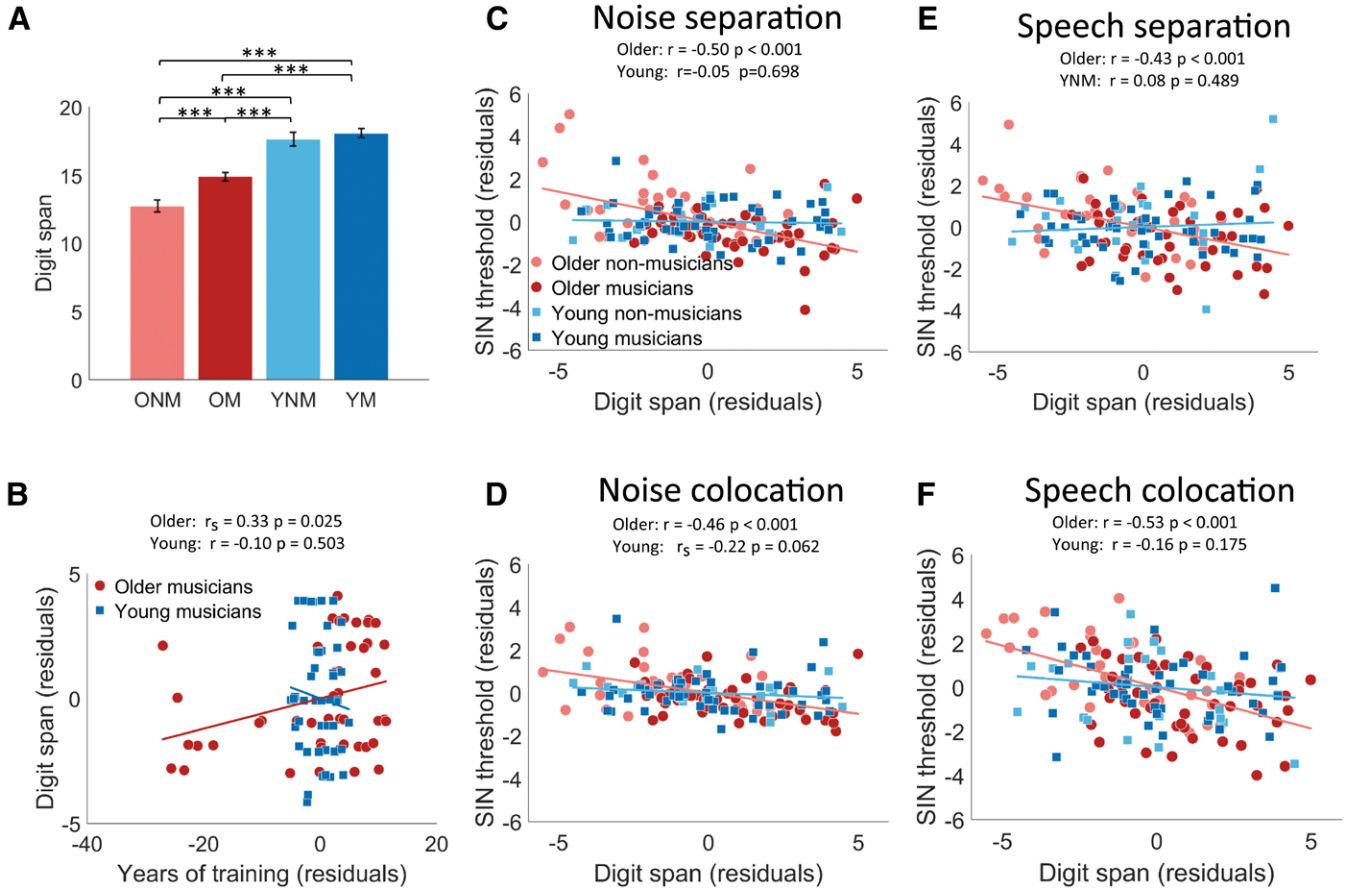

Fig. 5. (A) Auditory working memory (sum of forward and backward digit span) in older nonmusicians (ONM), older musicians (OM), young nonmusicians (YNM), and young musicians (YM). Error bars indicate standard error of the mean. *** $p < 0.001$, one-way ANOVA followed by Tukey's multiple comparison tests. (B) Pearson and Spearman partial correlations between auditory working memory and years of training in young musicians and older musicians after controlling for age, respectively. (C–F) Pearson and Spearman partial correlations between auditory working memory and speech-in-noise threshold under noise separation (C), noise colocation (D), speech separation (E), and speech colocation (F) after controlling for hearing level and age.

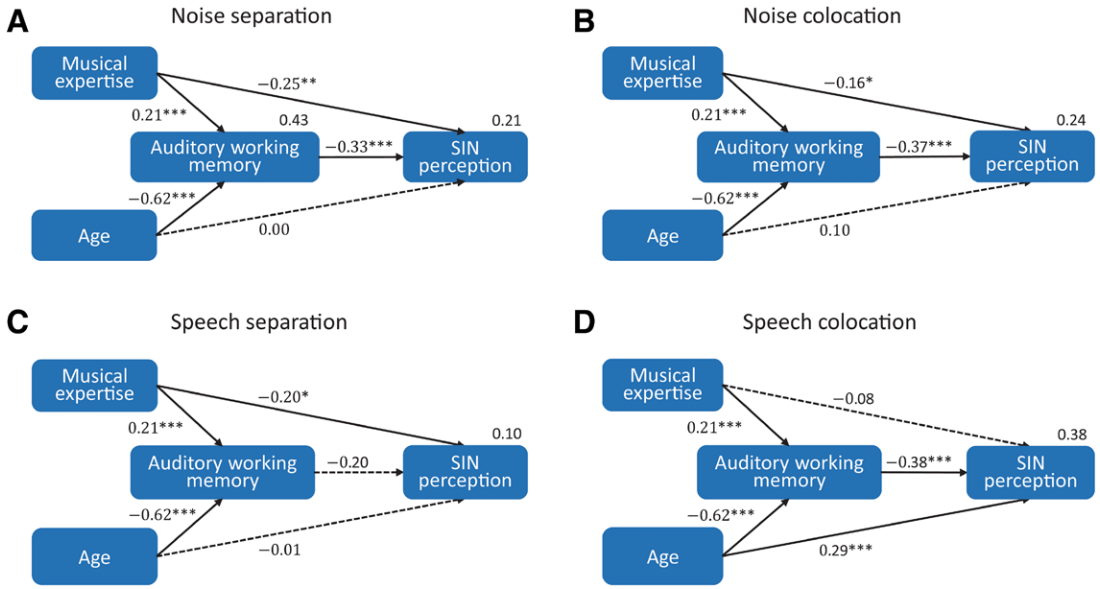

Fig. 6. Path models showing the effects of musical expertise (0, nonmusician; 1, musician) and age (0, young; 1, older) on speech-in-noise perception threshold via auditory working memory (sum of forward and backward digit span) as the mediator under four conditions: (A) noise separation, (B) noise colocation, (C) speech separation, and (D) speech colocation. Dotted lines indicate insignificant paths. Standardized path coefficients are displayed on direct paths. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

the path analysis of musical training and aging effects on SIN performance pointed to a critical mediation role of auditory working memory that may underlie a higher capacity for neurocognitive changes and a higher resilience to age-related SIN deficits in response to lifetime musical experience. In addition, the type of musical training did not substantially influence SIN performance, although older instrumentalists showed greater auditory working memory than older vocalists.

Difficulty in perceiving and comprehending speech in the absence of measurable hearing loss, especially when there are competing sound sources, is a ubiquitous part of aging. While musical training is believed to hold promise for delaying decline in cognitive functions later in life (Hanna-Pladdy & MacKay, 2011; Balbag et al., 2014), our finding that musicians excelled nonmusicians in SIN performance in older rather than young adults fits well with the differential preservation pattern of such a musician benefit (Alain et al., 2014). Indeed, only differential preservation indicates a protective effect of musical training against aging and an accumulating benefit of musical experience over time (Salthouse, 2006). Note that, regardless of listening effort which was not measured in this study, musicianship almost fully counteracted aging effect on speech perception threshold under three masking conditions except for the most difficult speech colocation condition, although older musicians' peripheral hearing was significantly worse than that of young adults. This suggests a powerful protective mechanism on central auditory and cognitive functions by long-term musical training for older adults in understanding speech at adverse listening conditions. In addition, although musician advantage in SIN perception has been repeatedly demonstrated in young adults, at levels from phonemes to sentences (Coffey et al., 2017b), the lack of musician effect in young adults here is consistent with the negative findings in recent studies when speech sentences were masked by speech spectrum noise or babble speech (Fuller et al., 2014; Boebinger et al., 2015; Escobar et al., 2020) and when nonsense sentences were masked by steady-state or amplitude-modulated noise (Ruggles et al., 2014). While hearing is a sense, listening is a skill that depends on higher-order cognition, such as working memory and attention, in grouping and segregating auditory streams (Alain et al., 2014). This is especially the case for SIN perception at the sentence level. Our result is therefore important to solve the mystery why musician advantage was observed in some studies but not others. Consistent with other recent studies (Ruggles et al., 2014; Boebinger et al., 2015; Escobar et al., 2020), the lack of training effect in young adults suggests that when musicians and nonmusicians are matched on cognitive abilities such as auditory working memory and nonverbal IQ, the group difference in SIN performance will not arise, at least not for SIN tasks where performance is measured via sentence recall. Whether the type of speech materials, such as sensical vs. nonsense sentences, will affect the musician advantage needs further investigation.

Auditory working memory—the ability to actively store behaviorally relevant sound information "in mind" over a period of seconds—has been found to play a central role in SIN perception in young, middle-aged and older adults (Parbery-Clark et al., 2011; Anderson et al., 2013; Yeend et al., 2019; Escobar et al., 2020). Although improved auditory working memory has been found in young and middle-aged musicians which correlates with better SIN performance (Parbery-Clark et al. 2009, 2011; Yoo & Bidelman 2019), the equalized auditory working memory and SIN perception between young musicians and nonmusicians in the current work echoes prior negative findings in young adults (Boebinger et al. 2015; Escobar et al., 2020). Importantly, despite the fact that older adults with short-term or lifetime musical training showed improvement in SIN perception (Zendel & Alain, 2012; Dubinsky et al., 2019; Zendel et al., 2019), no study has directly tested whether long-term musical training is related to enhanced auditory working memory which in turn contributes to better SIN perception in older adults. Here,

we first demonstrated a positive correlation between years of musical training and auditory working memory as well as correlations between auditory working memory and SIN thresholds in older adults, which replicate previous studies in young adults (Parbery-Clark et al., 2009; Yoo & Bidelman, 2019). Moreover, we used path analysis to directly reveal the indirect effects of age and musical expertise on SIN performance using auditory working memory as a mediator. Except for the speech colocation condition, auditory working memory played a full mediation role in age-related deficit of SIN perception, which suggests that auditory working memory was the most important mediator for speech understanding in healthy older adults with normal peripheral hearing below 3 kHz and normal cognitive ability in general (MOCA score ≥ 26). The effect sizes of the path coefficients show that the effect of musical experience was smaller than that of age on auditory working memory, namely, musical training could only counteract the deteriorated auditory working memory with aging to some extent, rather than fully reversing it. Since working memory training is well known for inefficiency and a recent study failed to find the transfer effect of a short-term adaptive working memory training to SIN perception in older adults (Wayne et al., 2016), musical training is more promising than pure working memory training in offsetting the SIN decline with aging.

Auditory working memory played a partial mediation role in musical training-related SIN enhancement in three conditions except for speech colocation, which means that other factors may exist as mediators. Since both cognition and auditory central processing supported SIN perception in older adults (Anderson et al., 2013), cognitive abilities like selective attention (Strait & Kraus, 2011) and nonverbal IQ (Ruggles et al., 2014; Boebinger et al., 2015), auditory skills like pitch discrimination (Dubinsky et al., 2019) and faithful encoding of speech spectrotemporal features (Kraus & Chandrasekaran, 2010), as well as sensorimotor integration (Du & Zatorre, 2017) are likely to play a mediation role in musical improvement of SIN perception. In particular, the contribution of selective attention to musical training-related SIN benefit could be inferred from the fact that the direct effect of musical experience on SIN thresholds was larger under spatial separation conditions than that under colocation conditions ($\beta = -0.25$ vs. $\beta = -0.16$ for noise masker; $\beta = -0.20$ vs. $\beta = -0.08$ for speech masker). When there was a perceived spatial separation between maskers and target speech, listeners could take advantage of location information to segregate streams in auditory scene analysis and pay selective attention to target's location which releases speech perception from masking (Bregman, 1990; Wu et al., 2005). As musical training is associated with better auditory selective attention (Strait & Kraus, 2011), musicians could take better advantage of spatial attention in facilitating SIN perception under separation conditions than nonmusicians, leading to a larger direct effect of musical training on SIN performance in addition to the indirect effect mediated by auditory working memory.

In addition, how different kinds of musical training affect SIN perception in different age groups is getting attention recently. In Slater and Kraus (2016), the only study that has compared different musician types in perceiving SIN, young percussionists but not vocalists outperformed young nonmusicians in SIN perception, with no significant difference between percussionists and vocalists. Consistent with their finding, we did not find a significant effect of musician type on SIN performance in either young or older adults, although the percussionist advantage in relative to nonmusicians was not replicated in young adults here. The discrepancy may due to balanced cognitive abilities including auditory working memory and nonverbal IQ in our young groups, and distinct rhythmic and prosodic characteristics between English and Mandarin Chinese (e.g., nontonal language vs. tonal language, Liang & Du, 2018). Note that we did not directly compare percussionists and vocalists in our dataset, we could not rule out the possibility that listeners may benefit more from rhythmic training than melodic training in understanding SIN under certain conditions. As for older adults, although both benefit by short-term vocal training (Dubinsky et al., 2019) and no effect of short-term piano training (Fleming et al., 2019) on SIN performance have been revealed, no prior study has directly tested the effect of training type on SIN processing. Here, we provide the first evidence that long-term vocal training was equally effective as long-term instrumental training in offsetting the age-related SIN deficit, even when vocalists received fewer years of training than instrumentalists (42 vs. 51.5 years). Although superior inhibitory control was found in young percussionists in comparison to young vocalists (Slate et al., 2017), we did not observe any difference in inhibitory control between older instrumentalists and older vocalists. Interestingly, different from our hypothesis, older vocalists had worse auditory working memory than older instrumentalists even after matching years of training, although years of training may be confounded by training frequency and intensity. Since almost all vocalists played in chorus, it would be an interesting question whether choir sing provides older adults extra advantage in stream segregation and selective attention to human voice in a multispeaker senario in addition to auditory working memory, which contributes to equivalent SIN perception as instrumental training. Future research is need to refine the impacts of different musicianship on SIN perception using large sample size, various training and playing styles, different SIN materials, paradigms and cognitive measures, and wide age groups. Nonetheless, instead of training type, what matters more in the current study is whether extensive musical training could ameliorate speech comprehension difficulties for older adults, and the answer is yes.

Although the present findings provide evidence that auditory working memory plays a mediating role in explaining the musician benefit on SIN perception in older adults, some limitations exist. First, while path analysis is a great approach for inferring the causal relationships among multiple variables, this method cannot affirm the causal link. Longitudinal studies with delicate design (e.g., Randomized Controlled Trial) are required to verify the causal contribution of musical training in offsetting the age-related decline in working memory and SIN perception. Second, the sample size (n = 149) was just acceptable for the path analysis. It is recommended to use sample sizes of at least 20 or 40 cases per parameter (Kline 2011). Further studies should enlarge the sample size and include more variables into the path model to comprehensively investigate the relationships between age, musical training, cognitive abilities (auditory working memory, nonverbal IQ, etc.), auditory skills, life experience, and SIN abilities. Third, the musician types were not matched in young and older adults. More types of musical training and playing styles (e.g., playing by ear vs. reading from a score, solo vs. playing with others) should be considered in different age groups to better understand how various musical

experiences impact distinct aspects of speech perception (e.g., rhythm vs. prosody) and cognitive abilities (e.g., auditory working memory, attention, inhibitory control). For instance, if auditory working memory is truly a key mediator in explaining the musical training effect in counteracting age-related SIN deficit, one would expect that musicians who are good at playing by ear would have better auditory working memory and SIN performance than musicians who cannot. Fourth, older vocalists and older instrumentalists were not matched in training length. Although in older musicians years of training did not correlate with SIN threshold under all four conditions, it predicted auditory working memory. Future research needs to disentangle the effects of training type versus the amount of training on SIN perception as well as cognitive abilities. Moreover, individuals with a wide range of training length should be included and the amount of training should be refined in total hours instead of years to better explore the effects of training volume and age of training onset on SIN perception. Last, due to the language difference, our results relating to the musician type effect on SIN perception may not directly be generalized to other populations, including speakers of nontonal languages.

Music is the art that almost everyone can appreciate and participate in. Besides the musical reward and esthetic experience, musical training could provide potential benefits to speech comprehension and cognition, especially for the elderly. This study sheds some light on the causal relationship that musical training mitigates the age-related decline in understanding speech in noisy situations by improving strategic listening skills represented by auditory working memory. These findings support musical training as an intervention to slow or attenuate cognitive decline and communication difficulty that often emerge later in life.

ACKNOWLEDGMENTS

This research was supported by grants from the National Natural Science Foundation of China (Grant No. 31671172 and 31822024), the "Thousand Talent Program for Young Outstanding Scientists," and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB32010300).

L. Z., X. Y. F., D. L., and L. D. S. X. performed the testing and data collection. L. Z. conducted the data analyses, contributed to the interpretation of the results, and drafted the manuscript. Y. D. designed the study, contributed to the interpretation of the results, and wrote the manuscript. All authors approved the final version of the manuscript for submission.

The data that support the findings of this study are available upon request from Dr. Yi Du (duyi@psych.ac.cn).

The authors have no conflicts of interest to declare.

Address for correspondence: Yi Du, Institute of Psychology Chinese Academy of Sciences, 16 Lincui Road, Chaoyang, Beijing, 100101, China. E-mail: duyi@psych.ac.cn.

Received October 30, 2019; accepted June 7, 2020.

REFERENCES

Alain, C., Zendel, B. R., Hutka, S., Bidelman, G. M. (2014). Turning down the noise: The benefit of musical training on the aging auditory brain. *Hear Res, 308*, 162–173.

Anderson, S., & Kraus, N. (2010). Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *J Am Acad Audiol, 21*, 575–585.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear Res, 300*, 18–32.

Balbag, M. A., Pedersen, N. L., Gatz, M. (2014). Playing a musical instrument as a protective factor against dementia and cognitive impairment: A population-based twin study. *Int J Alzheimers Dis, 2014*, 836748.

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci, 35*, 1240–1249.

Boebinger, D., Evans, S., Rosen, S., et al. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J Acoust Soc Am, 137*, 378–387.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.

Cattell, R., & Cattell, A. (1960). *Handbook for the Individual or Group Culture Fair Intelligence Test*. Institute for Personality and Ability Testing.

Coffey, E. B. J., Chepesiuk, A. M. P., Herholz, S. C., et al. (2017a). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Front Neurosci, 11*, 479.

Coffey, E. B. J., Mogilever, N. B., Zatorre, R. J. (2017b). Speech-in-noise perception in musicians: A review. *Hear Res, 352*, 49–69.

Coffey, E. B. J., Herholz, S. C., Scala, S., et al. (2011). Montreal Music History Questionnaire: A tool for the assessment of music-related experience in music cognition research. In The Neurosciences and Music IV: Learning and Memory, Conference, Edinburgh, UK, June 9–12.

Dubinsky, E., Wood, E. A., Nespoli, G., Russo, F. A. (2019). Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. *Front Neurosci, 13*, 1153.

Du, Y., Buchsbaum, B. R., Grady, C. L., Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci USA, 111*, 7126–7131.

Du, Y., Buchsbaum, B. R., Grady, C. L., Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat Commun, 7*, 12241.

Du, Y., Kong, L., Wang, Q., et al. (2011). Auditory frequency-following response: A neurophysiological measure for studying the "cocktail-party problem." *Neurosci Biobehav Rev*, *35*, 2046–2057.

Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proc Natl Acad Sci USA, 114*, 13579–13584.

Escobar, J., Mussoi, B. S., Silberer, A. B. (2020). The effect of musical training and working memory in adverse listening situations. *Ear Hear*, *41*, 278–228.

Faul, F., Erdfelder, E., Buchner, A., et al. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods*, *41*, 1149–1160.

Fleming, D., Belleville, S., Peretz, I., et al. (2019). The effects of short-term musical training on the neural processing of speech-in-noise in older adults. *Brain Cogn, 136*, 103592.

Freyman, R. L., Balakrishnan, U., Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *J Acoust Soc Am, 109*(5 Pt 1), 2112–2122.

Fuller, C. D., Galvin, J. J. 3rd, Maat, B., et al. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Front Neurosci, 8*, 179.

Gong, Y. X. (1992). *Manual of Wechsler Adult Intelligence Scale-Chinese Version*. Chinese Map Press.

Gordon-Salant, S., & Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear Hear, 37*, 593–602.

Hanna-Pladdy, B., & Mackay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology*, *25*, 378–386

Helfer, K.S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *J Sp Lan Hear Res*, *40*, 432–443.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear Hear, 29*, 87–98.

Hooper, D., Coughlan, J., Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electron J Bus Res Methods*, *6*, 53–60.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). Guilford Press.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat Rev Neurosci, 11*, 599–605.

Kraus, N., Strait, D. L., Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Ann N Y Acad Sci, 1252*, 100–107.

Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Front Neurosci, 12*, 495.

Merrett, D. L., Peretz, I., Wilson, S. J. (2013). Moderating variables of music training-induced neuroplasticity: A review and discussion. *Front Psychol, 4*, 606.

Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiol Aging, 33*, 1483.e1–1483.e4.

Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear, 30*, 653–661.

Parbery-Clark, A., Strait, D. L., Anderson, S., et al. (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS One*, *6*, 1–11.

Pearl, J. (2012). The causal foundations of structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of Structural Equation Modeling (pp.* 68–91*)*. Guilford Press.

Puschmann, S., Baillet, S., Zatorre, R. J. (2019). Musicians at the cocktail party: Neural substrates of musical training during selective listening in multispeaker situations. *Cereb Cortex*, *29*, 3253–3265.

Ruggles, D. R., Freyman, R. L., Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One, 9*, e86980.

Salthouse, T. A. (2006). Mental exercise and mental aging: Evaluating the validity of the "use it or lose it" hypothesis. *Perspect Psychol Sci, 1*, 68–87.

Slater, J., Azem, A., Nicol, T., et al. (2017). Variations on the theme of musical expertise: Cognitive and sensory processing in percussionists, vocalists and non-musicians. *Eur J Neurosci, 45*, 952–963.

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cogn Process, 17*, 79–87.

Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front Psychol, 2*, 113.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J Exp Psychol*, *18*, 643–662.

Turner, C. W., & Cummings, K. J. (1999). Speech audibility for listeners with high-frequency hearing loss. *Am J Audiol, 8*, 47–56.

Uhlmann, R. F., Larson, E. B., Rees, T. S., et al. (1989). Relationship of hearing impairment to dementia and cognitive dysfunction in older adults. *JAMA*, *261*, 1916–1919.

Wallach, H., Newman, E. B., Rosenzweig, M. R. (1949). The precedence effect in sound localization. *Am J Psychol, 62*, 315–336.

Wayne, R. V., Hamilton, C., Jones Huyck, J., Johnsrude, I. S. (2016). Working memory training and speech in noise comprehension in older adults. *Front Aging Neurosci, 8*, 49.

Wilson, R. H., McArdle, R. A., Smith, S. L. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res, 50*, 844–856.

Wingfield, A., & Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *J Am Acad Audiol, 18*, 548–558.

Wolfram, S. (1992). *Mathematica: A System for Doing Mathematics by Computer: User's Guide for Microsoft Windows*. Addison-Wesley.

Wu, X., Wang, C., Chen, J., et al. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hear Res*, *199*, 1–10.

Yeend, I., Beach, E. F., Sharma, M. (2019). Working memory and extended high-frequency hearing in adults: Diagnostic predictors of speech-in-noise perception. *Ear Hear, 40*, 458–467.

Yoo, J., & Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear Res, 377*, 189–195.

Yu, J., Li, J., Huang, X. (2012). The Beijing version of the Montreal Cognitive Assessment as a brief screening tool for mild cognitive impairment: A community-based study. *BMC Psychiatry, 12*, 156.

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychol Aging, 27*, 410–417.

Zendel, B. R., & Alain, C. (2014). Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiol Aging, 35*, 55–63.

Zendel, B. R., West, G., Belleville, S., et al. (2019). Music training improves the ability to understand speech-in-noise in older adults. *Neurobiol Aging*, *81*, 102–115.