Hearing Research 377 (2019) 189–195

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception

Jessica Yoo [a], Gavin M. Bidelman [a, b, c, *]

[a] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[b] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[c] University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

ARTICLE INFO

Article history:
Received 25 September 2018
Received in revised form
20 March 2019
Accepted 27 March 2019
Available online 29 March 2019

Keywords:
Experience-dependent plasticity
Speech-in-noise (SIN) perception
Cocktail party listening
Central auditory processing
Auditory masking

ABSTRACT

Previous studies have reported better speech-in-noise (SIN) recognition in musicians relative to nonmusicians while others have failed to observe this “musician SIN advantage.” Here, we aimed to clarify equivocal findings and determine the most relevant perceptual and cognitive factors that do and do not account for musicians' benefits in SIN processing. We measured behavioral performance in musicians and nonmusicians on a battery of SIN recognition, auditory backward masking (a marker of attention), fluid intelligence (IQ), and working memory tasks. We found that musicians outperformed nonmusicians in SIN recognition but also demonstrated better performance in IQ, working memory, and attention. SIN advantages were restricted to more complex speech tasks featuring sentence-level recognition with speech-on-speech masking (i.e., QuickSIN) whereas no group differences were observed in non-speech simultaneous (noise-on-tone) masking. This suggests musicians' advantage is limited to cases where the noise interference is linguistic in nature. Correlations showed SIN scores were associated with working memory, reinforcing the importance of general cognition to degraded speech perception. Lastly, listeners' years of music training predicted auditory attention scores, working memory skills, general fluid intelligence, and SIN perception (i.e., QuickSIN scores), implying that extensive musical training enhances perceptual and cognitive skills. Overall, our results suggest (i) enhanced SIN recognition in musicians is due to improved parsing of competing linguistic signals rather than signal-in-noise extraction, *per se,* and (ii) cognitive factors (working memory, attention, IQ) at least partially drive musicians' SIN advantages.

1. Introduction

The human brain is capable of structural and functional changes. As a model of neuroplasticity, musical experience has been shown to influence cognitive functions related to music and language processing (Moreno and Bidelman, 2014; Schön et al., 2004). Most notably, musicians' benefits in speech-language processing have been widely studied through behavioral and electrophysiological studies [(Bidelman and Krishnan, 2010; Chan et al., 1998; Moreno and Bidelman, 2014; Parbery-Clark et al., 2009; Shahin et al., 2003; Slevc and Miyake, 2006; Swaminathan et al., 2015); for review see Moreno and Bidelman (2014)]. Musicians' perceptual advantages with speech, for instance, are thought to stem from their long-term training that enhances auditory perceptual skills (Rammsayer and Altenmüller, 2006) and hones top-down processing (Strait et al., 2010) to enable finer detection of subtle changes in pitch, timbre, and timing of complex auditory signals (Kraus and Chandrasekaran, 2010).

Among its putative auditory-linguistic benefits, musicianship has been associated with enhanced speech-in-noise (SIN) recognition (Anaya et al., 2016; Bidelman and Krishnan, 2010; Clayton et al., 2016; Deroche et al., 2017; Du and Zatorre, 2017; Mankel and Bidelman, 2018; Parbery-Clark et al., 2011; Swaminathan et al., 2015); for review see Coffey et al. (2017)]. For example, amateur musicians with ~10 years of training show superior identification and discrimination of target speech amidst acoustic interferences including reverberation (Bidelman and Krishnan, 2010) and noise babble (Parbery-Clark et al., 2009). Bidelman and Krishnan (2010) found that perceptual discrimination for voice

---

[*] Corresponding author. School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN, 38152, USA.
E-mail address: gmbdlman@memphis.edu (G.M. Bidelman).

https://doi.org/10.1016/j.heares.2019.03.021
0378-5955/

pitch and formant cues was roughly 2–4 times better (i.e., smaller difference limens) in musically-trained listeners than nonmusician controls for vowel sounds presented in both quiet and in the presence of reverberation. Complementary effects were reported by Parbery-Clark et al. (2009) who observed that musicians could tolerate ~1 dB more noise on the Hearing in Noise Test (HINT) than their nonmusician peers during degraded speech recognition. Similar results were replicated using the QuickSIN (Mankel and Bidelman, 2018; Zendel and Alain, 2012) and spectrotemporally-degraded non-speech stimuli (Başkent et al., 2018; Fuller et al., 2014). While the significance of such benefits may not be readily apparent, a 1 dB change in signal-to-noise ratio (SNR) can equate to an improvement in speech recognition performance by as much as 10–15% (Middelweerd et al., 1990). Extending these behavioral results, Zendel et al. (2015) measured auditory evoked potentials in musicians and nonmusicians while they performed SIN tasks. They found that the N400 (related to lexical-semantic access) showed noise-related changes in nonmusicians while it remained stable in musicians, suggesting superior SIN processing and higher-level lexical access in musically trained ears. Collectively, these studies imply that musicianship is associated with improved cocktail party listening skills.

However, other studies have reported contradictory results and failed to find a musician SIN advantage (Boebinger et al., 2015; Madsen et al., 2017; Ruggles et al., 2014; Yeend et al., 2017). For example, using an informational masking paradigm, Boebinger et al. (2015) showed that musicians performed better than nonmusicians in frequency discrimination tasks but not in masked-speech perception. A significant relationship between IQ and SIN performance was also observed, indicating that cognitive skills play a substantial role in SIN perception. Furthermore, Ruggles et al. (2014) showed that musicians displayed higher performance in pitch discrimination but no advantage in deciphering voiced or whispered nonsense sentences in noise (energetic masking) nor clinical SIN tests (i.e., QuickSIN, HINT). In contrast to Boebinger et al. (2015), Ruggles and colleagues found no relationship between IQ and SIN reception thresholds (Ruggles et al., 2014). These findings have led some investigators to suggest musicians' auditory benefits may be rooted principally in pitch perception, rather than advantages in global cognition (Fuller et al., 2014).

On the contrary, a growing number of studies have reported superior cognitive abilities in musicians including working memory (WM) performance (Bidelman et al., 2013; Talamini et al., 2016) and cognitive control (Pallesen et al., 2010). Using Baddeley's framework of working memory (Baddeley and Hitch, 1974), Roden et al. (2014) found that children who underwent weekly music training showed a significant benefit in cognitive performance in auditory processing, especially in tasks tapping the phonological loop and the capacity to store and manipulate auditory-verbal information. Along with enhanced WM, music training has also been associated with positive changes in intelligence (IQ) and executive function (Bugos et al., 2007; Degé et al., 2011; Moreno et al., 2011; Schellenberg, 2006, 2011). Relevant to the present study, enhanced auditory WM in musicians also correlates with their SIN performance (Grassi et al., 2017; Parbery-Clark et al., 2009, 2011). Thus, musicians' SIN advantages may not be rooted in auditory, linguistic, or speech abilities, *per se*. Rather, music-related SIN benefits might be epiphenomenal, a byproduct of general cognitive abilities (Moreno and Bidelman, 2014; Patel, 2011) fortified by their years of experience integrating tactile, visual, auditory, and motor information necessary for producing music (Zatorre, 2005).

From the extant literature, the conditions under which one observes a "musician advantage" in SIN processing remain equivocal. To this end, we aimed to reevaluate musicians' SIN benefits, placing new attention on how other aspects of cognitive performance might relate to degraded speech processing. Specifically, we examined performance across a battery of speech and non-speech masking tasks in musician and nonmusician listeners. Our experimental design adopted the paradigm from Krizman et al. (2017), who assessed (in bilinguals) masking under tasks differing in the "linguistic" content of the noise interference. This allowed us to assess musicians' signal-in-noise perception along a quasi-continuum where the masker interference varied from a speech (multi-talker babble) to non-speech (bandpass noise) signal. We also included a series of cognitive tests (IQ, attention, WM) to assess factor(s) that might contribute to musicians' benefits in SIN perception. We hypothesized that if musicians have superior auditory skills overall, they would display higher performance compared to nonmusicians on all perceptual masking tasks, regardless of whether the masker was a speech or non-nonspeech signal. Alternatively, if musicians SIN benefits are driven more by cognitive abilities, their performance should differ from nonmusicians only under noise conditions that carried linguistically-relevant information (e.g., speech-on-speech masking).

2. Methods

2.1. Participants

Thirty-one young (age range: 18–35 years; *mean* ± *SD*: 25.4 ± 4.2 years), normal-hearing adults were recruited for the study. The sample was divided into two groups based on musical experience. Sixteen musicians (*M;* 10 females, 6 males) were defined as individuals with at least 8 years of continuous training (15.81 ± 4.82 yrs) on a musical instrument starting before age 10 (7.0 ± 2.37 yrs). Fifteen nonmusicians (*NM;* 10 females, 5 males) were defined as individuals with $\leq$3 years (0.37 ± 0.79 yrs) of lifetime music training on any combination of instruments. All participants showed normal hearing sensitivity (puretone audiometric thresholds $\leq$ 25 dB HL; 250–8000 Hz), had no previous history of brain injury or psychiatric problems, and were English-speaking monolinguals with no fluency in other languages. The two groups were otherwise matched in right-handedness (Oldfield, 1971) as measured by the Edinburgh Handedness inventory [$t(29) = -1.17$ $p = 0.25$], gender (Fisher's exact test: $p = 1.0$), formal education [$t(29) = 1.73$, $p = 0.09$], and socioeconomic status [$t(29) = 1.11$, $p = 0.28$; scored based on highest level of parental education: 1(high school without diploma or GED) - 6(doctoral degree)] (Mankel and Bidelman, 2018; Norton et al., 2005). Each completed a written informed consent form approved by the Institutional Review Board at The University of Memphis and were compensated $10/hr for their time.

2.2. Psychophysical and cognitive measures

Overview. Speech in noise recognition was evaluated via the QuickSIN (Niquette et al., 2001), Words-in-Noise test (WIN) (Wilson et al., 2007), and Hearing-in-Noise Test (HINT) (Nilsson et al., 1994). We used multiple speech-in-noise tests to parametrically vary task complexity and assess the "musician SIN advantage" using a continuum of non-speech to speech maskers. This design allowed us to identify which types of SIN tasks (competing maskers) yield perceptual enhancements in musicians (for comparable study design in bilinguals, see Krizman et al., 2017). Cognitive skills (i.e., fluid intelligence and working memory) were evaluated using Raven's Progressive Matrices (Raven, 1998) and backwards digit span (Wechsler et al., 2008), respectively. Backwards and simultaneous masking tasks were used as a proxy measure of auditory attention (Krizman et al., 2017; Strait et al., 2010). All auditory tasks were presented binaurally via

Sennheiser HD 280 circumaural headphones. Two repetitions were run for each task (described below). The first was used as a familiarization phase and subsequent analyses were conducted on the second run.

QuickSIN. Lists of six English sentences spoken by a female talker were presented amidst a background of four-talker babble noise. Target sentences were presented at 70 dB HL and the signal-to-noise ratio (SNR) decreased parametrically in 5 dB steps from 25 dB SNR to 0 dB SNR. At each SNR, participants were instructed to verbally repeat the sentence. Correctly recalled keywords were logged by the tester. Following the QuickSIN documentation, we computed the final SNR loss score by subtracting the number of correctly recalled target words from 25.5 (i.e., SNR loss = 25.5-Total Correct). The QuickSIN provided a non-adaptive measure of noise-degraded sentence recognition.

Hearing in noise test (HINT). The HINT consisted of lists of 20 simple sentences presented in speech-shaped noise. SNR was varied using an adaptive tracking method (4 dB-up/2 dB-down step). The noise level was set to 65 dB SPL and the target stimulus level was adjusted on each trial according to the participant's response. Participants were instructed to repeat each sentence. The SNR at threshold was measured corresponding to 50% sentence recognition performance. The HINT provided a similar measure of SIN processing as the QuickSIN but used an adaptive threshold measure and different masker characteristics (i.e., speech-shaped vs. babble noise).

Words in noise (WIN). The WIN consisted of lists of 40 monosyllabic words spoken by a female talker and were presented at 55 dB HL with four-talker babble background noise. The words were presented starting from 24 dB SNR and decreased by 4 dB steps every five words. Participants were instructed to repeat the words they heard. The final score, reflecting threshold performance, was based on the number of correctly recalled words. The WIN provided a measure of single-word perception in noise.

Backward and *simultaneous masking (BM noise, SIM).* These psychophysical tests were used to assess signal-in-noise perception devoid of linguistic content and determine how basic auditory detection for non-speech signals might play a role in influencing SIN performance. Backward masking is also known to assess top-down, cognitive processing such as attention (Strait et al., 2010; Tallal et al., 1993). Both masking paradigms were implemented in the PsyAcoustX GUI, programmed in MATLAB (The MathWorks, Inc.) (Bidelman et al., 2015). For each task, a three-interval forced choice (3IFC) adaptive tracking task was applied. Only one out of three intervals contained the probe tone (two others contained only noise), and the participants were instructed to listen and choose the interval containing the probe. A response window was shown on the computer screen to assist in visualizing the presentation order of the three intervals and logging responses. For backwards masking, the masker-target delay was 0 ms. The initial masker noise level was set at 25 dB SPL with 300 ms in duration, with the target probe tone (1000 Hz, 20 ms duration) set at 30 dB SPL. On each trial, the noise level was increased (made harder) following a correct response and decreased (made easier) following an incorrect response according to a 2-down—1 up adaptive tracking rule (Levitt, 1971). For simultaneous masking, the initial masker level was 15 dB SPL, with 300 ms in duration. The target probe tone level was the same as in the backwards masking task only the probe onset was contiguous with the masker onset. Masked threshold was then determined using an identical tracking procedure as in the backwards paradigm. For both backward and simultaneous masking, the masker was a 500–1500 Hz bandpass noise.

Raven's matrices. Raven's Progressive Matrices (Raven, 1998) was used to assess listeners' non-verbal fluid intelligence. Each question contained a 3x3 matrix of abstract patterns and shapes. Participants were instructed to select the missing pattern from one of 8 options given in the answer choices. Items became progressively more difficult and required greater reasoning ability and intellectual capacity over the course of the test. Each participant was randomly distributed one of two test versions each containing 29 questions. They were given 10 min to complete the task. Raw scores (number correct) were recorded and used in subsequent analyses.

Digits span. Backwards digit span was used to assess working memory ability. The backward digits test consisted of 7 questions with each question containing two repetitions. A series of digits was verbally presented to listeners (~1/sec) which varied in sequence length. The length started with two digits (e.g., 2, 5) and progressively increased up to eight digits (e.g., 4, 5, 2, 1, 3, 6, 2, 4) as the questions progressed. Participants were required to recall the sequence in backwards order from the presentation. The longest span length correctly recalled was recorded as individuals' auditory WM capacity.

2.3. Statistical analysis

Group differences (musician vs. nonmusician) were evaluated for each task (QuickSIN, WIN, HINT, backward masking noise, SIM masking, Raven's, Digits backward) using independent samples *t*-tests. To compare the performance between the tasks (which each have different measurement scales), raw scores were converted into z-scores to allow for a standardized comparison across tasks (Krizman et al., 2017). We then conducted a two-way, mixed-model ANOVA (group x task; subjects=random factor). Tukey-Kramer adjusted comparisons were used for multiple comparisons. Pearson correlations and regression analyses were used to determine (1) the relation between performance on the different auditory and cognitive tasks and (2) whether individuals' years of music training predicted perceptual-cognitive skills. Robust regression was performed using the 'fitlm' function in MATLAB and bisquare weighting.

3. Results

Fig. 1A–C depicts group performance in cognitive measures reflecting fluid intelligence (Raven's), auditory attention (backwards masking), and working memory (digit span), respectively. We found significant group differences between musicians and nonmusicians for Raven's scores [*M*: 24.19 ± 2.95 (score), *NM:* 22.13 ± 2.07; $t(29) = 2.23$, $p = 0.03$], backward masking [*M*: 57.81 ± 31.13 (threshold dB), *NM:* 38.48 ± 22.90; $t(29) = 1.96$, $p = 0.059$], and backward digits span [*M*: 11.63 ± 1.96 9 (score), *NM:* 7.47 ± 2.33; $t(29) = 5.39$, $p < 0.0001$]. That is, musicians demonstrated better performance than nonmusicians on cognitive measures including IQ, working memory, and attention.

Group comparisons for non-speech masking and SIN measures are shown in Fig. 1D–G. Musicians outperformed nonmusicians on the QuickSIN, achieving SIN thresholds that were ~2 dB lower [*M*: 1.56 ± 1.81 dB SNR loss dB, *NM:* 0.7 ± 1.32; $t(29) = -3.96$, $p < 0.0005$]. No group differences were observed on the HINT [*M*: 6.5 ± 2.0 dB, *NM:* 6.13 ± 2.07; $t(29) = -0.50$, $p = 0.62$], WIN [*M*: 1.60 ± 1.40 dB, *NM:* 1.84 ± 1.14; $t(29) = -0.52$, $p = 0.61$], or SIM masking thresholds [*M*: 26.88 ± 2.55 dB, *NM:* 24.07 ± 7.90; $t(29) = 1.35$, $p = 0.19$]. These results suggest a task-dependent benefit of musicianship on signal-in-noise processing that is largely limited to more complex SIN tasks requiring sentence-level recognition with linguistic maskers.

Fig. 2 shows z-transformed (normalized) scores for direct comparisons between groups on all perceptual-cognitive measures. A two-way ANOVA showed group*task interaction [$F(6,174) = 6.82$,

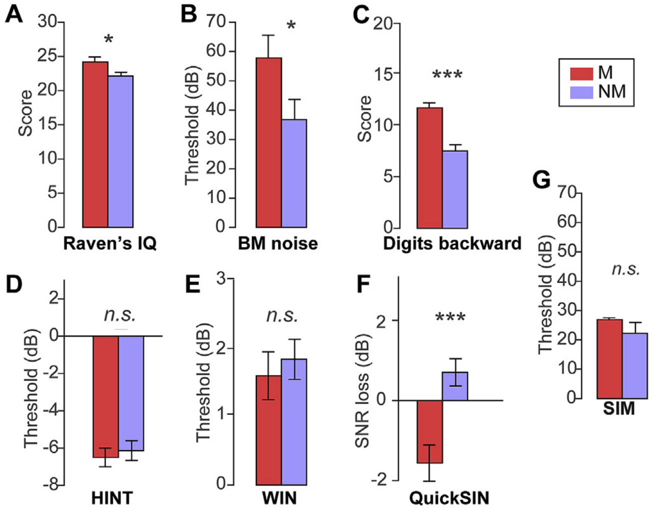

Fig. 1. Group comparisons of task performance between musicians and nonmusicians. (**A-C**) Cognitive measures. (**D-G**) SIN perceptual tasks. Group differences are observed for Raven's (IQ), backwards masking (attention), backwards digit span (WM), and QuickSIN (SIN recognition). BM noise, backward masking; HINT, hearing-in-noise test; WIN, words-in-noise test; SIM, simultaneous (noise-on-tone) masking. errorbars = ±1 s. e.m. $^{*}p < 0.05$, $^{***}p < 0.001$.

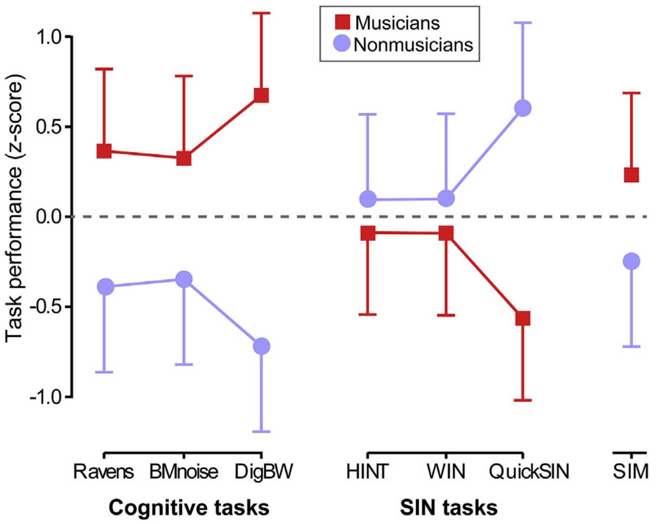

Fig. 2. Standardized (z-scored) task performance for musicians and nonmusicians. The dotted line denotes average across both groups. For “cognitive” and “SIM” tasks, higher scores reflect better performance. For the “SIN” tasks, lower scores are better. Otherwise as in Fig. 1. BMnoise, backward masking; DigBW, backward digits span; SIM, simultaneous (noise-on-tone) masking; HINT, hearing-in-noise test; WIN, words-in-noise test. errorbars = ±95% CI.

$p < 0.0001$]. Similar to the previous anlayses, multiple comparisons revealed musicians outperformed nonmusicians on backward masking [$t(174) = 2.02$, $p = 0.04$], backward digits span [$t(174) = 4.18$, $p < 0.0001$], QuickSIN [$t(174) = -3.50$, $p < 0.0006$], and Raven's [$t(174) = 2.26$, $p = 0.02$].

We then evaluated links between musical training, SIN perception, and perceptual-cognitive skills via correlational analyses (Figs. 3–4). These analyses focused on associations with the QuickSIN as this was the only SIN test that showed significant group differences (see Fig. 1). We found a significant correlation between QuickSIN scores and digits backward score (Fig. 3C; $r = -0.50$, $p = 0.0043$), confirming well-established links between working memory and degraded speech perception abilities (Dryden et al., 2017; Füllgrabe and Rosen, 2016; Pisoni and Geers, 2000). No other tasks showed correlations with QuickSIN scores nor was backwards WM span correlated with HINT or WIN scores. In addition to these task-task correlations, years of music training was positively correlated with backward masking thresholds (Fig. 4A; $r = 0.48$, $p = 0.0063$), such that longer music engagement was associated with superior auditory attention. Musical training also predicted backward digits WM (Fig. 4B; $r = 0.72$, $p < 0.0001$), QuickSIN (Fig. 4C; $r = -0.60$, $p < 0.0001$),[1] and Raven's scores (Fig. 4D; $r = 0.36$, $p = 0.04$). Collectively, these results demonstrate that listeners with longer durations of music training have enhanced auditory-perceptual and cognitive abilities that enables better QuickSIN recognition, auditory working memory, attention, and fluid IQ.

Lastly, we assessed correlations between QuickSIN and musical training, partialing out backward digits scores (WM) which also predicted SIN processing (see Fig. 3C). Musicianship and SIN performance remained correlated even after controlling for WM ($r = -0.39$, $p = 0.0311$). In terms of variance explained (R^2), this indicates that years of musical training accounts for ~15% ($R^2 = 0.152$) of the variance in SIN scores after accounting for WM, a drop from 36% ($R^2 = 0.36$; Fig. 4C) when WM is not factored into the model.

4. Discussion

Results of the current study relate to three main observations: (1) Musicians have enhanced SIN processing but this benefit is limited to speech-on-speech masking rather than figure-ground perception, *per se*; (2) working memory is strongly associated with SIN performance; (3) musicianship was predictive of not only of certain aspects of SIN processing but attention, WM, and IQ, suggesting that both auditory perceptual and broader domain-general cognitive skills are malleable to experience-dependent plasticity.

4.1. Musicians' advantages in noise-degraded hearing are limited to linguistically relevant signals/maskers

Our findings show that musicians have enhanced SIN processing. However, this benefit is largely restricted to specific speech-on-speech masking conditions where *both* the target signal and interference are linguistically relevant. Despite similar maskers between the QuickSIN and WIN, only the former produced group differences. This pattern was also observed by Krizman et al. (2017) in their assessment of bilinguals' SIN perception and may be related to differences in the complexity of the targets (QuickSIN: sentences; WIN: monosyllabic words) and/or contextual cues available in the QuickSIN but not WIN. Indeed, the much better (lower) SIN thresholds for the QuickSIN compared to WIN suggest this much (cf. Fig. 1E vs. F). The QuckSIN might also be more sensitive at revealing group differences in speech recognition than other SIN tests (Wilson et al., 2007). Arguably, the QuickSIN recruits more working memory (Parbery-Clark et al., 2009) than the HINT or WIN since it comprises more complex, longer, and harder to predict sentences (Wilson et al., 2007). Thus, our data suggest that musicians' SIN advantage is limited to conditions with *linguistic* maskers that arguably involve heavier use of cognitive functions. This notion is supported by the findings of Anaya et al. (2016), who observed that musicians outperformed nonmusicians in recognizing speech

---

[1] Correlations between musical training and the other SIN tests were not significant (WIN: $r = 0.00$, $p = 0.98$; HINT: $r = 0.01$, $p = 0.96$).

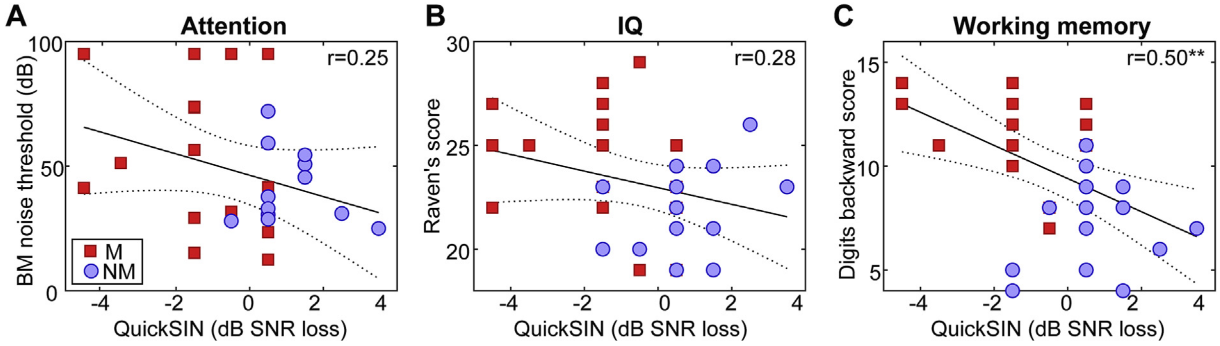

Fig. 3. Degraded speech perception is related to auditory working memory but not IQ and attention. QuickSIN scores are not related to **(A)** attention or **(B)** IQ. (**C**) Lower (i.e., better) QuickSIN scores predict larger auditory working memory capacity (i.e., longer backwards digit span). Solid lines, regression fits; dotted lines; 95% CI intervals. $^{**}p < 0.01$.

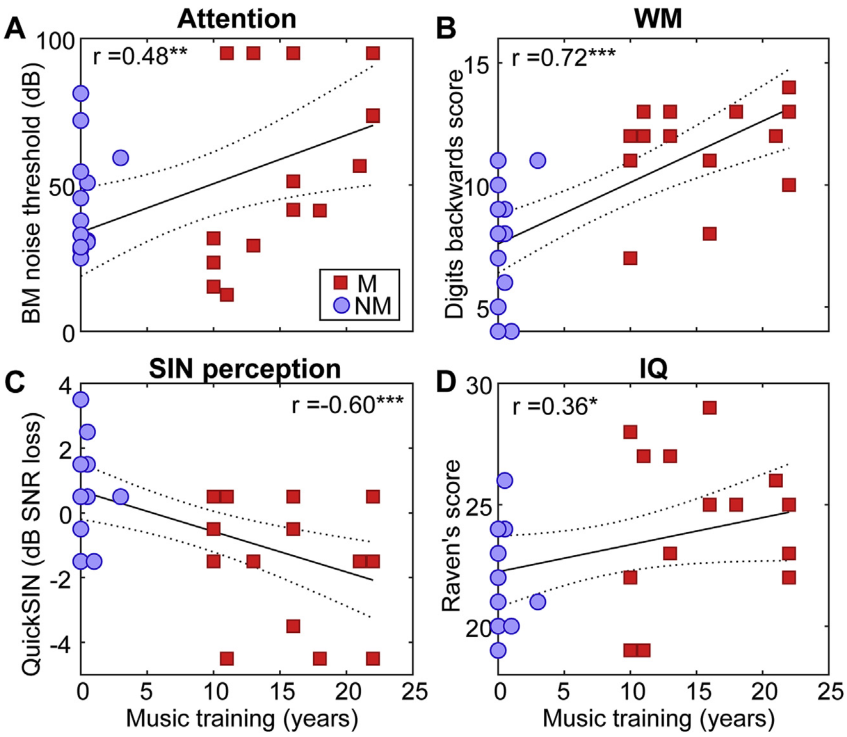

Fig. 4. Years of formal music training predicts musicians' perceptual-cognitive advantages in (**A**) auditory attention, (**B**) working memory, (**C**) SIN recognition, and (**D**) IQ. Both M and NMs are included as training was considered a continuous variable in these analyses. Longer duration of training is associated with superior auditory attention, larger WM capacity, SIN recognition at lower SNRs, and higher fluid intelligence. Solid lines, regression fits; dotted lines; 95% CI intervals. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

but not environmental (non-speech) sounds in noise. Indeed, other studies that have failed to find SIN advantages in musicians have used less complex sentence material (Boebinger et al., 2015), which may account for the absence of masked-speech benefit.

In contrast, our data reveal a musician QuickSIN benefit on the order of ~2 dB SNR, in agreement with prior studies (Mankel and Bidelman, 2018; Zendel and Alain, 2012). The discrepancy between studies utilizing the QuickSIN is unclear but may be due to differences between the musician demographics and/or test administration [e.g., eight (Ruggles et al., 2014) vs. two (present study) blocks of testing]. For example, our shorter QuickSIN assessment may have enabled participants to be more attentive during testing thereby revealing SIN benefits compared to longer paradigms that may fatigue listeners (Ruggles et al., 2014). Additionally, musicians may be more susceptible to noise damage (Phillips et al., 2010) (but see Bidelman et al., 2017; Brashears et al., 2003; Schmidt et al., 2014) and undocumented "hidden hearing loss" in previous studies on musicians may diminish their putative benefit in SIN listening (Skoe et al., 2018). However, we find this notion unlikely as there is currently little evidence that so-called hidden hearing loss relates to SIN processing, let alone exists in human listeners (e.g., Guest et al., 2017; Johannesen et al., 2019). Moreover, there is some suggestion that musicianship enhances cochlear gain control which may actually help protect the ear against acoustic overexposure (Bidelman et al., 2017). While our data suggest a SIN advantage in musicians, we qualify this finding with the following points: (i) although behaviorally relevant to speech recognition (Middelweerd et al., 1990), musicians' SIN advantage is small (i.e., several dB SNR); (ii) the benefit is largely circumscribed to speech-on-speech conditions that require the parsing of multiple linguistically-relevant signals (e.g. Anaya et al., 2016; Swaminathan et al., 2015).

Different masker characteristics might also account for musicians' SIN benefits. Maskers in our psychophysical masking tasks, WIN, HINT, and QuickSIN varied in their spectrotemporal characteristics (e.g., simultaneous masking = bandpass noise; QuickSIN = multi-talker babble) and thus, their degree of energetic masking (EM). EM is related to the interference of cochlear excitation patterns of the signal and masker and thus, is thought to reflect peripheral hearing function. In contrast, informational masking (IM) is defined as the non-energetic aspect of masking interference that occurs for similar/confusable target and masker sounds (e.g., speech-on-speech) and represents central-cognitive aspects of figure-ground perception (Moore, 2012). In this vein, Swaminathan et al. (2015) recently demonstrated that musicians' SIN benefits are stronger when the signal and masker are both speech and contain a higher level of IM. Our data agree with these findings by demonstrating a musician advantage only in the QuickSIN (which has high levels of linguistic/IM masking) compared to simple noise-on-tone tasks (largely EM). The fact that musicians' SIN advantages are most prevalent under conditions of IM is further supported by the lack of group difference in the HINT, which uses a speech-shaped noise masker and thus more EM compared to the QuickSIN (where group differences emerged). Along these lines, Oxenham et al. (2003) concluded that musicians were less affected by IM, which may be due to their enhanced analytic listening skills, and suggested that factors such as attention and expectation about the signal aid their performance. Musician advantages are most apparent when tasks tap broad cognitive processes such as executive function and attention (Bialystok and DePape, 2009). Thus, the fact that musicians are able to better juggle two different speech streams with higher levels of IM (as in the QuickSIN) suggests their SIN benefits are aided not by enhanced auditory figure-ground perception *per se* (cf. Bidelman et al., 2014) but superior domain-general cognitive mechanisms (Bialystok and DePape, 2009; Moreno and Bidelman, 2014; Sares et al., 2018; Strait and Kraus, 2011).

4.2. Auditory vs. cognitive mechanisms underlying musicians' SIN advantage

Our data identify clear links between auditory and cognitive abilities when it comes to musicians' SIN advantage. We find that duration of musical training predicted not only SIN perception but also cognitive measures which implies that these perceptual-cognitive skills may be driven by experience-dependent plasticity. Furthermore, we found correlations between listeners' QuickSIN scores and (i) their musical training and (ii) WM. This three way relation makes it difficult to tease apart whether music training or WM, *per se*, drives SIN performance. Comparing correlation models helps tease apart this three-way relation; although variance explained dropped (36%–15%) when WM was taken out of the model, music training and SIN performance remained correlated suggesting that factors beyond (or in addition to) WM contribute to the relation between musicianship and SIN benefits. Schellenberg (2011) found that musicianship was associated with IQ and Digit Span (WM and attention), but not other executive function tests. Strait and Kraus (2011) also demonstrated relationships between auditory attention and SIN performance (see also Sares et al., 2018). IQ, WM, and attention presumably play a large role in SIN. Indeed, we find musician enhancements in all three of these factors (Fig. 1), each of which might aid their degraded speech-listening skills.

Nevertheless, other factors (not tested here) might also contribute to musicians' benefits in SIN processing. Namely, enhancements in basic auditory perceptual abilities (e.g., pitch and spectrotemporal discrimination) have been noted in several psychophysical reports (Bidelman and Krishnan, 2010; Bidelman et al., 2011, 2014; Kishon-Rabin et al., 2001; Micheyl et al., 2006). Additionally, vocabulary knowledge may differ in musicians which might partially account for their SIN benefits observed here (Anaya et al., 2016). Alternatively, musicians' enhanced SIN processing could also relate to innate auditory skills, irrespective of musical training or experience, *per se* (Bidelman and Mankel, 2019; Mankel and Bidelman, 2018). To the extent that musician SIN benefits do exist, future studies are needed to fully tease apart the relative contributions of psychophysical, cognitive, and pre-existing factors underlying musicians' perceptual advantages (cf. Mankel and Bidelman, 2018). It would also be interesting to determine if musician enhancements in SIN perception observed here exist in more realistic "cocktail party" scenarios with a competitive mixture of talkers (e.g., 3D acoustic environment). Current studies are underway in our laboratory to test this possibility.

Collectively, our findings align with notions that the plasticity associated with musicianship extends beyond the auditory domain to improve broader cognitive functions (Moreno and Bidelman, 2014). For example, in addition to enhanced WM and IQ observed in this and previous studies (Bidelman et al., 2013; Bugos et al., 2007; Degé et al., 2011; Moreno et al., 2011; Schellenberg, 2006, 2011; Talamini et al., 2016), musical training has been associated with enhanced audiovisual processing (Bidelman, 2016; Lee and Noppeney, 2011) and even the perception of *visually* degraded text (Anaya et al., 2016). Our results are best cast in a framework which views music-related plasticity as a *multidimensional continuum* of cognitive transfer effects. This suggests the repeated exposure and experience with manipulating sound patterns tunes not only auditory but also broader analytic skills by enhancing domain general functions (e.g., WM, attention executive processing) (Moreno and Bidelman, 2014). Under this model, the amount of benefit from music to linguistic processing (i.e., SIN perception) should be mediated by an individual's general cognitive capacity. While we find evidence for this proposition in the current data (i.e., musicians' better WM and IQ scores and their correlations with SIN performance) further studies are needed to fully test this possibility as well as determine the dosage of music training necessary for such cognitive benefits to emerge (Alain et al., 2014; Kraus et al., 2014).

Acknowledgments

This work was supported by grant NIH/NIDCD R01DC016267 awarded to G.M.B. The authors thank Caitlin Price and Kelsey Mankel for comments on earlier versions of this manuscript.

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.

Anaya, E.M., Pisoni, D.B., Kronenberger, W.G., 2016. Long-term musical experience and auditory and visual perceptual abilities under adverse conditions. J. Acoust. Soc. Am. 140, 2074–2081.

Baddeley, A.D., Hitch, G., 1974. Working memory. In: Bower, G.H. (Ed.), Psychology of Learning and Motivation, vol. 8. Academic Press, pp. 47–89.

Başkent, D., Fuller, C.D., Galvin III, J.J., Schepel, L., Gaudrain, E., Free, R.H., 2018. Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents. J. Acoust. Soc. Am. 143, EL311–EL316.

Bialystok, E., DePape, A.-M., 2009. Musical expertise, bilingualism, and executive functioning. J. Exp. Psychol. Hum. Percept. Perform. 35, 565.

Bidelman, G.M., 2016. Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. Exp. Brain Res. 234, 3037–3047.

Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355, 112–125.

Bidelman, G.M., Mankel, K., 2019. Reply to Schellenberg: is there more to auditory plasticity than meets the ear? Proc. Nat. Acad. Sci. U. S. A. 116, 2785–2786 [Letter to the Editor].

Bidelman, G.M., Krishnan, A., Gandour, J.T., 2011. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. Eur. J. Neurosci. 33, 530–538.

Bidelman, G.M., Hutka, S., Moreno, S., 2013. Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. PLoS One 8, e60676.

Bidelman, G.M., Jennings, S.G., Strickland, E.A., 2015. PsyAcoustX: a flexible MATLAB((R)) package for psychoacoustics research. Front. Psychol. 6, 1498.

Bidelman, G.M., Schug, J.M., Jennings, S.G., Bhagat, S.P., 2014. Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. J. Acoust. Soc. Am. 136, EL33–EL39.

Bidelman, G.M., Schneider, A.D., Heitzmann, V.R., Bhagat, S.P., 2017. Musicianship enhances ipsilateral and contralateral efferent gain control to the cochlea. Hear. Res. 344, 275–283.

Boebinger, D., Evans, S., Rosen, S., Lima, C.F., Manly, T., Scott, S.K., 2015. Musicians and non-musicians are equally adept at perceiving masked speech. J. Acoust. Soc. Am. 137, 378–387.

Brashears, S.M., Morlet, T.G., Berlin, C.I., Hood, L.J., 2003. Olivocochlear efferent suppression in classical musicians. J. Am. Acad. Audiol. 14, 314–324.

Bugos, J.A., Perlstein, W.M., McCrae, C.S., Brophy, T.S., Bedenbaugh, P.H., 2007. Individualized Piano Instruction enhances executive functioning and working memory in older adults. Aging Ment. Health 11, 464–471.

Chan, A.S., Ho, Y.-C., Cheung, M.-C., 1998. Music training improves verbal memory. Nature 396, 128.

Clayton, K.K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A.D., Kidd Jr., G., 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. PLoS One 11, e0157638.

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

Degé, F., Kubicek, C., Schwarzer, G., 2011. Music lessons and intelligence: a relation mediated by executive functions. Music Percept. 29, 195.

Deroche, M.L., Limb, C.J., Chatterjee, M., Gracco, V.L., 2017. Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. J. Acoust. Soc. Am. 142, 1739–1755.

Dryden, A., Allen, H.A., Henshaw, H., Heinrich, A., 2017. The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic literature review and meta-analysis. Trends Hear. 21, 2331216517744675.

Du, Y., Zatorre, R.J., 2017. Musical training sharpens and bonds ears and tongue to hear speech better. Proc. Natl. Acad. Sci. Unit. States Am. 201712223.

Fuller, C.D., Galvin III, J.J., Maat, B., Free, R.H., Başkent, D., 2014. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? Front. Neurosci. 8, 179.

Füllgrabe, C., Rosen, S., 2016. On the (un) importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. Front. Psychol. 7, 1268.

Grassi, M., Meneghetti, C., Toffalini, E., Borella, E., 2017. Auditory and cognitive performance in elderly musicians and nonmusicians. PLoS One 12, e0187881.

Guest, H., Munro, K.J., Prendergast, G., Howe, S., Plack, C.J., 2017. Tinnitus with a normal audiogram: relation to noise exposure but no evidence for cochlear synaptopathy. Hear. Res. 344, 265–274.

Johannesen, P.T., Buzo, B.C., Lopez-Poveda, E.A., 2019. Evidence for age-related cochlear synaptopathy in humans unconnected to speech-in-noise intelligibility deficits. Hear. Res. 374, 35–48.

Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y., 2001. Pitch discrimination: are professional musicians better than non-musicians? J. Basic Clin. Physiol. Pharmacol. 12, 125–144.

Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci. 11, 599.

Kraus, N., Slater, J., Thompson, E.C., Hornickel, J., Strait, D.L., Nicol, T., White-Schwoch, T., 2014. Music enrichment programs improve the neural encoding of speech in at-risk children. J. Neurosci. 34, 11913–11918.

Krizman, J., Bradlow, A.R., Lam, S.S.-Y., Kraus, N., 2017. How bilinguals listen in noise: linguistic and non-linguistic factors. Biling. Lang. Cognit. 20, 834–843.

Lee, H., Noppeney, U., 2011. Long-term music training tunes how the brain temporally binds signals from multiple senses. Proc. Natl. Acad. Sci. Unit. States Am. 108, E1441–E1450.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467–477.

Madsen, S.M.K., Whiteford, K.L., Oxenham, A.J., 2017. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. Sci. Rep. 7, 12624.

Mankel, K., Bidelman, G.M., 2018. Inherent auditory skills rather than formal music training shape the neural encoding of speech. Proc. Natl. Acad. Sci. U.S.A. 115, 13129–13134.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219, 36–47.

Middelweerd, M.J., Festen, J.M., Plomp, R., 1990. Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. Audiology 29, 1–7.

Moore, B.C., 2012. An Introduction to the Psychology of Hearing. Brill.

Moreno, S., Bidelman, G.M., 2014. Examining neural plasticity and cognitive benefit through the unique lens of musical training. Hear. Res. 308, 84–97.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E.G., Cepeda, N.J., Chau, T., 2011. Short-term music training enhances verbal intelligence and executive function. Psychol. Sci. 22, 1425–1433.

Nilsson, M., Soli, S.D., Sullivan, J.A., 1994. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. 95, 1085–1099.

Niquette, P., Gudmundsen, G., Killion, M., 2001. QuickSIN Speech-In-Noise Test Version 1.3. Etymotic Research, Elk Grove Village, IL.

Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D.J., Schlaug, G., 2005. Are there pre-existing neural, cognitive, or motoric markers for musical ability? Brain Cogn. 59, 124–134.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Oxenham, A.J., Fligor, B.J., Mason, C.R., Kidd Jr., G., 2003. Informational masking and musical training. J. Acoust. Soc. Am. 114, 1543–1549.

Pallesen, K.J., Brattico, E., Bailey, C.J., Korvenoja, A., Koivisto, J., Gjedde, A., Carlson, S., 2010. Cognitive control in auditory working memory is enhanced in musicians. PLoS One 5, e11120.

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., Hittner, E., Kraus, N., 2011. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. PLoS One 6, e18082.

Patel, A.D., 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front. Psychol. 2, 142.

Phillips, S.L., Henrich, V.C., Mace, S.T., 2010. Prevalence of noise-induced hearing loss in student musicians. Int. J. Audiol. 49, 309–316.

Pisoni, D.D., Geers, A.E., 2000. Working memory in deaf children with cochlear implants: correlations between digit span and measures of spoken language processing. Ann. Otol. Rhinol. Laryngol. Suppl. 185, 92.

Rammsayer, T., Altenmüller, E., 2006. Temporal information processing in musicians and nonmusicians. Music Percept.: Interdisciplinary J. 24, 37.

Raven, J., 1998. Raven Manual Section 4: Advanced Progressive Matrices. Oxford Psychologists Press, Oxford.

Roden, I., Grube, D., Bongard, S., Kreutz, G., 2014. Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. Psychol. Music 42, 284–298.

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.

Sares, A.G., Foster, N.E.V., Allen, K., Hyde, K.L., 2018. Pitch and time processing in speech and tones: the effects of musical training and attention. J. Speech Lang. Hear. Res. 61, 496–509.

Schellenberg, E.G., 2006. Long-term positive associations between music lessons and IQ. J. Educ. Psychol. 98, 457.

Schellenberg, E.G., 2011. Music lessons, emotional intelligence, and IQ. Music Percept. 29, 185–194.

Schmidt, J.H., Pedersen, E.R., Paarup, H.M., Christensen-Dalsgaard, J., Andersen, T., Poulsen, T., Baelum, J., 2014. Hearing loss in relation to sound exposure of professional symphony orchestra musicians. Ear Hear. 35, 448–460.

Schön, D., Magne, C., Besson, M., 2004. The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology 41, 341–349.

Shahin, A., Bosnyak, D.J., Trainor, L.J., Roberts, L.E., 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J. Neurosci. 23, 5545–5552.

Skoe, E., Camera, S., Tufts, J., 2018. Noise exposure may diminish the musician advantage for perceiving speech in noise. Ear Hear. https://doi.org/10.1097/AUD.0000000000000665.

Slevc, L.R., Miyake, A., 2006. Individual differences in second-language proficiency: does musical ability matter? Psychol. Sci. 17, 675–681.

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.

Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear. Res. 261, 22–29.

Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V., Kidd Jr., G., Patel, A.D., 2015. Musical training, individual differences and the cocktail party problem. Sci. Rep. 5, 11628.

Talamini, F., Carretti, B., Grassi, M., 2016. The working memory of musicians and nonmusicians. Music Percept. 34, 183–191.

Tallal, P., Miller, S., Fitch, R.H., 1993. Neurobiological basis of speech: a case for the preeminence of temporal processing. Ann. N. Y. Acad. Sci. 682, 27–47.

Wechsler, D., Coalson, D.L., Raiford, S.E., 2008. WAIS-IV: Wechsler Adult Intelligence Scale. Pearson, San Antonio, TX.

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.

Yeend, I., Beach, E.F., Sharma, M., Dillon, H., 2017. The effects of noise exposure and musical training on suprathreshold auditory processing and speech perception in noise. Hear. Res. 353, 224–236.

Zatorre, R., 2005. Music, the food of neuroscience? Nature 434, 312.

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

Zendel, B.R., Tremblay, C.-D., Belleville, S., Peretz, I., 2015. The impact of musicianship on the cortical mechanisms related to separating speech from background noise. J. Cogn. Neurosci. 27, 1044–1059.