Check for updates

Auditory working memory mechanisms mediating the relationship between musicianship and auditory stream segregation

Martha Liu[1]*, Isabelle Arseneau-Bruneau[1,2], Marcel Farrés Franch[1,2], Marie-Elise Latorre[1], Joshua Samuels[1], Emily Issa[1], Alexandre Payumo[1], Nayemur Rahman[1], Naíma Loureiro[1], Tsz Chun Matthew Leung[3], Karli M. Nave[3], Kristi M. von Handorf[3], Joshua D. Hoddinott[3], Emily B. J. Coffey[4], Jessica Grahn[3] and Robert J. Zatorre[1,2]*

[1]Montreal Neurological Institute, McGill University, Montreal, QC, Canada, [2]Centre for Research in Brain, Language and Music, Montreal, QC, Canada, [3]Department of Psychology and Centre for Brain and Mind, Western University, London, ON, Canada, [4]Department of Psychology, Concordia University, Montreal, QC, Canada

This study investigates the interactions between musicianship and two auditory cognitive processes: auditory working memory (AWM) and stream segregation. The primary hypothesis is that AWM could mediate a relationship between musical training and enhanced stream segregation capabilities. Two groups of listeners were tested: the first aimed to establish the relationship between the three variables, and the second aimed to replicate the effect in an independent sample. Music experience history and behavioral data were collected from a total of 145 healthy young adults with normal binaural hearing. The AWM task involved the manipulation of tonal patterns in working memory, while the Music-in-Noise Task (MINT) measured stream segregation abilities in a tonal context. The MINT expands measurements beyond traditional Speech-in-Noise assessments by capturing auditory subskills (rhythm, visual, spatial attention, prediction) relevant to stream segregation. Our results showed that musical training is associated with enhanced AWM and MINT performance and that this effect is replicable across independent samples. Moreover, we found in both samples that the enhancement of stream segregation was largely mediated by AWM capacity. The results suggest that musical training and/or aptitude enhances stream segregation by way of improved AWM capacity.

KEYWORDS

auditory stream segregation, auditory working memory, hearing-in-noise, musical training, music perception

Introduction

Navigating the symphony of sounds that simultaneously converge upon our ears poses a multifaceted challenge to the human auditory system's ability to distinguish distinct perceptual objects (Bregman, 1990), thus playing a pivotal role in organizing our auditory perception (Shamma and Micheyl, 2010). This cognitive function is influenced by both stimulus-driven

grouping strategies (Noorden, 1975; Deroche et al., 2017; Bregman and Pinker, 1978) and cognitive top-down factors (Anderson et al., 2013; Davis and Johnsrude, 2007; Thompson et al., 2011). At the cognitive level, stream segregation involves various factors such as the listener's attention and attentional load (Heinrich et al., 2008; Thompson et al., 2011), prior knowledge (Davis and Johnsrude, 2007), inhibitory control (Lewis et al., 2021; Stenbäck et al., 2022), and schematic expectations (Bey and McAdams, 2002). In particular, auditory working memory (AWM), the active mental workspace that allows the temporary storage and manipulation of short-term acoustic information (Baddeley, 1992), has been suggested to play a crucial role in auditory stream segregation (Bey and McAdams, 2002; Heinrich et al., 2008; Dalton et al., 2009; Escobar et al., 2020), accounting for individual differences in this capacity (Gordon-Salant and Cole, 2016; Parbery-Clark et al., 2009a).

The influence of musicianship on AWM and SIN perception

Musicians have emerged as a distinctive population of interest due to their constant exposure and attunement to complex auditory patterns (Herholz and Zatorre, 2012; Brown et al., 2015). Musical activities such as practice and performance are proposed to lead to improved stream segregation abilities (Swaminathan et al., 2015; for review, see Coffey et al., 2017) and enhanced working memory, especially for tonal stimuli (for meta-analysis, see Talamini et al., 2017; for review, see Yurgil et al., 2020).

Traditionally, the relationship between musicianship, AWM, and stream segregation has been examined using a variety of Speech-in-Noise (SIN) tests (Nilsson et al., 1994; Killion et al., 2004). Many, though not all, studies have reported that musical training is correlated with a better perception of speech-in-noise (for review, see Coffey et al., 2017). Parbery-Clark et al. (2009b) specifically reported a strong relationship between AWM and SIN perceptual abilities across age groups in musicians, suggesting that the AWM enhancement of musicians mediates their better performance in SIN.

Several studies suggest the possibility that AWM may be related to SIN performance in musicians. Research using Mandarin nonsense sentence stimuli has shown a mediating role of AWM in ameliorating SIN perception loss in older, but not younger musicians, as demonstrated through path analysis (Zhang et al., 2021). Other research reported musicians' SIN advantage and correlation between SIN scores and working memory, although the associations are limited to cases where the noise induces linguistic interference (Yoo and Bidelman, 2019). Escobar et al. (2020) reported that after equating for AWM capacity, there was no difference between musicians and non-musicians; however AWM was correlated with performance on several SIN tests.

The varied findings in SIN tests could be related to variations in task design, criteria for musicianship, and different scoring methods (for further explanation, see Coffey et al., 2019). More critically, these SIN tasks fall short of providing the granularity required to assess individual perceptual components and top-down cues involved in stream segregation, which could potentially be affected by training or other interventions. Furthermore, SIN assessments in prior studies exclusively focused on sentence or word detection, which limits the generalizability of the findings on hearing-in-noise to the speech modality alone.

Music-in-Noise Task

The Music-in-Noise Task (MINT) is a stream segregation paradigm designed to eliminate linguistic influences, expand measures beyond speech perception, and assess different top-down processes (Coffey et al., 2019). By using a melodic target embedded within a mix of musical sounds as informational masking, MINT enables the systematic disentangling of critical auditory sub-skills involved in effective stream segregation (Slater and Kraus, 2016; Coffey et al., 2019), including rhythmic, visual, spatial attentional, and predictive cues. Paralleling the findings in SIN research, Coffey et al. (2019) reported significant correlations between cumulative musical practice hours and music-in-noise perception, particularly in rhythm, prediction, and visual conditions. The study also showed a significant relationship between AWM and overall MINT performance. However, AWM capability in that study was only accounted for as a covariate in analyzing musical training's impact on MINT sub-conditions, along with other factors such as pitch discrimination and multilingualism. Consequently, there remains a gap in the literature regarding the interaction between musical training, AWM, and music-in-noise perception.

Specific aims and hypothesis

The goal of the present study was (1) to determine if the purported musician advantage in auditory stream segregation could be consistently observed, and (2) specifically to test the hypothesis that such an effect is mediated by enhanced AWM. We implemented a test-replication research design where the same study was conducted in two phases with independent samples. This approach allows for testing the robustness of the findings across cohorts of different distributions of musicianship. In Experiment 1 (Initial Phase) the phenomenon of interest was identified and analyzed. Experiment 2 (Replication Phase) tested whether the initial findings could be replicated in a more heterogeneous sample, thus ensuring that the observed effects are robust and not solely related to the specific sample used in the first phase.

To test the effects of music training on both MINT and AWM, we carried out correlational analyses using cumulative practice hours as the independent variable; for additional verification and to account for possible nonlinear effects, we also carried out categorical comparisons of musicians vs. non-musicians. We hypothesized a positive relationship between musical training and AWM and MINT task performance. Finally, we aimed to test the hypothesis that musical training fosters improvements in MINT through the enhancement of AWM capabilities, as suggested but not fully confirmed by the literature, positioning AWM as a mediating factor in this relationship. We therefore used statistical mediation analysis to understand the underlying process by which musical training influences music-in-noise perception, delineating direct and indirect effects through the mediator (AWM).

Experiment (1)

Methods and materials

Participants

In the initial phase, we recruited 82 healthy young adults with either minimal or extensive piano experience. Participants were recruited from various advertisement sources (social media, flyers, etc.), while some expert musicians were specifically recruited through a snowball sampling method. As part of a broader study not detailed here, participants completed a comprehensive battery of tasks and were compensated with cash for their overall time. All Participants had completed at least 1 year of university-level education, and their demographic information is provided in Table 1. To conduct group comparisons on the effects of musical training, we defined subjects with >10 cumulative years of music training and > 4,000 h of lifetime practice as Musicians (N = 42), and subjects with <2 years of musical activity as Non-Musicians (N = 20) (Table 1).

Subjects provided informed consent and were compensated for their participation and time. All experimental procedures were approved by the McGill University Faculty of Medicine Research Ethics Board. All participants were screened to have normal or corrected-to-normal vision and reported no history of neurological disorders. Normal binaural hearing was confirmed by an audiometric test which measured pure-tone thresholds from 250 to 8,000 Hz (less than 25 dB SL). Participants with binaural hearing thresholds above 25 dB HL did not proceed with the study as deficiencies in the frequency range may influence their task performance. Out of the 82 participants from Experiment 1 who completed all parts of the study, 4 were excluded from the MINT analysis due to their inability to process basic musical content (with 2 or more out of 6 incorrect responses for the MINT task Control condition, see description below).

Procedure

Prior to the testing session, participants confirmed eligibility and completed the Montreal Music History Questionnaire (MMHQ) (Coffey et al., 2011). The MMHQ provides the subject's self-reported information regarding overall musical experience (instruments played, total cumulative practice hours), language proficiency, basic demographics, etc. The tasks were administered in the context of a larger test battery that will not be reported here. Each testing session began with an audiometry hearing test, followed by a series of behavioral tasks, including the AWM task (Albouy et al., 2017) and the MINT task (Coffey et al., 2019); see the following section for descriptions. The visual component of each task was presented on a computer screen and sounds were presented binaurally through headphones (ATH-M50x, Audio-Technica). A comfortable sound level set at 73 dB was determined during pilot testing and kept constant for all subjects and both tasks.

Measures and Behavioral Tasks

(1) To test for individual AWM abilities and eliminate linguistic influences, we implemented an AWM task that measures individuals' auditory retention and manipulation capabilities with sets of tonal stimuli (Albouy et al., 2017). This AWM task uses a discrimination design that involves the detection of a local pitch change within two tonal patterns differing in temporal order, described as the "Manipulation Task" in Albouy et al. (2017). On each trial, participants first listened to three sequentially presented 250 ms tones, which were followed after a 2000 ms silent retention interval by a probe consisting of another set of three tones (Figure 1). The task was to determine whether the sequence of the second set of three tones was a perfect reverse of the first set or not. The structure of this task engages AWM capabilities, requiring participants to retain the initial set of tones and inversely manipulate them in their mental workspace during the retention interval (Albouy et al., 2017; Foster et al., 2013; Zatorre et al., 2010). Six practice trials with feedback were provided, followed by 100 experimental trials without feedback. Task trials are randomized with a maximum of 3 consecutive trials with the same condition. The average accuracy score was then computed based on the percentage of responses correct.

(2) The Music-in-Noise Task (MINT) assesses stream segregation, involving the detection of a target musical melody embedded in irrelevant musical background noise (Coffey et al., 2019). Employing a match-mismatch discrimination design, each trial features one melodic line embedded in masking noise, and a melodic line presented in silence (Figure 2). Participants were asked to judge if the two presented melodies were the same or different. The MINT consists of five conditions which capture auditory sub-skills and the influence of perceptual cues: (1) Baseline (Pitch; Figure 2A), where the target-noise mixture is first presented, followed by the comparison melody in silence, without additional cues; (2) Rhythm (Figure 2B), the target is a rhythmic pattern with no pitch variation; (3) Spatial (Figure 2D), an additional spatial attentional cue is presented for the participant to attend to sounds coming from their left or right side (the perception of which is manipulated via interaural sound level difference); (4) Visual (Figure 2E), an additional visual cue outlining the melody's contour is presented to facilitate target detection within the mixture; and (5) Prediction (Figure 2C), subjects hear the target melody in silence first, followed by the comparison melody in noise. There is also a control condition with both melodies presented in silence to screen out participants incapable of discriminating the musical content of the MINT task, and who may therefore have amusia (Peretz et al., 2002).

TABLE 1 Experiment 1.

Variable	Total (N = 82)	Musicians (N = 42)	Non-musicians (N = 20)
Age (mean ± SD)	25.5 ± 6.8	24.3 ± 5.6	29.2 ± 7.8
Age range	18–45	18–44	21–45
Sex	30 males, 51 females, 1 non-binary	16 males, 25 females, 1 non-binary	7 males, 13 females
Cumulative practice hours (mean ± SD)	5,300 ± 5,900	9,400 ± 5,700	90 ± 200
Cumulative practice range	0–30,000	4,100–30,000	0–620
Age of onset (mean ± SD)	N/A	5.0 ± 1.3	N/A

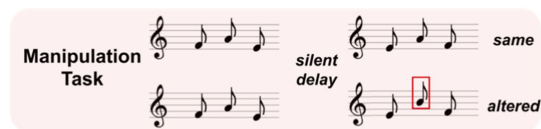Descriptive statistics of demographic and musical training variables.

FIGURE 1
Illustration of AWM task (adapted from Albouy et al., 2017). "Match" trials: the second sequence of melody was presented in a reversed temporal order of the first melody; "mismatch" trials: the second melody was presented in reversed temporal order, with one local pitch change. This required the retention and manipulation of auditory information.
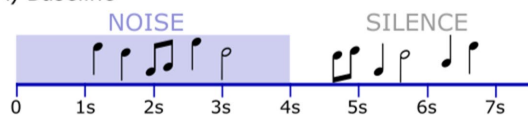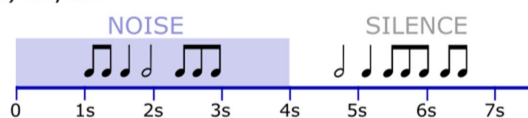
FIGURE 2
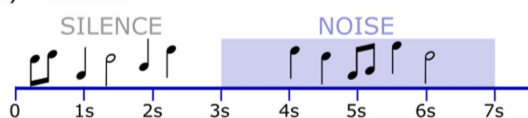Illustration for MINT (adapted from Coffey et al., 2019). "Match" trials: the melody mixed in noise is identical to the melody presented in silence; "mismatch" trials: the melody mixed in noise is not identical to the melody presented in silence. MINT consists of five conditions: **(A)** Baseline (Pitch), **(B)** Rhythm, **(C)** Prediction, **(D)** Spatial, and **(E)** Visual. In the Spatial condition **(D)**, an icon on one side of the screen directed the listener to attend to the corresponding ear. In the Visual condition **(E)**, a scrolling graphic representation outlines the timing and melodic contour of the target melody.

All conditions were tested at three different signal-to-noise (SNR) levels (0, −3, and −6 dB). Each condition involved 2 practice trials, followed by 20 experimental trials presented in a randomized block order across subjects. The accuracy score for each individual condition and overall performance is calculated by averaging the percentage of correct responses across all SNR levels within the respective condition(s); and the accuracy score for performance at each SNR level is computed by averaging the percentage of correct responses across all conditions at that specific SNR level (for further procedural details, see Coffey et al., 2019).

Data analysis

Data analyses were conducted using IBM SPSS Statistics (version 29.0.2.0) to perform correlation and mediation analyses. To examine the correlational relationships between cumulative practice hours, AWM, and MINT scores, both parametric (Pearson's r) and non-parametric (Spearman's rho) tests were conducted. For consistency with the mediation analysis, which uses raw values rather than ranks, only Pearson's correlation coefficients are reported. Nonetheless, all tests produced comparable significant results (see Supplementary material for non-parametric correlations). Comparisons between Experiment 1 and Experiment 2 were performed using independent samples *t*-tests, while comparisons between Musicians and Non-Musicians in each experiment were conducted using the Mann–Whitney U test, concerning the non-normal distribution and small sample size.

PROCESS macro (version 4.2 beta release) for SPSS by A. F. Hayes was used for mediation analysis. PROCESS is an observed variable ordinary least square and logistic regression path analysis tool that provides estimation of direct and indirect effects within both single and multiple mediator models (Preacher and Hayes, 2004). It allows us to estimate the conditional indirect effects of AWM as a mediator between music training and MINT outcomes. All mediation models were tested for statistical significance through bootstrapping with 5,000 resamples, conducted with replacement, with significance determined by confidence intervals (Preacher and Hayes, 2008). Bootstrapping is a common procedure in mediation analysis that enhances statistical power and robustness against non-normal distributions, small sample sizes, and outliers.

Results

Musical training and AWM

Descriptive statistics for AWM score (% correct) for all participants, as well as for the Musician and Non-Musician groups, are presented in Table 2. Pearson correlation indicates a significant relationship between cumulative hours of practice and AWM task performance ($r = 0.399$, $p < 0.001$; Figure 3A; see Supplementary Table S1 for Spearman's results). Mann–Whitney U test indicates a significant difference between Musicians and Non-Musicians groups on AWM score ($U = 792$, $p < 0.001$) (Figure 3C).

Musical training and MINT outcomes

Descriptive statistics for overall MINT performance are presented in Table 2. The mean accuracy scores for each MINT sub-condition were: Baseline (Pitch) = 80.94 ($SD = 11.95$), Rhythm = 63.85 ($SD = 14.56$), Spatial = 84.02 ($SD = 10.49$), Visual = 90.60 ($SD = 10.27$), and Prediction = 90.85 ($SD = 8.80$). The mean accuracy scores for each SNR level were: SNR 0 = 84.77 ($SD = 10.74$), SNR −3 = 83.79 ($SD = 9.32$), and SNR −6 = 77.59 ($SD = 9.93$). Pearson correlation analysis between cumulative practice hours and overall MINT task performance revealed a significant correlation, with a *r*-value of 0.363 ($p < 0.001$; Figure 3B; see Supplementary Table S1 for Spearman's results). Cumulative hours of practice were also correlated with the Baseline (Pitch) ($r = 0.22$, $p = 0.025$), Prediction ($r = 0.26$, $p = 0.010$), Rhythm ($r = 0.28$, $p = 0.007$), and Visual ($r = 0.29$, $p = 0.005$) sub-conditions. In addition, cumulative hours of practice

TABLE 2 Experiment 1.

Variable	Total (N = 78)	Musicians (N = 42)	Non-musicians (N = 20)
AWM (mean ± SD)	78.75 ± 18.52	86.94 ± 13.11	56.65 ± 12.57
Range	45–100	57–100	45–92
MINT (mean ± SD)	82.05 ± 7.60	84.64 ± 5.86	75.13 ± 8.36
Range	57.33–92.00	70.67–92.00	57.33–89.33

Descriptive statistics for AWM and MINT scores.

correlated with all SNR levels: SNR 0 ($r = 0.24$, $p = 0.019$), SNR −3 ($r = 0.31$, $p = 0.003$), and SNR −6 ($r = 0.29$, $p = 0.006$). Mann–Whitney U test shows a significant difference in MINT performance between Musicians and Non-Musicians (U = 696, $p < 0.001$) (Figure 3C).

AWM and MINT performance

Pearson correlation analysis evaluated the relationship between performance on the AWM and MINT tasks. The AWM scores significantly correlated with the overall MINT scores ($r = 0.584$, $p < 0.001$) (Figure 3D). The AWM correlated with all MINT sub-conditions, as listed in Table 3A (see also Supplementary Table S2A). Moreover, AWM was correlated with all the SNR levels, as presented in Table 3B (see also Supplementary Table S2B). Fisher's test performed to compare the differences between the z-transformations of each pair of correlations demonstrated that none of the correlations were significantly larger than the others.

Mediating role of AWM

Regression analyses with bootstrapping were performed to assess each component of the proposed mediation model. First, it was found that cumulative music training hours were positively associated with both MINT performance [$R = 0.36$, $F(1, 76) = 11.56$, $p = 0.001$] and AWM performance [$R = 0.40$, $F(1, 76) = 14.42$, $p < 0.001$]. It was also found that the mediator, AWM ability, was positively related to the MINT test score [$R = 0.58$, $F(1, 76) = 39.42$, $p < 0.001$]. Lastly, multiple regression analysis was conducted to examine the effects of hours of musical training (X_1) and AWM (X_2) on MINT performance (Y). Results indicated that the overall regression model was significant [$R = 0.601$, $F(2, 75) = 21.24$, $p < 0.001$], with VIF = 1.19, MSE = 37.88, and $\eta^2 = 0.362$ (Figure 4). Both predictors contributed to better MINT performance ($\beta_1 = 0.16$, $p = 0.129$; $\beta_2 = 0.52$, $p < 0.001$).

Because the general model, the a-path (music training to AWM), and the b-path (AWM to MINT) were significant, mediation analysis was tested using the bootstrapping method with bias-corrected confidence estimates (refer to Methods and Materials section Data Analysis; Preacher and Hayes, 2004). The 95% confidence interval of the indirect effect was obtained with 5,000 bootstrap samples and confirmed the significant mediating role of AWM in the relationship between music training and MINT task performance (Figure 4). Regression results also indicated that the direct effect of music training on MINT becomes non-significant ($p = 0.11$) when controlling for AWM, thus suggesting full mediation. Moreover, confidence intervals derived from bootstrapping mediation analysis revealed mediation

effects of AWM on the Baseline (Pitch), Prediction, Rhythm, and Visual sub-conditions. Our results also indicate significant mediating effects of AWM on MINT performance at the SNR 0 and SNR −6 levels.

Interim discussion

The main findings from Experiment 1 aligned with our hypothesis, highlighting a clear advantage for musicians in both AWM abilities and music-in-noise perception. The results reveal a positive correlation between the number of practice hours and AWM task performance, and that the Musician group consistently outperformed Non-Musicians in AWM abilities. Additionally, both correlational and group comparison analyses illustrate a significant association between musical experience and enhanced music-in-noise performance. The bootstrapping analysis concerning practice hours, AWM and MINT further supports our mediation hypothesis, suggesting that AWM ability substantially mediates the relationship between musical experience and music-in-noise perception.

It is important to note that the majority of subjects from Experiment 1 were selected to fall into either non-musicians or expert musicians categories. Consequently, the dataset includes fewer subjects with moderate exposure to music and thus may be less reflective of the general population's musical experience distribution. Although Pearson's correlations indicate a notable parametric association between music training and both AWM and music-in-noise abilities, replicating the main effects observed in Experiment 1 based on a more normative and representative dataset would strengthen the statistical robustness and generalizability of the results.

In addition, results from the MINT task in Experiment 1 showed that participants performed optimally around the 80% mark, suggesting that the SNR range tested (0, −3, and − 6) may not fully challenge their music-in-noise capabilities. In light of these findings, we devised a second phase of the study to extend the difficulty of the MINT task with SNR levels of −3, −6, and − 9. By adjusting the noise ratio, we aim to better understand how musicianship affects MINT performance under more demanding conditions and to assess whether the effects observed in Experiment 1 persist with increased task demand. This modification should provide an assessment of the consistency of musical training effects across a wider range of noise interference challenges.

Based on the main correlational results from Experiment 1, we determined the minimum sample size required for Experiment 2 to achieve the desired statistical power. Using an expected correlation coefficient (ρ) of 0.40, a significance level (α) of 0.05, and a power (1 - β) of 0.90, and applying the Fisher Transformation of the correlation coefficient, the minimum sample size required for Experiment 2 is calculated to be 66.

Experiment (2)

Methods and materials

Participants

In the replication phase, we recruited 73 subjects with a distributed range of music experience and expertise (Table 4). Recruitment
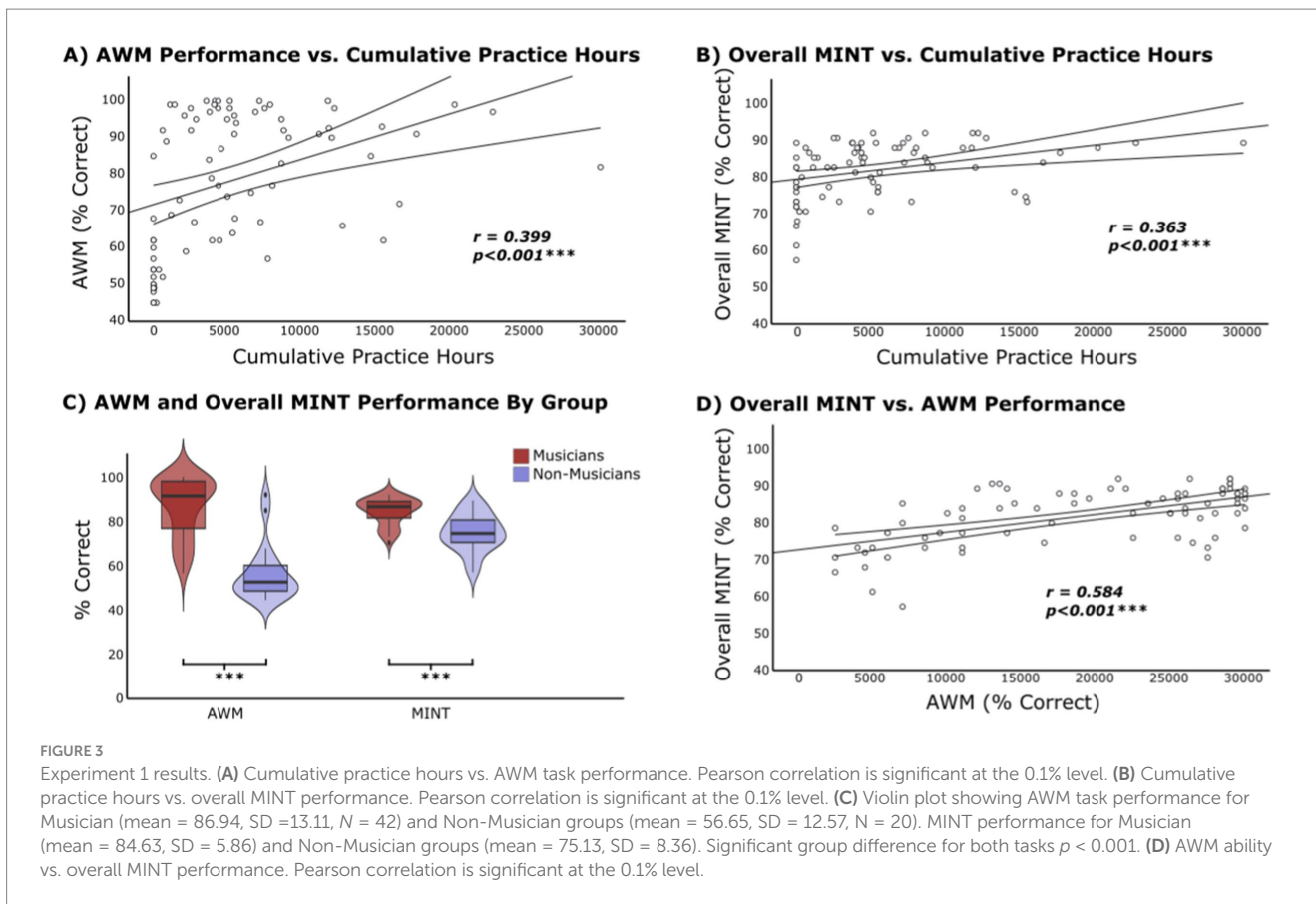
FIGURE 3

Experiment 1 results. **(A)** Cumulative practice hours vs. AWM task performance. Pearson correlation is significant at the 0.1% level. **(B)** Cumulative practice hours vs. overall MINT performance. Pearson correlation is significant at the 0.1% level. **(C)** Violin plot showing AWM task performance for Musician (mean = 86.94, SD =13.11, N = 42) and Non-Musician groups (mean = 56.65, SD = 12.57, N = 20). MINT performance for Musician (mean = 84.63, SD = 5.86) and Non-Musician groups (mean = 75.13, SD = 8.36). Significant group difference for both tasks p < 0.001. **(D)** AWM ability vs. overall MINT performance. Pearson correlation is significant at the 0.1% level.

TABLE 3 Experiment 1.

(A)						
	Pitch total	Prediction total	Rhythm total	Spatial total	Visual total	MINT overall
AWM task performance	0.424***	0.447***	0.373***	0.244*	0.509***	0.584***
Sig. (2-tailed)	<0.001	<0.001	<0.001	0.032	<0.001	<0.001

(B)			
	SNR−6	SNR−3	SNR 0
AWM task performance	0.538***	0.360***	0.432***
Sig. (2-tailed)	<0.001	<0.001	<0.001

(A) AWM task performance vs. MINT conditions. Pearson correlations based on AWM task percent correct and average MINT scores for the corresponding condition. (B) AWM task performance vs. MINT SNR Levels. Pearson correlations based on AWM task percent correct and average MINT scores for each SNR level across conditions. ***denotes statistical significance at the 0.1% level. *denotes statistical significance at the 5% level.

methods and compensation were consistent with those used in Experiment 1. All participants had completed at least 1 year of university-level education. On average, subjects from Experiment 2 have fewer practice hours than those in Experiment 1; $t(143) = -1.84$, $p = 0.034$. Within the 73 subjects, 19 were categorized as Musicians according to the same criteria as above, and 18 were Non-Musicians (Table 4).

All procedures and screening criteria remained consistent with those in Experiment 1 and were approved by either the McGill University Faculty of Medicine Research Ethics Board or Western University Non-Medical Research Ethics Board. Out of the 73 subjects who completed all components of Experiment 2, 3 who could not process basic musical content were excluded from the MINT analysis.

Procedure

Refer to Experiment 1 Materials and methods section Procedure.

Measures and Behavioral Tasks

Refer to Experiment 1 Materials and methods section Measures and Behavioral Tasks.

Results

Musical training and AWM

The mean accuracy score (% correct) for the AWM task in the second sample was 66.33 (SD = 15.85, range: 41–100, N = 70). Results from a one-tailed Pearson correlation test indicated a trend toward significance in the association between musical training and AWM task performance (r = 0.191, p = 0.057). Potential outlier effects were suspected through examination of the data distribution, prompting the use of Spearman's rank-order correlation, which is more robust to extreme values. The Spearman's test revealed a significant monotonic relationship between AWM scores and cumulative hours of practice (ρ = 0.324, p = 0.003). The discrepancy between the rank-order and parametric test results suggests that the data may have been affected by extreme values. Upon comprehensive examination of the total 148 qualified subjects from Experiment 1 and 2 using linear regression (practice hours versus AWM performance), we identified two subjects from Experiment 2 with performance significantly deviating from the model's predictions. Specifically, one subject had a standardized residual of −2.45 and the other −2.40, while the standardized residuals for the remaining 146 subjects ranged between −1.67 and 1.70. Consequently, these two subjects are considered outliers and were excluded from subsequent analysis.

By removing the two outliers, the adjusted mean AWM accuracy score in the second sample is presented in Table 5. Independent samples t-test indicates a significantly lower AWM performance for the subjects in Experiment 2 compared to Experiment 1; $t(144)$ = −4.17, p < 0.001. A significant relationship between AWM

FIGURE 4
Mediation analysis results. Enhanced AWM was the significant mediator of the correlation between music training (cumulative practice hours) and MINT performance. Pearson correlation is significant at the 0.1% level.

score and cumulative hours of practice is demonstrated with Pearson's test (r = 0.370, p < 0.001) (Figure 5A; see Supplementary Table S3 for Spearman's results). In addition, the Mann–Whitney U test also indicates a group difference in AWM between Musicians and Non-Musicians (U = 302, p < 0.001; Table 5; Figure 5C).

Musical training and MINT outcomes

Descriptive statistics for overall MINT performance are presented in Table 5. Independent samples t-test between Experiment 1 and 2 indicates a significantly lower MINT score for the subject in Experiment 2; $t(144)$ = −5.36, p < 0.001. The mean accuracy scores for each MINT condition were: Baseline (Pitch) = 72.45 (SD = 16.25), Rhythm = 59.51 (SD = 12.95), Spatial = 71.76 (SD = 13.97), Visual = 83.43 (SD = 15.77), and Prediction = 81.47 (SD = 15.19). The mean accuracy scores for each SNR level were: SNR -3 = 79.29 (SD = 14.38), SNR −6 = 74.00 (SD = 12.51), and SNR −9 = 67.88 (SD = 13.04).

A correlation between cumulative practice hours and MINT task performance was tested with r = 0.293 (p = 0.008; Figure 5B; see Supplementary Table S3 for Spearman's results). Cumulative hours of practice were also correlated with the Baseline (Pitch) (r = 0.21, p = 0.040), Prediction (r = 0.29, p = 0.009), and Visual (r = 0.29, p < 0.009) sub-conditions. In addition, cumulative hours of practice correlated with SNR −3 (r = 0.34, p = 0.003) and SNR −9 (r = 0.21, p = 0.043). Mann–Whitney U test shows a significant difference in total MINT performance between Musicians and Non-Musicians (U = 281, p < 0.001; Table 5; Figure 5C).

AWM and MINT performance

Pearson correlation analysis evaluated the correlation between performance on the AWM and MINT tasks. AWM score significantly correlated with the overall MINT score (r = 0.573, p < 0.001; Figure 5D). The correlations between AWM also significantly correlated with all MINT sub-conditions and SNR levels, as presented in Table 6 (see also Supplementary Table S4). Fisher's test performed to compare the differences between the z-transformations of each pair of correlations demonstrated that none of the correlations are significantly larger than the others.

Mediating role of AWM

Regression analyses were conducted to assess each component of the mediation model proposed in Experiment 1. Linear regression with bootstrapping revealed a positive association between cumulative music training hours and both AWM performance [R = 0.370, F(1,
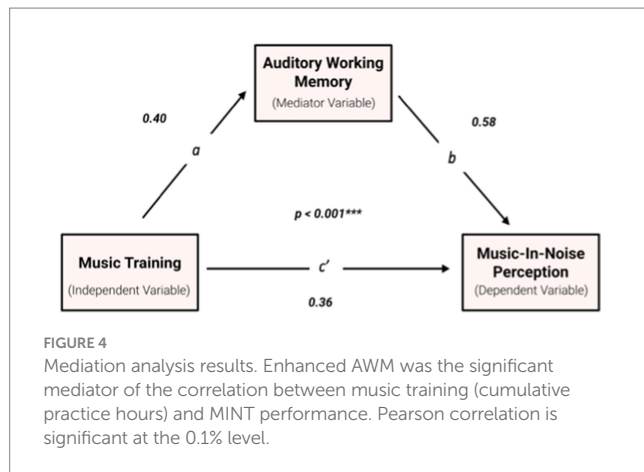
TABLE 4 Experiment 2.

Variable	Total (N = 73)	Musicians (N = 19)	Non-Musicians (N = 18)
Age (mean ± SD)	27.0 ± 6.6	28.3 ± 6.5	26.4 ± 6.3
Age range	18–49	20–41	20–45
Sex	37 males, 35 females, 1 non-binary	10 males, 8 females, 1 non-binary	10 males, 8 females
Cumulative practice hours (mean ± SD)	4,600 ± 6,600	12,000 ± 8,600	200 ± 600
Cumulative practice range	0–34,000	4,700–32,000	0–1900
Age of onset (mean ± SD)	N/A	6.0 ± 2.5	N/A

Descriptive statistics of demographic and musical training variables.

TABLE 5 Experiment 2.

Variable	Total (N = 68)	Musicians (N = 19)	Non-musicians (N = 18)
AWM (mean ± SD)	66.78 ± 15.83	76.68 ± 16.60	55.44 ± 6.82
Range	41–100	55–100	41–69
MINT (mean ± SD)	73.73 ± 11.05	78.74 ± 9.48	66.15 ± 11.54
Range	46.67–90.67	53.33–89.33	46.67–90.67

Descriptive statistics for AWM and MINT scores.

66) = 10.48, $p = 0.004$] and MINT performance [$R = 0.293$, $F(1, 66) = 6.18$, $p = 0.015$]. AWM ability as the proposed mediator was also positively related to MINT test scores [$R = 0.573$, $F(1, 66) = 32.24$, $p < 0.001$]. Subsequent multiple regression analysis was performed to assess the effects of musical training hours (X_1) and AWM (X_2) on MINT performance (Y). The overall regression model was significant [$R = 0.579$, $F(2, 65) = 16.42$, $p < 0.001$], with VIF = 1.16, MSE = 83.62, $\eta^2 = 0.336$, and predictors contributing to improved MINT performance ($\beta_1 = 0.093$, $p = 0.394$; $\beta_2 = 0.538$, $p < 0.001$).

Given that the multiple regression model and the paths were significant and consistent with Experiment 1, mediation analysis was conducted using the same bootstrapping method (refer to Methods and Materials section Data Analysis). A 95% confidence interval for the indirect effect was derived from bootstrap samples and demonstrated a significant mediating role of AWM in the relationship between music training and MINT task performance. Results also show that the direct effect of music training on MINT became non-significant (p = 0.394) when controlling for AWM. Additional bootstrapping analysis also revealed a mediating effect of AWM on the Baseline (Pitch), Prediction, Spatial and Visual sub-conditions, and across all SNR levels.

Discussion

Effects of musical training on AWM and music-in-noise perception

The findings from both Experiments 1 and 2 provide compelling evidence that supports our hypothesis of a musician's advantage in both AWM abilities and music-in-noise perception. Importantly, the musician advantage was consistently observed across two distinct samples, which differed in overall musical experience, proportion of musicianship, and average performance on both tasks.

A meta-analysis by Talamini et al. (2017) demonstrated that musicians outperform non-musicians across various memory domains, including long-term, short-term, and working memory, with a particularly pronounced advantage for tonal stimuli. To investigate this tonal aspect of AWM in which musicians excel, the task in this study required participants to detect a local pitch change between two tonal patterns that differed in temporal order. This AWM task not only captured auditory retention capabilities but also assessed the ability to mentally manipulate the stimuli (i.e., serial order processing; Albouy et al., 2017; Foster et al., 2013), along with related cognitive skills such as decision-making, attention, processing speed, etc. Correlational analyses between cumulative practice hours and AWM task performance from Experiment 1 indicated a positive

association between music experience and AWM abilities (Figure 3A), a finding that was replicated in Experiment 2 (Figure 5A). Moreover, the group comparison underscores this advantage, as musicians from both studies consistently outperformed their non-musician counterparts on the standardized measures of AWM (Figures 3C, 5C). These results are supported by existing literature, which consistently demonstrates behavioral, electrophysiological (event-related potential), and neuro-oscillatory evidence for the superiority of musicians in AWM abilities (Albouy et al., 2017; Foster et al., 2013; George and Coch, 2011).

Analyses of the overall MINT task performance in Experiment 1 in relation to cumulative practice hours suggests a clear association between musical experience and improved music-in-noise perception (Figure 3B). Although subjects in Experiment 2 showed an overall poorer performance on the MINT—potentially due to differences in musicianship and/or increased task SNR difficulty—the correlation between musical experience and MINT performance remained consistent and significant (Figure 5B). In other words, the relationship between musical experience and music-in-noise perception is stable across different signal-to-noise levels tested. Furthermore, significant musician advantage on music-in-noise perception was also observed in both studies when comparing the musician and non-musician group differences in the overall MINT performance (Figures 3C, 5C). These results are in line with the findings from Coffey et al. (2019) original MINT study and the subsequent MINT results by Hsieh et al. (2022), further validating the MINT's reliability and supporting the cognitive benefits of musical expertise amid varying levels of noise interference.

AWM ability and music-in-noise perception

Musicians' music-in-noise benefits may arise from improvements in both auditory perception and cognitive processing. On the perceptual side, musicians demonstrate an increased sensitivity to fundamental acoustic features critical for music perception, such as pitch discrimination and temporal fine structure (Micheyl et al., 2006; Mishra et al., 2015). Cognitively, studies have shown a connection between musicianship and enhancements in cognitive faculties including working memory and attention (Bidelman and Yoo, 2020; Yoo and Bidelman, 2019), which may be linked to stream segregation improvements.

Evidence suggesting that AWM plays a crucial role in music-in-noise perception stems from the strong correlation between performance on the AWM task and the overall MINT score, observed in both Experiment 1 (Figure 3D) and Experiment 2 (Figure 5D). This finding replicates the original MINT study (Coffey et al., 2019), and is consistent with the majority of the SIN literature which suggests that working memory for phonological or tonal information is linked to improved speech segregation abilities (Bidelman and Yoo, 2020; Escobar et al., 2020; Lad et al., 2020; Mattys et al., 2012; Yoo and Bidelman, 2019).

The mediation analysis conducted in Experiment 1 supports our hypothesis that AWM ability significantly mediates the relationship between musical experience and music-in-noise perception (Figure 4). This mediation model was successfully replicated in Experiment 2, which included a more heterogeneous musician population. The comparable results from Experiment 2 reinforce the reliability and generalizability of our initial findings. Overall, our results suggest that musicians' enhanced AWM skills are a crucial driving force behind
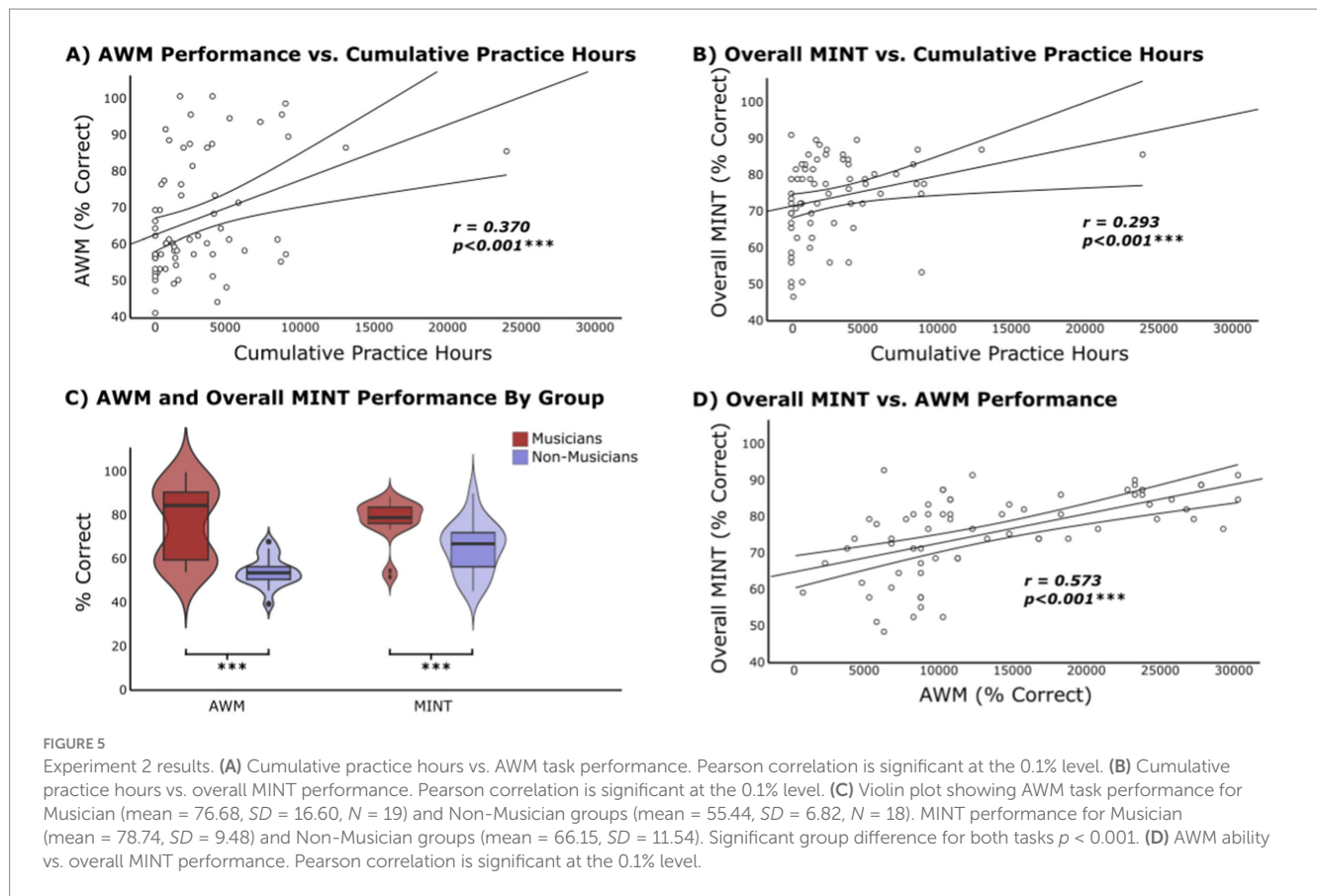
FIGURE 5
Experiment 2 results. **(A)** Cumulative practice hours vs. AWM task performance. Pearson correlation is significant at the 0.1% level. **(B)** Cumulative practice hours vs. overall MINT performance. Pearson correlation is significant at the 0.1% level. **(C)** Violin plot showing AWM task performance for Musician (mean = 76.68, *SD* = 16.60, *N* = 19) and Non-Musician groups (mean = 55.44, *SD* = 6.82, *N* = 18). MINT performance for Musician (mean = 78.74, *SD* = 9.48) and Non-Musician groups (mean = 66.15, *SD* = 11.54). Significant group difference for both tasks *p* < 0.001. **(D)** AWM ability vs. overall MINT performance. Pearson correlation is significant at the 0.1% level.

TABLE 6 Experiment 2.

	(A)					
	Pitch total	Prediction total	Rhythm total	Spatial total	Visual total	MINT overall
AWM task performance	0.451***	0.424***	0.292***	0.438***	0.507***	0.573***
Sig. (1-tailed)	<0.001	<0.001	0.008	<0.001	<0.001	<0.001

	(B)		
	SNR−9	SNR−6	SNR−3
AWM task performance	0.435***	0.481***	0.508***
Sig. (1-tailed)	<0.001	<0.001	<0.001

(A) AWM task performance vs. MINT conditions. Pearson correlations based on AWM task percent correct and average MINT scores for the corresponding condition. (B) AWM Task Performance vs. MINT SNR Levels. Pearson correlations based on AWM task percent correct and average MINT scores for each SNR level across conditions. ***denotes statistical significance at the 0.1% level.

their enhanced MINT performance, and that musical training is associated with improvements in the performance of auditory stream segregation tasks through the enhancement of AWM capabilities. This mediating effect of AWM in music-in-noise performance parallels the mediation model proposed for AWM's role in SIN performance (Kraus et al., 2012; Parbery-Clark et al., 2009b). Parbery-Clark et al. (2009b) demonstrated that musicians possess superior AWM skills, which those authors identify as a significant factor behind the group's improved SIN performance. In addition, Bidelman and Yoo (2020) found that the relationship between musicianship and performance on a complex SIN task did not remain significant after controlling for

working memory, which is associated with the listener's year of musical training. This finding supports the concept that auditory stream segregation superiority is driven heavily by the enhanced working memory capacity, likely developed through musical training, although aptitude may also play role.

Although evidence supports the importance of AWM in overall stream segregation, the precise mechanisms underlying its contribution remain unclear. The predominant literature on SIN has focused on the role of AWM in facilitating the understanding of linguistic context (Kraus et al., 2012). For example, the Ease of Language Understanding (ELU) model by Rönnberg et al. (2013)

posits that working memory enables the listener to hold a schematic representation of speech while processing contextual information, using linguistic knowledge to compensate for missing information in adverse listening environments. In addition, the ELU model states that individuals with enhanced working memory capacity can apply more mental resources to resolve the phonological and semantic aspects of a listening task (Rönnberg et al., 2013). It follows that the advantage offered by AWM in aiding SIN processing may depend largely on the redundancy of linguistic contextual cues (e.g., phonological, lexical, syntactic, and semantic information) of the speech signal tested (Gordon-Salant and Cole, 2016). However, given the consistent relationship between AWM and MINT performance—which is not influenced by linguistic factors—our study provides evidence that the benefits of AWM in stream segregation extend beyond the speech domain, pointing to more fundamental mechanisms that are more generally involved in stream segregation processing.

AWM's association with perceptual and cognitive components in stream segregation

One advantage of the MINT over standard SIN tests is its ability to assess specific cues and auditory sub-skills related to stream segregation (Coffey et al., 2019), offering insights into how AWM may interact with the perceptual and cognitive elements involved in this process. The original MINT study indicated that AWM has the most significant contribution to the Prediction condition, and the relationship between musical training and the Prediction task diminished in significance when AWM performance was factored in as a covariate in the analysis (Coffey et al., 2019). Prior research also supports the role of AWM in musical predictive processing, highlighting its importance in top-down schematic expectations—the concept that knowing the pattern to be segregated *a priori* facilitates subsequent detection (Bey and McAdams, 2002).

However, contrary to earlier findings, here we did not observe a more important contribution of AWM to the MINT Prediction condition compared to other conditions. Instead, there were significant and consistent correlations between AWM and all MINT sub-tasks in both Experiments (Table 3A; Table 6A). This finding suggests that AWM's contribution is only one among many factors that modulate stream segregation situations.

One possible explanation for the contribution of AWM to general stream segregation is that enhanced AWM allows a more precise representation of acoustic signals in the mental workspace (Kraus et al., 2012). Research suggests that working memory is linked to improved performance on a rhythm synchronization task, where participants are required to reproduce the temporal structure of the presented rhythms (Bailey and Penhune, 2010). It is also indicated that individuals who can effectively retain auditory source properties, such as frequency and temporal fluctuations over time have a perceptual advantage in SIN tasks (Lad et al., 2020; Lad et al., 2024). It is therefore plausible that the ability to maintain acoustic information accurately aids the sequential segregation processes essential for stream intelligibility (Bregman, 1990).

Another perspective involves attention. Dalton et al. (2009) manipulated the working memory load during a distractor interference task, demonstrating a causal role for the availability of working memory in auditory selective attention. In addition, it is suggested that segregating auditory streams from background noise draws upon attentional resources (Heinrich et al., 2008), and accomplishing such tasks necessitates the allocation of one's limited cognitive resources to balance the competing demands of attention, processing, and storage (Wingfield and Tun, 2007). It is therefore plausible that enhanced AWM proficiency promotes the maintenance and encoding of auditory signals, which in turn allows for more efficient use of attention resources to extract and recall the target stream.

In addition, the advantages of AWM can be understood through the temporal aspects of information processing: temporal integration and serial order processing. On the one hand, it is proposed that working memory aids the linkage between recent past and imminent future events, thus serving both a retrospective role in information retention and a prospective role in anticipation (Fuster and Bressler, 2012). Specifically, prior literature proposes that working memory is important for minimizing distractor interference through the active maintenance of current stimulus-processing priorities (Dalton et al., 2009; Lavie, 2005). In stream segregation, AWM may therefore enable individuals to hold fragments of auditory information while processing, integrating, and anticipating degraded target signals.

On the other hand, the AWM task used in this study, which requires temporal reversal, captures item-based retention and serial order processing, which have been shown to be distinct processes. Serial ordering, in particular, is thought to be a domain-general process based on positional codes, as observed in verbal and musical working memory studies (Hurlstone et al., 2014; Majerus, 2019; Gorin, 2022). Since melodic retention and prediction did not appear to play a special role in stream segregation, serial ordering may serve as an alternative key factor, contributing to tracking the sequence of items over time and thereby enhancing the ability to organize auditory streams. It will be of interest in future research to study the contribution of AWM when measured with tasks that do not require serial order processing, such as for example musical transposition.

The auditory dorsal stream and its implications for musician enhancement

The dorsal stream of auditory processing, which involves the parietal lobe, dorsal premotor cortex, and dorsolateral frontal regions, is central to higher-order cognitive auditory functions. It supports the manipulation of sound patterns in working memory, auditory-motor integration, abstract temporal representations, and predictive coding (Rauschecker and Scott, 2009; Zatorre, 2024). Neuroimaging studies highlight the dorsal stream's key role in AWM, with activations in parietal regions associated with various kinds of mental transformation and manipulation processes (Foster et al., 2013; Zatorre et al., 2010). Moreover, Albouy et al. (2017) have observed that sustained evoked activity in the bilateral dorsal streams, particularly through long-range theta phase locking and increased local theta power in the IPS, is associated with successful AWM manipulation. Furthermore, when theta power is boosted in the dorsal stream via rhythmic brain stimulation (Albouy et al., 2017) or via flickering visual rotating stimuli (Albouy et al., 2022), AWM performance is also enhanced.

While perceiving auditory signals in background noise heavily engages primary and non-primary auditory regions (Holmes et al., 2021; Kell and McDermott, 2019; Puschmann et al., 2019), research indicates that motor and somatosensory areas are also more actively recruited under challenging listening conditions (for review, see

Skipper et al., 2017). This suggests a compensatory mechanism of dorsal steam activity for reduced processing specificity in the auditory system (Du et al., 2014). Importantly, a study comparing musicians and non-musicians found that the benefits of musical training on SIN perception in difficult listening contexts were related to activity in the motor cortices of the auditory dorsal streams (Du and Zatorre, 2017).

Further research has shown that music training enhances functional connectivity within the dorsal auditory stream (Jünemann et al., 2023). Musicians also exhibit greater structural connectivity in the white matter tracts of the dorsal stream (i.e., arcuate fasciculus and superior longitudinal fasciculus; Halwani et al., 2011; Oechslin et al., 2010). Differences in the microstructural plasticity of dorsal white matter are suggested to underlie musicians' improved SIN perception (Li et al., 2021). Considering the role of the auditory dorsal stream in AWM and SIN perception, we thus infer that the musician enhancements in these abilities may be rooted in this stream, although the exact mechanisms warrant further exploration.

Implications for age-related hearing loss

Auditory functioning is one of the most prevalently affected sensory modalities in the elderly population (Yamasoba et al., 2013; Davis et al., 2016). In addition, older adults show deficits in speech recognition in noisy environments and AWM (Dubno et al., 1984; Humes and Floyd, 2005). Previous studies demonstrated that older musicians exhibit enhanced performance in AWM and SIN perception compared to their non-musician counterparts, suggesting that musical experience may mitigate age-related hearing challenges (Zhang et al., 2021; Zendel et al., 2019).

Recent longitudinal studies assigning older adults to musical activities (piano/choir) have also demonstrated behavioral, neurophysiological, and neuro-oscillatory evidence of improvements in SIN perception (Worschech et al., 2021; Hennessy et al., 2021; Dubinsky et al., 2019; Gray et al., 2022). Shedding light onto the mediating role of AWM in stream segregation, we propose that future music programs designed to address hearing challenges in older adults should focus on enhancing AWM to achieve optimal intervention outcomes.

Limitations and future directions

Limitations of the current study include reliance on self-report music history questionnaire responses and the challenge of precisely controlling for the nuanced variations of individual musical experiences and expertise (e.g., learning styles, extent of practice). Moreover, the correlational design of the study does not address issues regarding self-selection and the direction of causality, particularly considering evidence suggesting that auditory and musical expertise arises from a combination of genetic predispositions and experience-driven plasticity (Schellenberg, 2015; Zatorre, 2013). The inherent predispositions for AWM or stream segregation abilities could potentially influence one's path toward musical engagement, an aspect that warrants further investigation.

Longitudinal studies with school-aged children (as well as the elderly, as described in the preceding section) provide evidence that music instruction is in fact causally associated with moderate benefits in SIN and AWM abilities (Slater et al., 2015; Nie et al., 2022), but of course, this does not mean that predisposing factors do not exist. Moreover, research has demonstrated a relationship between music and language performance in elementary school children (Zuk et al., 2013), which is primarily driven by temporal processing (Andrade et al., 2024). These findings suggest some shared contributions, yet the extent of transfer effects from music to speech and phonological processing remains unclear. Therefore, future research directions entail conducting longitudinal studies to examine the development of both speech-in-noise and music-in-noise perception, further unraveling the relationship between musical training, AWM, and overall auditory stream segregation. Such endeavors will also help elucidate experience-dependent plasticity in the auditory domain and contribute to a deeper understanding of the development of higher-level auditory cognitive mechanisms.

Conclusion

This study explores the influence of musical training on two auditory cognitive processes: AWM and stream segregation. As hypothesized, our findings provide support for a musician advantage in AWM abilities and music-in-noise perception. We show using replication across two samples that musicians' enhanced AWM skill is one of the driving forces behind their better music-in-noise performance, suggesting that musicianship fosters improvements in stream segregation through the enhancement of AWM capabilities. In addition, the study's two-phase design strengthens the generalizability of the results across various populations and conditions. Together, these findings shed light on the relationship between musical training, AWM, and stream segregation, underscoring the potential for music-based interventions to enhance auditory processing abilities.

Data availability statement

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

Ethics statement

The studies involving humans were approved by McGill University Faculty of Medicine Research Ethics Board Western University Nomedical Research Ethics Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

ML: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. IA-B: Investigation, Writing – review & editing. MF: Investigation, Writing – review & editing. M-EL: Investigation, Writing – review & editing. JS: Investigation, Writing – review & editing. EI: Investigation, Writing – review & editing.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1538511/full#supplementary-material

References

Albouy, P., Martinez-Moreno, Z. E., Hoyer, R. S., Zatorre, R. J., and Baillet, S. (2022). Supramodality of neural entrainment: rhythmic visual stimulation causally enhances auditory working memory performance. *Sci. Adv.* 8:eabj9782. doi: 10.1126/sciadv.abj9782

Albouy, P., Weiss, A., Baillet, S., and Zatorre, R. J. (2017). Selective entrainment of Theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron* 94, 193–206.e5. doi: 10.1016/j.neuron.2017.03.015

Anderson, S., White-Schwoch, T., Parbery-Clark, A., and Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear. Res.* 300, 18–32. doi: 10.1016/j.heares.2013.03.006

Andrade, P. E., Müllensiefen, D., Andrade, O. V. C. A., Dunstan, J., Zuk, J., and Gaab, N. (2024). Sequence processing in music predicts Reading skills in young readers: a longitudinal study. *J. Learn. Disabil.* 57, 43–60. doi: 10.1177/00222194231157722

Bailey, J. A., and Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Exp. Brain Res.* 204, 91–101. doi: 10.1007/s00221-010-2299-y

Baddeley, A. (1992). Working memory. *Science*, 255, 556–559. doi: 10.1126/science.1736359

Bey, C., and McAdams, S. (2002). Schema-based processing in auditory scene analysis. *Percept. Psychophys.* 64, 844–854. doi: 10.3758/BF03194750

Bidelman, G. M., and Yoo, J. (2020). Musicians show improved speech segregation in competitive, multi-talker cocktail party scenarios. *Front. Psychol.* 11:1927. doi: 10.3389/fpsyg.2020.01927

Bregman, A. S. (1990). Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: The MIT Press.

Bregman, A. S., and Pinker, S. (1978). Auditory streaming and the building of timbre. *Can. J. Psychol.* 32, 19–31. doi: 10.1037/h0081664

Brown, R. M., Zatorre, R. J., and Penhune, V. B. (2015). Expert music performance: cognitive, neural, and developmental bases. *Prog. Brain Res.* 217, 57–86. doi: 10.1016/bs.pbr.2014.11.021

Coffey, E. B. J., Arseneau-Bruneau, I., Zhang, X., and Zatorre, R. J. (2019). The music-in-noise task (MINT): a tool for dissecting complex auditory perception. *Front. Neurosci.* 13:199. doi: 10.3389/fnins.2019.00199

Coffey, E. B. J., Herholz, S. C., Scala, S., and Zatorre, R. J. (2011). "Montreal music history questionnaire: a tool for the assessment of music-related experience in music cognition research," in Proceedings of the Neurosciences and Music IV: Learning and Memory, Conference Edinburgh.

Coffey, E. B. J., Mogilever, N. B., and Zatorre, R. J. (2017). Speech-in-noise perception in musicians: a review. *Hear. Res.* 352, 49–69. doi: 10.1016/j.heares.2017.02.006

Dalton, P., Santangelo, V., and Spence, C. (2009). The role of working memory in auditory selective attention. *Q. J. Exp. Psychol.* 62, 2126–2132. doi: 10.1080/17470210903023646

Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147. doi: 10.1016/j.heares.2007.01.014

Davis, A., McMahon, C. M., Pichora-Fuller, K. M., Russ, S., Lin, F., Olusanya, B. O., et al. (2016). Aging and hearing health: the life-course approach. *The Gerontologist* 56, S256–S267. doi: 10.1093/geront/gnw033

Deroche, M. L. D., Limb, C. J., Chatterjee, M., and Gracco, V. L. (2017). Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. *J. Acoust. Soc. Am.* 142, 1739–1755. doi: 10.1121/1.5005496

Du, Y., Buchsbaum, B. R., Grady, C. L., and Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. USA* 111, 7126–7131. doi: 10.1073/pnas.1318738111

Du, Y., and Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proc. Natl. Acad. Sci. USA* 114, 13579–13584. doi: 10.1073/pnas.1712223114

Dubinsky, E., Wood, E. A., Nespoli, G., and 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. doi: 10.3389/fnins.2019.01153

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76, 87–96. doi: 10.1121/1.391011

Escobar, J., Mussoi, B. S., and Silberer, A. B. (2020). The effect of musical training and working memory in adverse listening situations. *Ear Hear.* 41, 278–288. doi: 10.1097/AUD.0000000000000754

Foster, N. E. V., Halpern, A. R., and Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *Neuroimage* 75, 27–35. doi: 10.1016/j.neuroimage.2013.02.044

Fuster, J. M., and Bressler, S. L. (2012). Cognit activation: a mechanism enabling temporal integration in working memory. *Trends Cogn. Sci.* 16, 207–218. doi: 10.1016/j.tics.2012.03.005

George, E. M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49, 1083–1094. doi: 10.1016/j.neuropsychologia.2011.02.001

Gordon-Salant, S., and 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. doi: 10.1097/AUD.0000000000000316

Gorin, S. (2022). Temporal grouping effects in verbal and musical short-term memory: is serial order representation domain-general? *Q. J. Exp. Psycholo.* 75, 1603–1627. doi: 10.1177/17470218211057466

Gray, R., Sarampalis, A., Başkent, D., and Harding, E. E. (2022). Working-memory, alpha-theta oscillations and musical training in older age: research perspectives for speech-on-speech perception. *Front. Aging Neurosci.* 14:806439. doi: 10.3389/fnagi.2022.806439

Halwani, G. F., Loui, P., Rüber, T., and Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Front. Psychol.* 2:156. doi: 10.3389/fpsyg.2011.00156

Heinrich, A., Schneider, B. A., and Craik, F. I. (2008). Investigating the influence of continuous babble on auditory short-term memory performance. *Q. J. Exp. Psychol.* 61, 735–751. doi: 10.1080/17470210701402372

Hennessy, S., Wood, A., Wilcox, R., and Habibi, A. (2021). Neurophysiological improvements in speech-in-noise task after short-term choir training in older adults. *Aging* 13, 9468–9495. doi: 10.18632/aging.202931

Herholz, S. C., and Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron* 76, 486–502. doi: 10.1016/j.neuron.2012.10.011

Holmes, E., Zeidman, P., Friston, K. J., and Griffiths, T. D. (2021). Difficulties with speech-in-noise perception related to fundamental grouping processes in auditory cortex. *Cereb. Cortex* 31, 1582–1596. doi: 10.1093/cercor/bhaa311

Hsieh, I. H., Tseng, H. C., and Liu, J. W. (2022). Domain-specific hearing-in-noise performance is associated with absolute pitch proficiency. *Sci. Rep.* 12:16344. doi: 10.1038/s41598-022-20869-2

Humes, L. E., and Floyd, S. S. (2005). Measures of working memory, sequence learning, and speech recognition in the elderly. *J. Speech Langu. Hear. Rese.* 48, 224–235. doi: 10.1044/1092-4388(2005/016)

Hurlstone, M. J., Hitch, G. J., and Baddeley, A. D. (2014). Memory for serial order across domains: an overview of the literature and directions for future research. *Psychol. Bull.* 140, 339–373. doi: 10.1037/a0034221

Jünemann, K., Engels, A., Marie, D., Worschech, F., Scholz, D. S., Grouiller, F., et al. (2023). Increased functional connectivity in the right dorsal auditory stream after a full year of piano training in healthy older adults. *Sci. Rep.* 13:19993. doi: 10.1038/s41598-023-46513-1

Kell, A. J. E., and McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nat. Commun.* 10:3958. doi: 10.1038/s41467-019-11710-y

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116, 2395–2405. doi: 10.1121/1.1784440

Kraus, N., Strait, D. L., and 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. doi: 10.1111/j.1749-6632.2012.06463.x

Lad, M., Holmes, E., Chu, A., and Griffiths, T. D. (2020). Speech-in-noise detection is related to auditory working memory precision for frequency. *Sci. Rep.* 10:13997. doi: 10.1038/s41598-020-70952-9

Lad, M., Taylor, J. P., and Griffiths, T. D. (2024). The contribution of short-term memory for sound features to speech-in-noise perception and cognition. *Hear. Res.* 451:109081. doi: 10.1016/j.heares.2024.109081

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends Cogn. Sci.* 9, 75–82. doi: 10.1016/j.tics.2004.12.004

Lewis, J. H., Castellanos, I., and Moberly, A. C. (2021). The impact of neurocognitive skills on recognition of spectrally degraded sentences. *J. Am. Acad. Audiol.* 32, 528–536. doi: 10.1055/s-0041-1732438

Li, X., Zatorre, R. J., and Du, Y. (2021). The microstructural plasticity of the arcuate fasciculus undergirds improved speech in noise perception in musicians. *Cereb. Cortex* 31, 3975–3985. doi: 10.1093/cercor/bhab063

Majerus, S. (2019). Verbal working memory and the phonological buffer: the question of serial order. *Cortex* 112, 122–133. doi: 10.1016/j.cortex.2018.04.016

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47. doi: 10.1016/j.heares.2006.05.004

Mishra, S. K., Panda, M. R., and Raj, S. (2015). Influence of musical training on sensitivity to temporal fine structure. *Int. J. Audiol.* 54, 220–226. doi: 10.3109/14992027.2014.969411

Nie, P., Wang, C., Rong, G., Du, B., Lu, J., Li, S., et al. (2022). Effects of music training on the auditory working memory of Chinese-speaking school-aged children: a longitudinal intervention study. *Front. Psychol.* 12:770425. doi: 10.3389/fpsyg.2021.770425

Nilsson, M., Soli, S. D., and 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. doi: 10.1121/1.408469

Noorden, L.V. (1975). Temporal coherence in the perception of tone sequences. [PhD Thesis, Institute for Perception Research, Eindhoven]. Technische Hogeschool Eindhoven.

Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., and Jäncke, L. (2010). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* 3:76. doi: 10.3389/neuro.09.076.2009

Parbery-Clark, A., Skoe, E., and Kraus, N. (2009a). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107. doi: 10.1523/JNEUROSCI.3256-09.2009

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009b). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661. doi: 10.1097/AUD.0b013e3181b412e9

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/s0896-6273(01)00580-3

Preacher, K. J., and Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav. Res. Methods Instrum. Comput.* 36, 717–731. doi: 10.3758/BF03206553

Preacher, K. J., and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40, 879–891. doi: 10.3758/BRM.40.3.879

Puschmann, S., Baillet, S., and 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. doi: 10.1093/cercor/bhy193

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The ease of language understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. *Annals of the New York Academy of Sciences*, 1337, 170–177. doi: 10.1111/nyas.12627

Shamma, S. A., and Micheyl, C. (2010). Behind the scenes of auditory perception. *Curr. Opin. Neurobiol.* 20, 361–366. doi: 10.1016/j.conb.2010.03.009

Skipper, J. I., Devlin, J. T., and Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. *Brain Lang.* 164, 77–105. doi: 10.1016/j.bandl.2016.10.004

Slater, J., and 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. doi: 10.1007/s10339-015-0740-7

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., and Kraus, N. (2015). Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behav. Brain Res.* 291, 244–252. doi: 10.1016/j.bbr.2015.05.026

Stenbäck, V., Marsja, E., Hällgren, M., Lyxell, B., and Larsby, B. (2022). Informational masking and listening effort in speech recognition in noise: the role of working memory capacity and inhibitory control in older adults with and without hearing impairment. *J Speech Lang Hear Res* 65, 4417–4428. doi: 10.1044/2022_JSLHR-21-00674

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G. Jr., and Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Sci. Rep.* 5:11628. doi: 10.1038/srep11628

Talamini, F., Altoè, G., Carretti, B., and Grassi, M. (2017). Musicians have better memory than nonmusicians: a meta-analysis. *PLoS One* 12:e0186773. doi: 10.1371/journal.pone.0186773

Thompson, S. K., Carlyon, R. P., and Cusack, R. (2011). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1253–1262. doi: 10.1037/a0021925

Wingfield, A., and Tun, P. A. (2007). Cognitive supports and cognitive constraints on comprehension of spoken language. *J. Am. Acad. Audiol.* 18, 548–558. doi: 10.3766/jaaa.18.7.3

Worschech, F., Marie, D., Jünemann, K., Sinke, C., Krüger, T. H. C., Großbach, M., et al. (2021). Improved speech in noise perception in the elderly after 6 months of musical instruction. *Front. Neurosci.* 15:696240. doi: 10.3389/fnins.2021.696240

Yamasoba, T., Lin, F. R., Someya, S., Kashio, A., Sakamoto, T., and Kondo, K. (2013). Current concepts in age-related hearing loss: epidemiology and mechanistic pathways. *Hear. Res.* 303, 30–38. doi: 10.1016/j.heares.2013.01.021

Yoo, J., and Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear. Res.* 377, 189–195. doi: 10.1016/j.heares.2019.03.021

Yurgil, K. A., Velasquez, M. A., Winston, J. L., Reichman, N. B., and Colombo, P. J. (2020). Music training, working memory, and neural oscillations: a review. *Front. Psychol.* 11:266. doi: 10.3389/fpsyg.2020.00266

Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science*, 342:585–589. doi: 10.1126/science.1238414

Zatorre, R. (2024). From perception to pleasure: the neuroscience of music and why we love it. *online* Edn. New York: Oxford Academic.

Zatorre, R. J., Halpern, A. R., and Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *J. Cogn. Neurosci.* 22, 775–789. doi: 10.1162/jocn.2009.21239

Zendel, B. R., West, G. L., Belleville, S., and Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiol. Aging* 81, 102–115. doi: 10.1016/j.neurobiolaging.2019.05.015

Zhang, L., Fu, X., Luo, D., Xing, L., and Du, Y. (2021). Musical experience offsets age-related decline in understanding speech-in-noise: type of training does not matter, working memory is the key. *Ear Hear.* 42, 258–270. doi: 10.1097/AUD.0000000000000921

Zuk, J., Andrade, P. E., Andrade, O. V., Gardiner, M., and Gaab, N. (2013). Musical, language, and reading abilities in early Portuguese readers. *Front. Psychol.* 4:288. doi: 10.3389/fpsyg.2013.00288