Title: Musical Training Increases Robustness of Sequence Learning and Increases Anticipatory Responses During Sequence Learning

Running title: Musical training speeds up motor planning and increases robustness of sequence learning

Li-Ann Leow[1,2,3] Jarrad Lum[3], Sara Johnson[2], Emily Corti[2], & Welber Marinovic[2]

[1]School of Arts and Humanities, Edith Cowan University, Joondalup, Western Australia, Australia

[2]School of Population Health, Curtin University, Bentley, Western Australia, Australia
[3]School of Psychology, Deakin University, Burwood, Victoria, Australia

Email: l.leow@ecu.edu.au or w.marinovic@curtin.edu.au

Abstract

Musicians demonstrate advantages in acquiring motor sequences, showing faster learning and better explicit sequence knowledge than non-musicians. However, it is unclear whether this advantage extends beyond acquisition to the consolidation phase, which is when newly learned skills stabilize and become resistant to interference. Additionally, while interference from executing competing motor tasks is well-established, less is known about whether purely sensory information presented after learning can disrupt consolidation of a bimodal motor sequence.

We investigated how post-acquisition sensory interference affects performance of a learned audio-visual sequence, and whether musical training moderates this vulnerability. Participants first learned an explicit sequence in a serial reaction time task using synchronous, informative audio-visual cues. After a brief consolidation period, they were randomly assigned to one of four observational conditions that manipulated the relationship between auditory and visual streams. Motor performance was then reassessed.

Post-acquisition sensory interference significantly impaired subsequent motor performance, but this effect was modality-specific: it was driven primarily by manipulations to the task-relevant visual stream, while auditory interference alone had no credible effect. Distributional analysis revealed that learning involved a strategic shift from reactive to anticipatory responding. Critically, participants with musical training made this shift significantly faster than those without, demonstrating a more rapid adoption of predictive motor control.

These findings demonstrate that newly formed sensorimotor memories are selectively vulnerable to interference in task-relevant modalities. Furthermore, our work provides a mechanistic explanation for the musician advantage in sequence learning, linking it to faster development of predictive motor strategies during consolidation.

Keywords: motor sequence learning; sensory interference; memory consolidation; multisensory integration; musical training.

Introduction

Our memories are typically multisensory, combining sensory experiences encoded by specialised areas of our brains. For example, imagine watching a pianist perform. You see a finger pressing down a key, an action that is immediately followed by the characteristic sound that key makes. Once you learn the association between keys and sounds, you can predict one from the other: what key produced the sound, and what sound that key will produce. Through this learning process, experienced musicians can guide sequences of actions based on the predictions of sensory inputs (Chen *et al.*, 2006; Zatorre *et al.*, 2007; Chen *et al.*, 2008b; a; 2009; Matsushita *et al.*, 2021). Similarly, in experimental studies of sequence learning, auditory stimuli can be used as a cue that predicts the required action (Han *et al.*, 2023; Leow *et al.*, 2025), or as a form of feedback after the motor action (Hoffmann *et al.*, 2001; Stocker *et al.*, 2003; Stocker & Hoffmann, 2004).

Musical training requires acquisition of spatial-temporal sequences that are of far longer lengths than the 8 to 20 element sequences that are used in behavioural paradigms used to quantify sequence learning, such as in the serial reaction time task (SRTT) (Janata & Grafton, 2003). Perhaps unsurprisingly, several studies demonstrate a musician advantage in acquisition of sequences (Palmer, 2005; Tierney *et al.*, 2008; Anaya *et al.*, 2017; Sobierajewicz *et al.*, 2018). Specifically, musical training is associated with a faster rate of training-related reduction in reaction times during sequence learning (Palmer, 2005; Schwizer Ashkenazi *et al.*, 2022) or reduced errors in explicitly recalling previously learnt sequences (Tierney *et al.*, 2008; Sobierajewicz *et al.*, 2018). For example, one study showed that children with musical training showed an advantage in acquiring explicit knowledge of auditory sequences, reproducing more items from recently executed auditory sequences compared to children with similar volume of gymnastics training or video-game training (Tierney *et al.*, 2008). Indeed, some studies suggest a broad advantage in acquiring explicit sequence knowledge, even when sequences were presented only in the visual modality (Pau *et al.*, 2013; Anaya *et al.*, 2017; Schwizer Ashkenazi *et al.*, 2022), when sequences were executed via eye movements (Schwizer Ashkenazi *et al.*, 2022). By contrast, it is not clear if music training improves implicit learning processes, as previous studies show mixed results (Loui *et al.*, 2010; Francois & Schön, 2011; Romano Bergstrom *et al.*, 2012; Thorpe *et al.*, 2020). Musical training thus appears to augment the capacity to acquire explicit knowledge of sequences.

Whilst there is an increasing body of work demonstrating that musical training can improve the acquisition of explicit knowledge of sequences, less is known about how the effects of music training on learning extends beyond the acquisition and to the consolidation phase. Specifically, during consolidation, the passage of time after initial acquisition increases the robustness of sequence memories, helping newly acquired memories become more resistant to interference and more reliably retrieved (McGaugh, 2000; Alberini *et al.*, 2013). A large body of work demonstrates that the passage of time after training consolidates explicit motor memories in sequence learning tasks (Müller & Pilzecker, 1900; McGaugh, 2000; Walker *et al.*, 2002; Kelly *et al.*, 2003; Truong *et al.*, 2023). Although this

consolidation has previously been assumed to require the passage of hours after learning, newer evidence demonstrates that a rapid form of consolidation occurs at much shorter timescales of minutes after initial learning (Bönstrup *et al.*, 2019). During this period, sequence memories are initially labile, as they can be disrupted by learning a competing motor sequence (Robertson, 2012), and may also be either enhanced or disrupted through the simple observation of congruent or incongruent actions (Kelly *et al.*, 2003). However, such sequence memories become increasingly robust with increasing time after learning (Robertson, 2012). It is unclear if musical training alters the consolidation of sequence learning, as there is currently no direct evidence addressing this question. It is possible that musicians show enhanced consolidation of learning, as musical training is associated with more transfer of learning from one auditory-visual sequence to a second auditory-visual sequence (Palmer, 2005). Alternatively, musical training might increase the rigidity of sequence learning, as extensive musical training can reduce flexibility in acquisition of new auditory sequences when there are counterintuitive mappings between the response and the auditory feedback (i.e., where higher-pitched feedback is associated with keys on the left) (Pfordresher & Chow, 2019).

In the present study, we investigated whether musical training alters the consolidation of acquired sequence memories. Specifically, we asked whether musical training renders motor memories more robust against interference after a post-training consolidation phase. To test consolidation of the sequence learning, we first had participants complete an acquisition phase, where they acquired a 10-item motor sequence that was cued by synchronous auditory-visual stimuli, to establish a strong initial learning effect, similar to our earlier work (Leow *et al.*, 2025). After a 2-minute consolidation break, we then interfered with this learning by having participants learn a new, interfering sequence via observation alone. As some previous studies demonstrated that music training has modality-specific effects on sequence learning, where there is a selective benefit for auditory and not visual sequence learning (Tierney *et al.*, 2008), we manipulated the modality of the new, interfering sequence. Participants were assigned to one of four interference conditions: (1) a **Visual Interference** group observed a new visual sequence paired with the original auditory sequence; (2) an **Auditory Interference** group observed the original visual sequence with a new auditory sequence (i.e., a new mapping of sounds to keypresses); (3) an **Audio-visual Interference** group observed a new visual sequence paired with a new auditory sequence; and (4) a **No Interference** control group observed the exact same audio-visual sequence from the acquisition phase. This interference phase required no motor responses from participants, but they were told that they would have to report that sequence later.

While sequence learning is often assessed through reductions in averaged reaction times (RTs), this can obscure underlying cognitive changes. For example, performance improvements may reflect not just faster motor execution, but also a shift from reactive to predictive control during sequence learning. Supporting this, oculomotor tracking studies have shown that anticipatory eye movements offer a rich, real-time index of multiple, simultaneous learning processes that are not captured by averaged reaction times alone (Tal *et*

al., 2021). Similarly, studies that tightly control movement preparation time in sequence learning tasks suggest that movements during sequence learning can be pre-planned at least 3 movements in advance (Ariani & Diedrichsen, 2019). Increasing evidence also shows that such movement pre-planning occurs during execution of an ongoing movement (Ariani & Diedrichsen, 2019; Kashefi *et al.*, 2024). To capture these subtler dynamics, we examined the full distribution of reaction times rather than relying solely on mean values. This approach, inspired by recent work demonstrating the complexity of sequential effects, enables identification of distinct cognitive mechanisms that influence different portions of the RT distribution (Voormann & Miller, 2024). Moving beyond average RT is thus essential, as even indirect behavioural markers (such as prediction-driven shifts in response time distribution) might reveal learning dynamics that would otherwise remain hidden.

Methods

Participants

A total of 151 undergraduate psychology students were recruited from the Curtin University Psychology Participant Pool and participated in exchange for course credit. The sample included 115 females, 34 males, and 2 participants who identified as other, with a mean age of 20.5 years (SD = 2.74). Regarding experience with musical training, 74 participants (49.0%) reported having no experience with formal musical training, while the remaining 77 participants (51.0%) had at least some experience with formal musical training. Participants were asked about musical training at the end of the study session. All participants had normal or corrected-to-normal vision and hearing, and no known cognitive or neurological conditions that could have affected their performance. The study was approved by the Curtin University Human Research Ethics Committee (HRE2018-0257), and all participants provided informed consent prior to taking part in the study.

Sample size

We did not conduct a formal a priori power analysis for the present study. Instead, we aimed to recruit a sample size comparable to our previous work on bimodal sequence learning. The present study includes approximately 38 participants per condition, which exceeds the sample size of ~32 participants per group used in Leow *et al.* (2025). We reasoned that this sample size would provide adequate power to detect the effects of interest related to sensory interference.

Apparatus and Materials

Visual and auditory stimuli were presented using Inquisit 6 software (Version 6.6.0). Participants were seated in front of a 24-inch computer monitor (1920×1200, 60Hz) and used a standard QWERTY keyboard for responses. Sounds were presented through Corsair HS50 stereo headphones at a fixed intensity of 65dBa.

Task and Stimuli

Participants performed a multisensory serial reaction time (SRT) task. They were seated in front of a computer and placed the middle and index fingers of their left hand on the 'V' and 'B' keys, and the index and middle fingers of their right hand on the 'N' and 'M' keys, respectively. The screen displayed four grey squares in a horizontal row, which mapped spatially from left to right to the four response keys. On each trial, one square changed colour for 100 ms. This colour change occurred synchronously with a 100 ms auditory tone. The auditory stimuli were four pure tones (C4, D4, E4, and F4) that were mapped informatively to the four visual locations, creating a consistent audio-visual pairing during the acquisition phase (detailed later). Participants were instructed to respond as quickly and accurately as possible by pressing the key corresponding to the square that changed colour. Following a correct response, the trial ended, and a 500 ms inter-trial interval occurred before the next trial began (i.e., there was a 500 ms response-stimulus interval). However, in the interference phase, the stimulus-response mapping was altered as per the task condition.

Procedure

The task comprised three main phases: acquisition, interference, and revision (see Figure 1B for an overview). First, participants were familiarized with the task by completing 50 practice trials with random stimuli. They then began the Acquisition Phase (Blocks 1-5), which consisted of five blocks of 100 trials each (each block = 10 repetitions of a 10-item sequence per block): this resulted in a total of 500 trials for the acquisition phase. In this phase, all participants learned the same sequence with synchronous and informative audio-visual stimuli (Figure 1A). Following acquisition, a two-minute Consolidation Break was administered, during which participants did not perform any task: this timeframe is sufficient to allow rapid offline consolidation for sequence learning (Bönstrup *et al.*, 2019). Next, in the Interference Phase, participants were randomly assigned to one of four interference conditions. This phase consisted of a single block (100 trials, or 10 repetitions of a new or the original 10-item sequence) that was purely observational, requiring no motor responses, similar to previous work by Howard and colleagues (Howard *et al.*, 1992). The four conditions were: (1) a Visual Interference condition, in which participants passively observed a new visual sequence paired with the original auditory sounds (the original auditory cue order and auditory-response mapping was maintained, but the responses were now mapped to spatially different visual cues) ; (2) an Auditory Interference condition, in which participants passively observed the original visual sequence with new auditory sounds (the original visual sequence order was maintained and the visual cue-response mapping was maintained, but the auditory-response mapping was now altered) ; (3) an Audio-visual Interference condition, in which participants observed a new visual sequence paired with new sounds (both the visual and auditory cue order was altered); and (4) a No Interference control condition: this group observed the exact same audio-visual sequence from the acquisition phase. Participants were instructed that they would have to report the sequence they observed in this phase later. Subsequently, in the Revision Phase (Block 6), all participants performed one block of 100 trials of the original acquisition sequence to measure the effect of the observational interference on their motor performance. Finally, in the Recall phase, participants were asked

to explicitly recall the original sequence as well as the interference sequence, using mouse clicks.

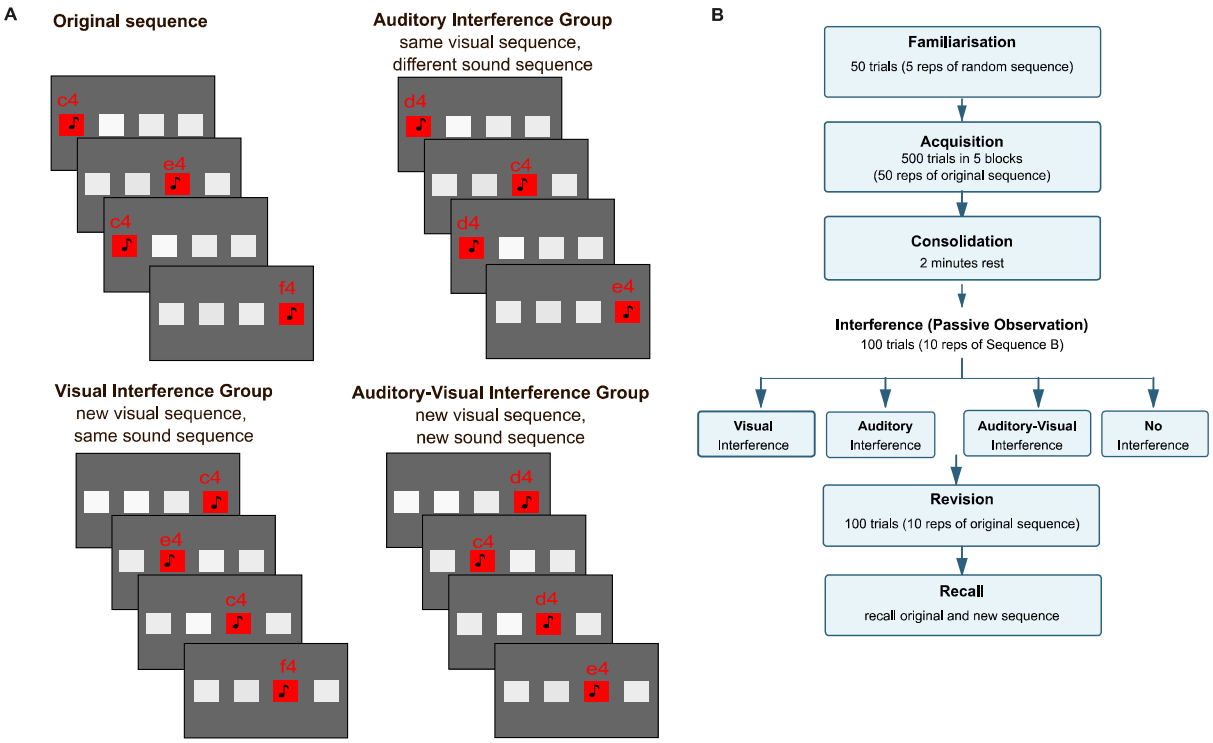

Figure 1. (A) ***Experimental design showing the four interference conditions****. Participants responded to visual cues (red squares) paired with corresponding auditory tones using keyboard presses with both hands. During acquisition, the original sequence (1,3,1,4,3,2,4,2,3,1) had each visual cue position informatively paired with a specific auditory tone (1=C4, 2=D4, 3=E4, 4=F4). Following acquisition, participants were assigned to one of four passive observation interference conditions that manipulated the congruence between visual and auditory cues. In the* ***Visual Interference group*** *(bottom left), participants passively observed a new visual sequence while the original auditory sequence was maintained. In the* ***Auditory Interference group*** *(top right), participants observed the original visual sequence paired with new auditory tones; the visual sequence order and visual cue-response mappings were preserved, but auditory-response mappings were altered. In the* ***Auditory-Visual Interference group*** *(bottom right), both visual and auditory sequences were changed, with responses mapped to new visual positions and paired with new auditory tones. A* ***No Interference control group*** *passively observed the original audiovisual sequence from the acquisition phase. (B) Schematic of the experimental procedure. Following familiarization (50 trials), participants completed an Acquisition phase (500 trials across 5 blocks) learning a 10-item sequence with synchronous audio-visual stimuli. After a 2-minute Consolidation break, participants were randomly assigned to one of four Interference conditions (100 trials of passive observation): Visual Interference (new visual sequence with original sounds), Auditory Interference (original visual sequence with new sounds), Auditory-Visual Interference (new visual and auditory sequences), or No Interference (original sequence). Subsequently, participants performed a Revision phase (100 trials) with the original sequence, followed by a Recall phase testing explicit memory for both sequences.*

Data Cleaning and Trial Exclusion

Prior to the main analyses, the data was processed to exclude trials with incorrect responses or slow reaction times (RT). From a total of 90,600 trials, responses were removed

if they were incorrect or if the RT exceeded a 1000 ms cutoff. This procedure led to the exclusion of 5,530 trials (6.10%) for incorrectness and 1,971 trials (2.18%) for exceeding the RT limit. After accounting for trials that met both criteria, a total of 6,862 unique trials were removed, representing 7.57% of the total data. Consequently, 83,738 trials (92.43%) were retained for the final analyses.

Data Analysis

All statistical analyses were performed using R (version 4.3.2) with the brms package (Bürkner, 2017) for Bayesian modelling. Our analytical strategy combined Bayes Factors (BF) with leave-one-out cross-validation (PSIS-LOO) to identify the most parsimonious models. Bayes Factors were interpreted using Jeffreys (1961) classification (e.g., BF > 3 indicates moderate evidence, BF > 10 strong evidence, BF > 30 very strong evidence). PSIS-LOO model comparisons were evaluated based on differences in expected log predictive density (ELPD); following Vehtari *et al.* (2017), a difference exceeding their ELPD by a factor of 2 or more provides evidence for improved out-of-sample predictive accuracy. For the selected models we report the median of the posterior distribution and the 95% Highest Posterior Density Interval (HPDI) for parameters of interest.

Following best practice in Bayesian modelling (Kruschke, 2021), we specified weakly informative, data-aware priors to improve estimation stability and guard against overfitting. For continuous outcomes, intercepts were given normal priors centred on the observed mean with a scale proportional to twice the observed standard deviation; slopes received normal priors centred on zero with a scale proportional to the observed standard deviation. Standard deviation parameters for group-level effects and residual error were assigned exponential priors scaled to the data, and Student-t likelihoods with a fixed degrees-of-freedom parameter were used where appropriate to provide robustness to outliers.

The main dependent variable for the primary learning analyses was normalised reaction time (RT), calculated as in Leow et al. (2025), to index learning gains relative to each participant's starting performance. RTs were aggregated to a single mean normalised RT per subject per experimental block. For the acquisition phase (Blocks 1–5), we compared a sequence of hierarchical models that included Block (centred), interference condition, and musical experience as predictors. All models incorporated random intercepts and random slopes for Block by subject (1 + Block ‖ subject) to capture individual learning trajectories. Model selection was based on a combination of Bayes Factors and PSIS-LOO. The best-supported model included two-way interactions between Block and the experimental factors, while adding a three-way interaction did not improve out-of-sample fit.

For the end-of-acquisition analysis (Block 5), we modelled mean normalised RT with interference condition and musical experience as fixed factors and subject as a random intercept. The main interference analysis (Block 6) was implemented as an ANCOVA with the mean normalised RT from Block 5 entered as a covariate to control for baseline performance. Competing models ranged from a null model with only the covariate, to models adding condition and musical experience main effects and their interaction.

A second set of analyses targeted explicit memory measures collected at the end of the experiment. We fitted Bayesian models (Student-t likelihood with robust priors) to explicit memory scores for the practiced and interference sequences. Finally, we examined distributional aspects of RTs. A two-component normal mixture model on log-transformed latencies provided a data-driven threshold separating fast, anticipatory responses from slower, reactive responses. Using this threshold, we computed (1) the proportion of anticipatory responses per block, analysed with a Bayesian linear mixed model, and (2) a binary bimodal classification for each participant and block, analysed with a Bernoulli Bayesian GLMM. This allowed us to characterise how the prevalence of anticipatory and bimodal response patterns evolved across practice and differed by musical experience.

Results

Acquisition Phase: Blocks 1–5

During the acquisition phase, participants in all groups experienced identical synchronous and informative audio-visual stimuli. Still, we sought to test for pre-existing group differences during this acquisition phase prior to the primary intervention in the Interference phase. To this end, we compared a series of Bayesian linear mixed-effects models with weakly informative priors. A model including two-way interactions between Block, condition, and musical experience was compared against a simpler main-effects-only model. This comparison yielded strong evidence in favour of the interaction model (BF = 5 × 10^38; PSIS-LOO ΔELPD ≈ +2.2). We then compared this model to a full three-way interaction model and found extreme evidence against including the three-way interaction (BF < 0.001; PSIS-LOO ΔELPD ≈ -0.7). All posterior diagnostics indicated excellent MCMC mixing (all max $\hat{R} \leq 1.01$, all effective sample sizes > 4,000, and zero divergent transitions).

Therefore, we interpreted the winning two-way interaction model. The analysis revealed two key findings. First, there was a credible interaction between Block and condition, indicating that the rate of learning differed significantly among the four experimental groups (see Figure 2A). Second, a credible Block × musical experience interaction (Median = 0.03, 95% HPDI [0.01, 0.06]) revealed that participants with music training learned at a significantly faster rate than those without.

Performance at the End of Acquisition (Block 5)

Given the different learning rates, we formally tested for performance differences at the end of the acquisition phase (Block 5). A Bayesian model was fitted to the mean normalised reaction times from this block alone. PSIS-LOO indicated that the simple model with condition and musical experience main effects adequately captured the data. Convergence diagnostics again indicated excellent mixing (max $\hat{R} \leq 1.01$, all effective sample sizes > 4,500, no divergent transitions). Pairwise comparisons between the conditions (averaged over musical experience) revealed one credible difference: participants in the Audio-visual Interference group were slower than those in the No Interference group (Median Difference = -0.14, 95% HPDI [-0.29, -0.01]), as shown in Figure 2B. These baseline differences confirm

the necessity of controlling for Block 5 performance in the subsequent analyses of the interference phase at Block 6, described later. Prior to describing these analyses, we first report on further analyses on the response time distributions during acquisition.

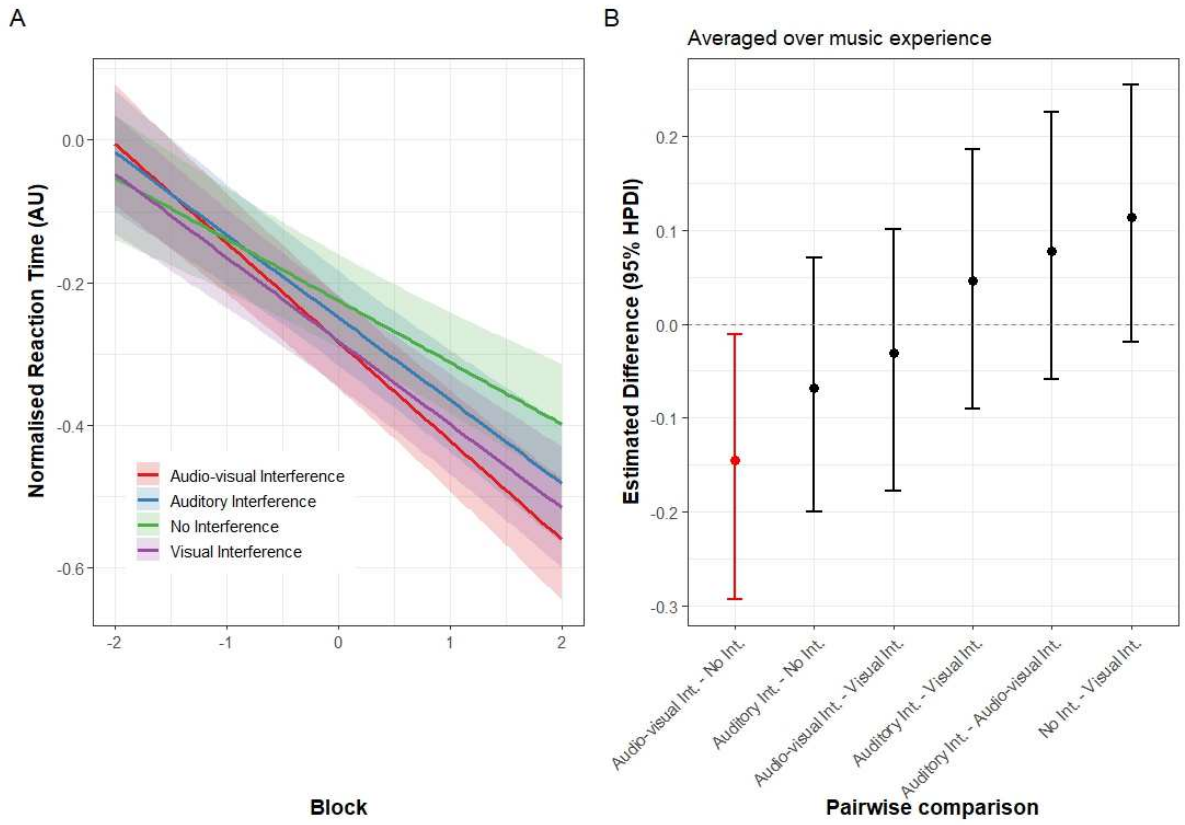

Figure 2. *Analysis of the Learning Phase and Performance at the End of Acquisition. (**A**) Conditional effects from the Bayesian model showing the estimated learning rate for each of the four Groups across the five acquisition blocks. Solid lines represent the posterior mean estimates for Normalized Reaction Time (AU), and the shaded ribbons represent the 95% credible intervals. The different slopes illustrate the significant interaction between Block and Condition, indicating that the groups learned at different rates. (**B**) Bayesian pairwise comparisons of the mean Normalized Reaction Time between conditions at the final acquisition block (Block 5). Points represent the median of the posterior distribution for the estimated difference, and the vertical bars represent the 95% Highest Posterior Density (HPD) intervals. The comparison highlighted in red is the only one where the 95% HPD interval does not cross zero, indicating a credible performance difference between the Audio-visual Interference and No Interference groups immediately before the manipulation phase.*

Analysis of Response Time Distributions During Acquisition

While the initial analyses focused on mean reaction time, an inspection of the trial-level data suggested a more complex underlying structure. Specifically, the distribution of reaction times appeared to be bimodal, with responses clustering into a fast, anticipatory mode and a slower, reactive mode (Figure 3A). To gain a more mechanistic understanding of how learning unfolded, we conducted a subsequent analysis to explicitly model these two response components.

Practice Increases Anticipatory Responding During Acquisition of Sequences

To quantify the shift from reactive to predictive control, we first defined a data-driven threshold to separate the two response modes. A two-component mixture model was fitted to the distribution of all log-transformed reaction times from the acquisition phase, yielding an anticipation threshold of 182.01 ms. As shown in Figure 4B, the distribution of reaction times clearly shifted across the acquisition phase, with a growing peak of fast, anticipatory responses emerging over time.

To formally test this, we first examined the overall trend across all participants. A Bayesian mixed-effects model revealed a decisive effect of practice, showing that the proportion of anticipatory responses (latencies < 182.01 ms) increased significantly across blocks (Median = 0.95, 95% HPDI [0.85, 1.05]; max R-hat = 1.001; Bulk ESS ≈ 1121; Tail ESS ≈ 2132; see Figure 4C). Similarly, we classified each participant on a block-by-block basis as either showing a bimodal latency distribution or not. A Bayesian logistic mixed-effects model confirmed that as participants practiced, they were significantly more likely to develop a bimodal response pattern (Median = 2.03, 95% HPDI [1.55, 2.62]; max R-hat = 1.001; Bulk ESS ≈ 4079; Tail ESS ≈ 6758; see Figure 3D).

Music Training Increases Anticipatory Responding During Acquisition of Sequences

To understand the source of the different learning rates observed in our initial analysis, we next examined how music training moderated this strategic shift towards anticipatory responding during acquisition.

For the proportion of anticipatory responses, we compared a model including an interaction between Block and music training against a simpler main-effects-only model. This yielded evidence from Bayes factors favouring the main-effects model (BF for interaction vs. main = 0.095). However, PSIS-LOO cross-validation favoured the interaction model, with an elpd difference of ~17.83 (SE = 7.08) and lower looic (7089.71 vs 7125.36), indicating superior out-of-sample predictive accuracy for the interaction model. As illustrated in Figure 3A, while both groups increased their use of predictive actions over time, the rate of this shift differed. The credible interaction term (Median = -0.04, 95% HPDI [-0.05, -0.02]) indicates that participants with formal musical training experience adopted an anticipatory strategy at a significantly faster rate than those without.

Next, we examined the likelihood of participants developing a bimodal reaction time distribution. Here, a model comparison provided evidence against including an interaction term (BF = 0.09; PSIS-LOO showed no meaningful difference, Δelpd ≈ –0.08, SE ≈ 0.77; looic 487.87 vs 488.04), indicating the main-effects model was preferred for parsimony. This winning model revealed two key effects. First, there was a decisive main effect of Block (Median = 1.98, 95% HPDI [1.51, 2.56]), confirming that all participants were more likely to become bimodal with practice. Second, there was a decisive main effect of musical experience (Median = -1.52, 95% HPDI [-2.74, -0.29]), indicating that, overall, participants with formal

musical experience were significantly more likely to be classified as bimodal responders than those without musical experience (Figure 5B).

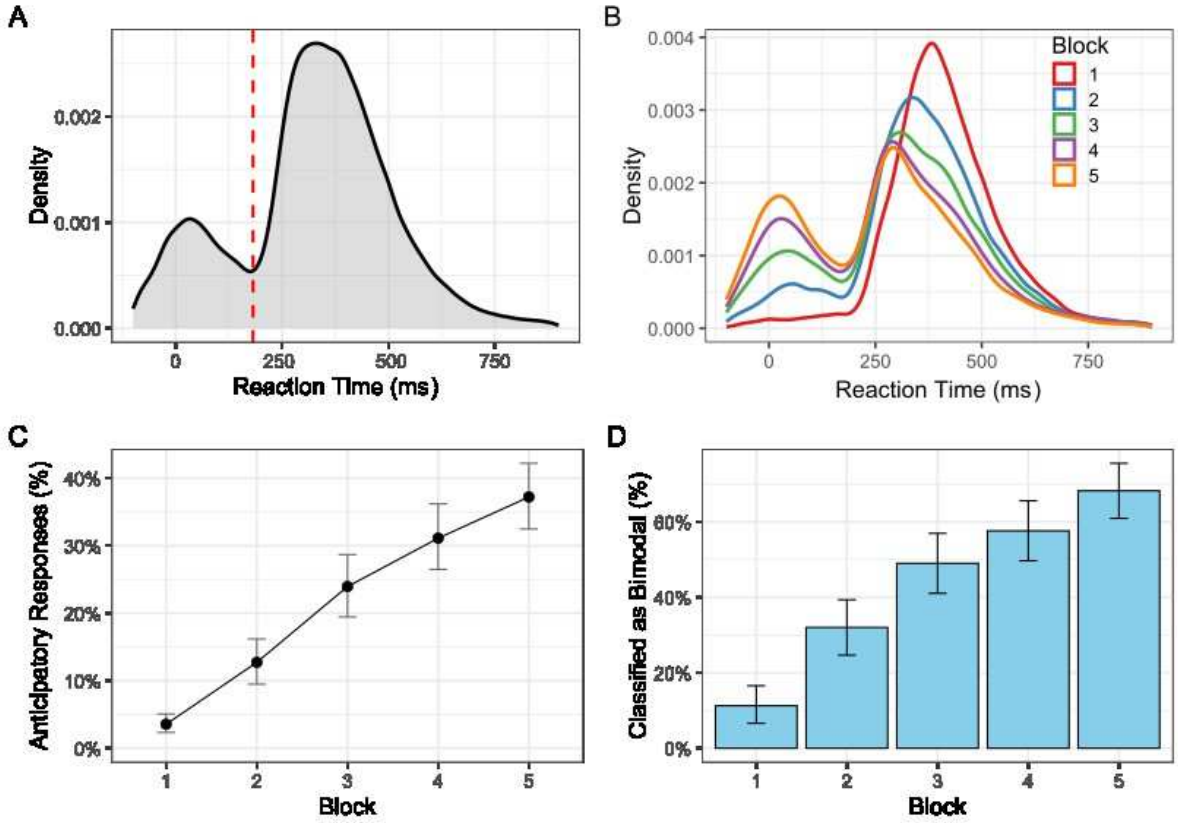

*Figure 3. Emergence of a Bimodal Distribution in Reaction Times During Learning. (**A**) The density plot of all reaction times (RTs) from the acquisition phase, collapsed across all blocks and participants. The plot clearly shows a bimodal distribution, suggesting the presence of two distinct response types: a fast, anticipatory mode and a slower, reactive mode. (**B**) The density plot of RTs separated by training Block. This panel illustrates the dynamic shift in response strategy during learning. In Block 1 (red line), the distribution is largely unimodal and centred on slower, reactive responses. As practice progresses to Block 5 (orange line), a distinct peak of fast, anticipatory responses emerges and grows, indicating an increasing reliance on predictive motor control. (**C**) The plot shows the mean proportion of anticipatory responses across the five acquisition blocks. Points represent the mean, and the vertical bars represent the 95% bootstrapped confidence intervals. (**D**) The bars represent the mean proportion of subjects classified as "bimodal" in each of the five acquisition blocks. Error bars represent the 95% bootstrapped confidence intervals.*

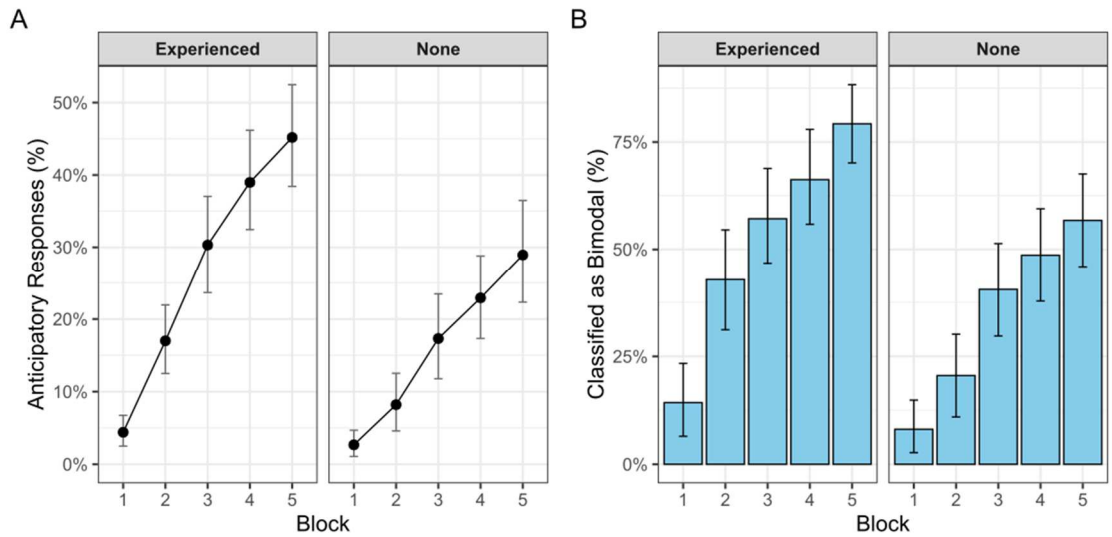

Figure 4. Moderating Effect of Experience with Musical Training on the Emergence of Anticipatory Responses During Acquisition. ***(A)*** *The proportion of anticipatory responses across the five acquisition blocks, plotted separately for participants with ("Experienced") and without ("None") musical training. Points are the means and error bars represent 95% bootstrapped confidence intervals.* ***(B)*** *The proportion of participants classified as having a bimodal RT distribution in each block, separated by experience with musical training. Bars represent the mean proportion and error bars represent the 95% bootstrapped confidence intervals.*

Interference from Observing a New Visual Sequence

We next tested how observing a new audio/visual/audio-visual sequence interfered with the newly formed sequence memory during the subsequent Revision Phase (Block 6). Given the baseline performance differences at acquisition, we statistically controlled for baseline performance (Block 5 RT) using a Bayesian ANCOVA. Specifically, a model including the main effects of condition and musical experience was compared to a null model containing only the baseline covariate. This provided weak but consistent evidence in favour of the main effects model (BF = 2.45), with PSIS-LOO indicating a clear predictive advantage (ΔELPD $\approx$ +8.6), suggesting that the experimental manipulations improved model fit beyond baseline. We then compared the main effects model to a full interaction model and found extreme evidence against including the interaction term (BF < 0.01; PSIS-LOO ΔELPD $\approx$ -3.3). Convergence diagnostics confirmed excellent MCMC mixing (all max $\hat{R} \leq 1.01$, all effective sample sizes > 4,700, and no divergent transitions).

Based on the winning main effects model, we found a credible main effect of condition on post-intervention performance. As shown in Figure 5, pairwise comparisons revealed that, after adjusting for baseline performance, the Visual Interference group (Median Difference vs. No Int. = 0.07, 95% HPDI [0.01, 0.14]) and the Audio-visual Interference group (Median Difference vs. No Int. = 0.08, 95% HPDI [0.01, 0.15]) were slower than the No Interference group. Furthermore, both the Visual Interference (Median Difference vs. Auditory Int. = 0.14, 95% HPDI [0.07, 0.21]) and Audio-visual Interference groups (Median Difference vs.

Auditory Int. = 0.14, 95% HPDI [0.07, 0.21]) were slower than the Auditory Interference group. There was no credible difference between the Auditory Interference and No Interference conditions (Median Difference = 0.06, 95% HPDI [-0.01, 0.13]). These results indicate that interference was mainly driven by manipulations of the task-relevant visual stream.

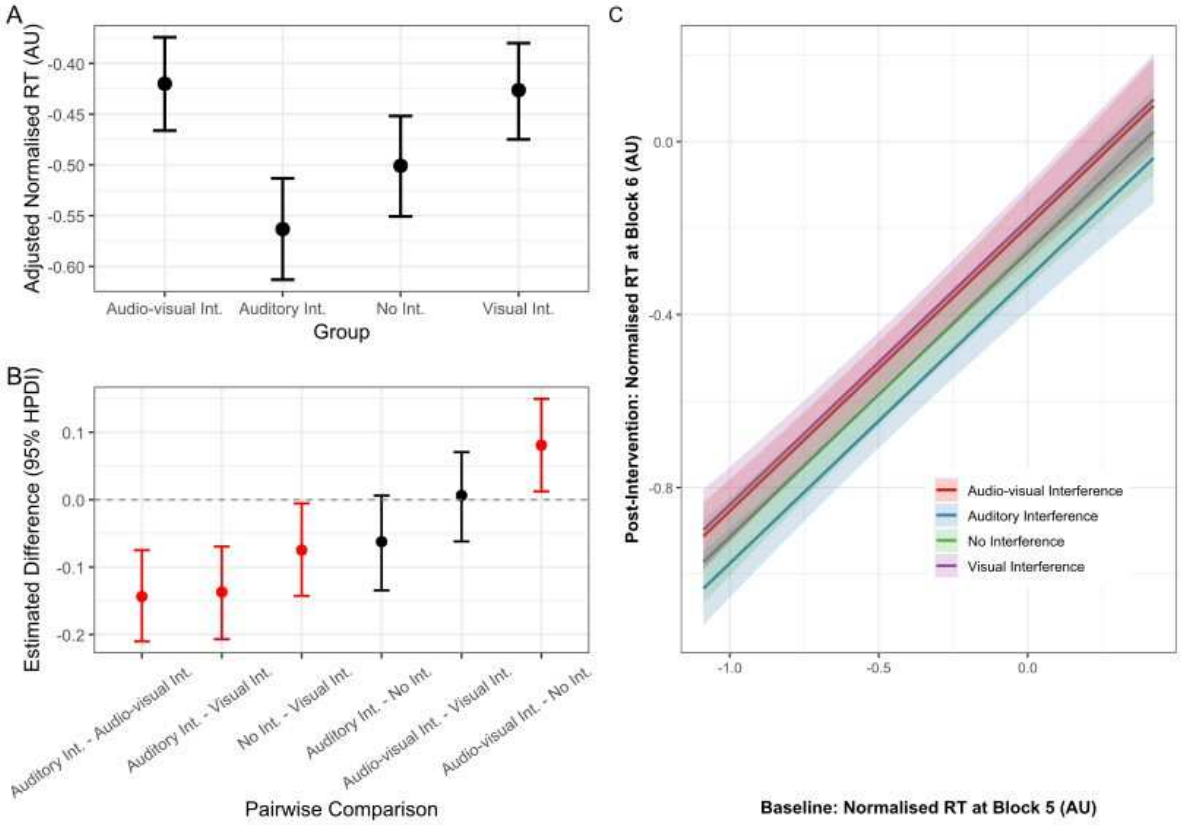

Figure 5. ***Bayesian ANCOVA Results for Post-Intervention Performance.*** *(**A**) The plot displays the estimated marginal means for Normalized Reaction Time (arbitrary units, A.U.) in the post-intervention Revision Phase (Block 6) for each of the four experimental conditions, where greater interference (i.e., increases in normalised reaction time) is evident in all groups who observed a new visual sequence (visual interference, auditory-visual interference) than groups who heard a new auditory sequence (auditory interference) and groups who re-observed the same auditory and visual sequence (the no interference group). All estimates are adjusted for baseline performance at Block 5. The points represent the posterior median estimates, and the vertical bars represent the 95% credible intervals. (**B**) This panel shows the Bayesian pairwise comparisons of the estimated differences between each experimental condition. Points represent the median of the posterior distribution for the estimated difference, and the vertical bars show the 95% Highest Posterior Density (HPD) intervals. Comparisons highlighted in red indicate a credible difference between groups, as their 95% HPD interval does not overlap the zero line. (**C**) Predicted post-intervention performance (Revision Phase, Block 6) is shown as a function of baseline performance (Block 5) for each condition. The parallel lines, derived from the ANCOVA model, illustrate that the intervention had a consistent effect across the range of baseline scores. The vertical separation between the lines represents the main effect of the interference condition,*

providing a visual demonstration that intervention impact is not explained by a regression to the mean.

Explicit Memory

Better Explicit Memory for the Practiced (Acquisition) Sequence With Music Training

To assess participants' explicit memory for the sequence they practiced in the Acquisition Phase, we quantified recall in the Revision Phase, where participants used mouse clicks to recall the practiced sequence. We fitted a Bayesian linear model with condition and musical experience as interacting predictors of the recall score. The analysis did not reveal a credible interaction between the interference condition and musical experience. As shown in Figure 4A (left panel), participants with and without experience with musical training demonstrated broadly similar recall scores across all conditions.

Pairwise comparisons of the difference between participants with and without formal musical training within each condition further confirmed the lack of a robust musical training advantage (Figure 6A, right panel). Median differences between those with and without formal musical experience were small and their 95% HPDIs all included zero (e.g., Audio-visual Interference: Median Difference ≈ 1.15, 95% HPDI [-1.10, 2.38]; Visual Interference: Median Difference ≈ 0.67, 95% HPDI [-0.57, 1.93]; No Interference: Median Difference ≈ 0.67, 95% HPDI [-0.55, 1.85]). All posterior diagnostics indicated excellent convergence and stability (all $\hat{R} \leq 1.01$, effective sample sizes > 4,300, no divergent transitions; PSIS-LOO elpd_loo = -341.2, all Pareto-k < 0.7). Overall, these results suggest that those with and without formal musical experience had comparable explicit recall of the originally practiced sequence, with no reliable evidence of a musician advantage in this task.

Better Explicit Memory for the Observed (Interference) Sequence With Music Training

We next analysed participants' memory for the second sequence observed during the manipulation phase. The results pointed towards an interaction between condition and musical experience (Figure 6B, left panel). Pairwise comparisons showed that participants experienced with formal music training outperformed those without music training in remembering the new sequence when it involved visual changes, with credible evidence for this difference in the Visual Interference group (Median Difference = 2.37, 95% HPDI [0.55, 4.32]) and the Audio-visual Interference group (Median Difference = 2.43, 95% HPDI [0.50, 4.28]). As expected, there was no musical experience advantage in the No Interference condition (Median Difference = 1.08, 95% HPDI [-0.609, 2.94]). All posterior diagnostics confirmed excellent mixing and convergence (all $\hat{R} \leq 1.01$, all effective sample sizes > 3,900, no divergent transitions). This indicates that non-musicians struggled more to encode the new, interfering visual sequence itself.

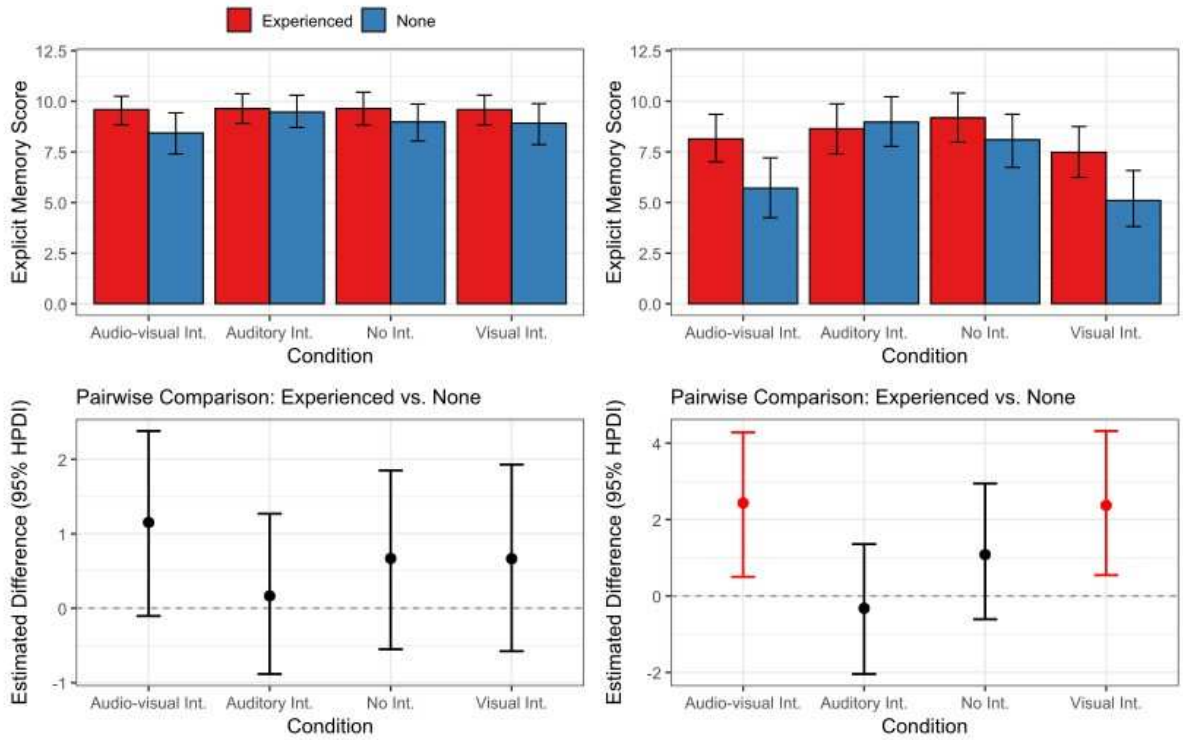

Figure 6. ***Bayesian Analysis of Explicit Memory Scores. (A) Memory for the Practiced (Acquisition) Sequence.*** *The top panel displays the estimated marginal mean recall scores for each condition, separated by musical experience. Bars represent the posterior median estimates, and the vertical error bars represent the 95% Highest Posterior Density Intervals (HPDIs). The bottom panel shows the pairwise comparisons of the estimated difference between experienced and non-experienced participants within each condition. Points represent the median of the posterior distribution for the estimated difference, and the vertical bars show the 95% HPDIs.* ***(B) Memory for the Observed (Interference) Sequence.*** *The top panel displays the estimated marginal mean recall scores for the second, observed sequence. The bottom panel shows the corresponding pairwise comparisons between experience groups for each condition. In both comparison plots, intervals not crossing the zero line indicate a credible performance difference between those with experience of music training and those without.*

Discussion

The present study explored the effect of musical training on the capacity to acquire and consolidate distinct sequence memories, and characterised how musical training alters anticipatory response patterns in sequence learning. We found that musical training not only increased the capacity to acquire explicit knowledge of the practiced sequence during the acquisition blocks, but also enhanced the consolidation of sequences, increasing the robustness of the practiced sequence from interference resulting from observing and learning a distinct visual sequence in the interference block. Furthermore, the response times distributions during

the acquisition block also differed depending on musical training. Response times distributions became increasingly bimodal with increasing practice of the sequence in the acquisition block, demonstrating a fast, anticipatory mode and a slower, reactive mode of response. Those with musical training adopted an anticipatory strategy at a faster rate than those without, and were also more likely to be classified as bimodal responders than those without musical training. Overall, our findings demonstrate that musical training confers multiple advantages in sequence learning: increasing the robustness of sequence memories against interference, increasing explicit recall of the interfering sequence, and increasing predictive control earlier during acquisition of a motor sequence.

Musical Training Increased the Acquisition and the Robustness of Explicit Sequence Learning

Musical training requires the concurrent acquisition of spatial sequences as well as temporal sequences, where sequences are typically of far longer lengths than the 8 to 20 element sequences that are used in sequence learning tasks like the SRTT (Janata & Grafton, 2003). It is perhaps unsurprising then that musicians show an advantage in sequence learning tasks with auditory-visual sequences (Palmer, 2005; Pau *et al.*, 2013; Anaya *et al.*, 2017), as this constitutes a form of near transfer. Importantly, whilst previous work has demonstrated that musical training improves acquisition of sequences, relatively little has been known about whether musical training alters the *consolidation* of sequence learning, such that memories becomes more robust against interference. Furthermore, exactly what components of learning are susceptible to interference remain incompletely understood. Previous work has shown that perceptual experience of a sequence is sufficient for acquisition of sequence learning (Howard *et al.*, 1992), although it is less clear if perceptual sequences can also cause interference. We show here that perceptual sequences can cause interference, but not all forms of perceptual stimuli cause interference. Only sequence stimuli presented in the visual modality interfered with previous learning, slowing reaction times when re-executing the original sequence, both in those with and without musical training. Listening to a different auditory sequence did not interfere with previous learning, similar to previous work in expert musicians (Brown & Palmer, 2013). Our results showed that musical training failed to render sequence learning immune from interference from observing a new visual or auditory-visual sequence, slowing reaction times when re-executing the original sequence. What might explain the modality-specificity of this interfering effect? We propose that despite presenting the original sequence with both auditory and visual stimulus information at acquisition, learning occurred predominantly by mapping spatial information of the visual cues to the spatial information of the upcoming action, whereas auditory stimuli served mainly as an accessory stimulus. In other words, learning occurred in a task-relevant domain: where the spatial locations of the cues helped encode the upcoming action. However, the interference effects shown from reaction time decrements in both those with and without musical training contrast with the explicit recall data, where those with musical training were more resilient against interfering effects from the visual sequence when explicitly recalling the initially acquired sequence, as well as the second, interfering sequence.

We note that our task allows concurrent engagement of implicit and explicit learning processes. It remains to be seen if the musical training advantage in both acquiring and consolidating explicit sequence memories here would also be evident for implicit sequence learning processes, as previous findings have been mixed (Loui *et al.*, 2010; Francois & Schön, 2011; Romano Bergstrom *et al.*, 2012; Thorpe *et al.*, 2020). The separable mechanisms of implicit and explicit sequence learning has long been proposed (Reber & Squire, 1998), and evidence suggests that consolidation processes for implicit and explicit sequence learning depend on distinct neural processes (Hirano *et al.*, 2017). For example, perturbational TMS experiments show that offline improvements following the explicit sequence learning depend on the primary motor cortex and not on the inferior parietal lobe, whilst offline improvements following the implicit sequence learning dependent on the inferior parietal lobe but not the primary motor cortex (Breton & Robertson, 2017). It is plausible that the explicit representation formed in our task is less reliant on this low-level motor trace and is more dependent on higher-order cognitive structures, the very networks implicated in long-term musical training (Ihalainen *et al.*, 2023). It remains to be seen if sequence learning paradigms that dissociate the implicit component of learning (e.g., probabilistic serial reaction time tasks, contextual cueing, or statistical learning paradigms) would reveal similar advantages for musicians. Examining these processes more systematically could help clarify whether musical expertise enhances domain-general implicit sensitivity to sequential regularities, or whether the advantage is more specific to explicit strategic learning and higher-order representational processes.

Prior Musical Training Increased Anticipatory Responding During Sequence Learning s

Our analysis of fast, anticipatory reaction times during sequence learning provides a more direct window into the development of predictive processing during sequence learning than standard approaches using mean reaction times. Our analysis supports a longstanding view that a characteristic feature of sequence learning is a shift from a reactive mode of responding to an anticipatory one (Ghilardi *et al.*, 2003). Importantly, this shift was faster for participants with musical training, supporting the premise that prior expertise can facilitate the development of predictive models for action. Although some previous studies have posited that a shift to anticipatory responding occurs more rapidly in individuals with musical training (Engel *et al.*, 1997; Palmer, 2005)— supporting the idea that prior expertise may facilitate the development of predictive models for action— there is relatively limited direct evidence to substantiate this claim. For example, Engel *et al.* (1997) demonstrated anticipatory adjustments in skilled pianists' hand and finger kinematics, but their small sample included a mix of experience levels and did not systematically test training effects; anticipatory modifications were notably task-dependent and variable across both pieces and individuals. Importantly, most available studies compare anticipatory response patterns during sequence learning in child and adults with varying musical experience, and do not isolate the effect of music training from age-related cognitive development. For example, Palmer and Pfordresher (2003) contrasted child novices with adult musicians, but the design did not control for age, making it difficult to attribute observed group differences in

anticipatory response patterns specifically to training rather than developmental differences. On the other hand, Drake and Palmer (2000) show that increases in anticipatory planning are closely linked to both age and accumulated musical experience, with age often exerting a stronger influence than formal training alone. Our findings are in line with a recent study that showed more anticipatory eye movements in musicians when sequences were learned via eye movements (Schwizer Ashkenazi *et al.*, 2022). Extending this line of work, we provide clear evidence for an anticipatory mode of responding during sequence learning that is augmented by practice and by musical experience.

Exactly what mechanisms drive this pattern of anticipatory responding in those with music training? One possibility is that music training results in a generalized improvement in predictive processing. Music training improves the brain's predictive processing of exogenous stimuli, such that musicians are better than non-musicians at predicting upcoming items in auditory sequences, even when those sequences have complex or high-order structures (Pesnot Lerousseau & Schön, 2021). Another possibility is that musicians are faster and better at integrating multisensory information (Landry & Champoux, 2017) which would allow them to form a more robust audio-visual-motor representations during acquisition.

Conclusion

In conclusion, our findings demonstrate that the stability of a new sensorimotor memory is contingent on the relevance of the sensory modality being challenged. The acquisition process itself is a strategic shift toward prediction, a shift that is accelerated in musicians. We speculate that this predictive pattern of behaviour results from an enhancement of latent processes with musical training, such as multisensory integration and/or predictive temporal attention.

References

Alberini, C.M., Johnson, S.A. & Ye, X. (2013) Chapter five - Memory Reconsolidation: Lingering Consolidation and the Dynamic Memory Trace. In Alberini, C.M. (ed) *Memory Reconsolidation*. Academic Press, San Diego, pp. 81-117.

Anaya, E.M., Pisoni, D.B. & Kronenberger, W.G. (2017) Visual-spatial sequence learning and memory in trained musicians. *Psychol Music*, **45**, 5-21.

Ariani, G. & Diedrichsen, J. (2019) Sequence learning is driven by improvements in motor planning. *Journal of Neurophysiology*, **121**, 2088-2100.

Bönstrup, M., Iturrate, I., Thompson, R., Cruciani, G., Censor, N. & Cohen, L.G. (2019) A rapid form of offline consolidation in skill learning. *Current Biology*, **29**, 1346-1351. e1344.

Breton, J. & Robertson, E.M. (2017) Dual enhancement mechanisms for overnight motor memory consolidation. *Nat Hum Behav*, **1**, 0111.

Brown, R.M. & Palmer, C. (2013) Auditory and motor imagery modulate learning in music performance. *Front Hum Neurosci*, **7**, 320.

Bürkner, P.-C. (2017) brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, **80**, 1-28.

Chen, J.L., Penhune, V.B. & Zatorre, R.J. (2008a) Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex*, **18**, 2844-2854.

Chen, J.L., Penhune, V.B. & Zatorre, R.J. (2008b) Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J Cogn Neurosci*, **20**, 226-239.

Chen, J.L., Penhune, V.B. & Zatorre, R.J. (2009) The Role of Auditory and Premotor Cortex in Sensorimotor Transformations. In DallaBella, S., Kraus, N., Overy, K., Pantev, C., Snyder, J.S., Tervaniemi, M., Tillmann, B., Schlaug, G. (eds) *Neurosciences and Music Iii: Disorders and Plasticity*, pp. 15-34.

Chen, J.L., Zatorre, R.J. & Penhune, V.B. (2006) Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, **32**, 1771-1781.

Drake, C. & Palmer, C. (2000) Skill acquisition in music performance: relations between planning and temporal control. *Cognition*, **74**, 1-32.

Engel, K.C., Flanders, M. & Soechting, J.F. (1997) Anticipatory and sequential motor control in piano playing. *Experimental brain research*, **113**, 189-199.

Francois, C. & Schön, D. (2011) Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, **21**, 2357-2365.

Ghilardi, M.-F., Eidelberg, D., Silvestri, G. & Ghez, C. (2003) The differential effect of PD and normal aging on early explicit sequence learning. *Neurology*, **60**, 1313-1319.

Han, Z., Sanchez, D., Levitan, C.A. & Sherman, A. (2023) Stimulus-locked auditory information facilitates real-time visuo-motor sequence learning. *Psychonomic Bulletin & Review*.

Hirano, M., Kubota, S., Koizume, Y., Tanaka, S. & Funase, K. (2017) Different Effects of Implicit and Explicit Motor Sequence Learning on Latency of Motor Evoked Potential Evoked by Transcranial Magnetic Stimulation on the Primary Motor Cortex. *Front Hum Neurosci*, **Volume 10 - 2016**.

Hoffmann, J., Sebald, A. & Stöcker, C. (2001) Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27**, 470.

Howard, J.H., Mutter, S.A. & Howard, D.V. (1992) Serial pattern learning by event observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 1029.

Ihalainen, R., Kotsaridis, G., Vivas, A.B. & Paraskevopoulos, E. (2023) The relationship between musical training and the processing of audiovisual correspondences: Evidence from a reaction time task. *PLoS One*, **18**, e0282691.

Janata, P. & Grafton, S.T. (2003) Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature neuroscience*, **6**, 682-687.

Jeffreys, H. (1961) *Theory of probability*. Clarendon Press, Oxford,.

Kashefi, M., Reschechtko, S., Ariani, G., Shahbazi, M., Tan, A., Diedrichsen, J. & Pruszynski, J.A. (2024) Future movement plans interact in sequential arm movements. *Elife*, **13**, RP94485.

Kelly, S.W., Burton, A., Riedel, B. & Lynch, E. (2003) Sequence learning by action and observation: Evidence for separate mechanisms. *British journal of psychology*, **94**, 355-372.

Kruschke, J.K. (2021) Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, **5**, 1282-1291.

Landry, S.P. & Champoux, F. (2017) Musicians react faster and are better multisensory integrators. *Brain and Cognition*, **111**, 156-162.

Leow, L.-A., Nguyen, A., Corti, E. & Marinovic, W. (2025) Informative Auditory Cues Enhance Motor Sequence Learning. *European Journal of Neuroscience*, **61**, e70140.

Loui, P., Wessel, D.L. & Kam, C.L.H. (2010) Humans rapidly learn grammatical structure in a new musical scale. *Music perception*, **27**, 377-388.

Matsushita, R., Puschmann, S., Baillet, S. & Zatorre, R.J. (2021) Inhibitory effect of tDCS on auditory evoked response: Simultaneous MEG-tDCS reveals causal role of right auditory cortex in pitch learning. *Neuroimage*, **233**, 117915.

McGaugh, J.L. (2000) Memory - A century of consolidation. *Science*, **287**, 248-251.

Müller, G.E. & Pilzecker, A. (1900) *Experimentelle beiträge zur lehre vom gedächtniss*. Ja Barth.

Palmer, C. (2005) Sequence memory in music performance. *Current Directions in Psychological Science*, **14**, 247-250.

Palmer, C. & Pfordresher, P.Q. (2003) Incremental planning in sequence production. *Psychological Review*, **110**, 683.

Pau, S., Jahn, G., Sakreida, K., Domin, M. & Lotze, M. (2013) Encoding and recall of finger sequences in experienced pianists compared with musically naive controls: a combined behavioral and functional imaging study. *Neuroimage*, **64**, 379-387.

Pesnot Lerousseau, J. & Schön, D. (2021) Musical expertise is associated with improved neural statistical learning in the auditory domain. *Cerebral Cortex*, **31**, 4877-4890.

Pfordresher, P.Q. & Chow, K. (2019) A cost of musical training? Sensorimotor flexibility in musical sequence learning. *Psychonomic Bulletin & Review*, **26**, 967-973.

Reber, P.J. & Squire, L.R. (1998) Encapsulation of implicit and explicit memory in sequence learning. *Journal of cognitive neuroscience*, **10**, 248-263.

Robertson, E.M. (2012) New insights in human memory interference and consolidation. *Current Biology*, **22**, R66-R71.

Romano Bergstrom, J.C., Howard Jr, J.H. & Howard, D.V. (2012) Enhanced implicit sequence learning in college-age video game players and musicians. *Applied Cognitive Psychology*, **26**, 91-96.

Schwizer Ashkenazi, S., Raiter-Avni, R. & Vakil, E. (2022) The benefit of assessing implicit sequence learning in pianists with an eye-tracked serial reaction time task. *Psychological Research*, **86**, 1426-1441.

Sobierajewicz, J., Naskręcki, R., Jaśkowski, W. & Van der Lubbe, R.H. (2018) Do musicians learn a fine sequential hand motor skill differently than non-musicians? *PLoS One*, **13**, e0207449.

Stocker, C. & Hoffmann, J. (2004) The ideomotor principle and motor sequence acquisition: tone effects facilitate movement chunking. *Psychol Res*, **68**, 126-137.

Stocker, C., Sebald, A. & Hoffmann, J. (2003) The influence of response--effect compatibility in a serial reaction time task. *Q J Exp Psychol A*, **56**, 685-703.

Tal, A., Bloch, A., Cohen-Dallal, H., Aviv, O., Schwizer Ashkenazi, S., Bar, M. & Vakil, E. (2021) Oculomotor anticipation reveals a multitude of learning processes underlying the serial reaction time task. *Scientific Reports*, **11**, 6190.

Thorpe, L., Cousins, M. & Bramwell, R. (2020) Implicit knowledge and memory for musical stimuli in musicians and non-musicians. *Psychology of Music*, **48**, 836-845.

Tierney, A.T., Bergeson-Dana, T.R. & Pisoni, D.B. (2008) Effects of early musical experience on auditory sequence memory. *Empirical musicology review: EMR*, **3**, 178.

Truong, C., Ruffino, C., Gaveau, J., White, O., Hilt, P.M. & Papaxanthis, C. (2023) Time of day and sleep effects on motor acquisition and consolidation. *npj Science of Learning*, **8**, 30.

Vehtari, A., Gelman, A. & Gabry, J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, **27**, 1413-1432.

Voormann, A. & Miller, J. (2024) Sequential effects on reaction time distributions: commonalities and differences across paradigms. *Journal of cognition*, **7**, 68.

Walker, M.P., Brakefield, T., Morgan, A., Hobson, J.A. & Stickgold, R. (2002) Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron*, **35**, 205-211.

Zatorre, R.J., Chen, J.L. & Penhune, V.B. (2007) When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci*, **8**, 547-558.