Musical Training Promotes the Interaction between Pitch and Temporal Dimensions in Auditory Working Memory

Authors: Zhou Linshu, Zhang Yuqing, Cai Danchao, Zhou Linshu, Cai Danchao

Date: 2025-06-16T12:33:13+00:00

Abstract

Music training can enhance individuals' sensitivity to temporal and non-temporal structures, but the joint influence of these structures in working memory remains unclear. This study investigates how pitch and rhythmic structures are processed in auditory working memory and the role of music training. The experiment manipulated pitch and rhythmic structures in melodies of varying lengths; musicians and non-musicians made same-different judgments based on changes in either the pitch or rhythmic dimension while suppressing interference from the other dimension. Results showed that in pitch retention tasks, non-musicians processed pitch and rhythmic structures independently, whereas musicians processed them interactively, and the interactive effect was positively correlated with music aptitude scores; in rhythm retention tasks, both groups processed structures independently, indicating that the effect of music training on structural integration is modulated by task type. Furthermore, the interactive effect was more pronounced in pitch retention tasks with shorter sequences, suggesting that this integration is further constrained by task properties and difficulty. These findings support dynamic attention theory, demonstrating that music training can enhance individuals' flexibility and adaptability in multidimensional information integration processing.

Full Text

Musical Training Enhances the Interaction Between Pitch and Time Dimensions in Auditory Working Memory

ZHOU Linshu[1]**, ZHANG Yuqing**[1]**, CAI Dan-Chao**[2]
([1] Music College, Shanghai Normal University, Shanghai 200234, China)

(^{2} Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China)

Abstract

Musical training enhances sensitivity to both temporal and non-temporal structures, yet the joint influence of these structures in working memory remains unclear. This study investigated how pitch and rhythmic structures are processed in auditory working memory and the role of musical training. We manipulated pitch and rhythmic structures in melodies of varying lengths, requiring musicians and nonmusicians to make same-different judgments based on either pitch or rhythm while suppressing interference from the other dimension. Results showed that in the pitch maintenance task, nonmusicians processed pitch and rhythmic structures independently, whereas musicians processed them interactively, with the interaction effect positively correlated with musical sophistication scores. In the rhythm maintenance task, both groups processed structures independently, indicating that musical training's effect on structural integration is modulated by task type. Moreover, the interaction effect was more pronounced in the pitch maintenance task with shorter sequences, suggesting that this integration is further constrained by task properties and difficulty. These findings support dynamic attending theory, demonstrating that musical training enhances flexibility and adaptability in multidimensional information integration.

Keywords: musical training, auditory working memory, temporal regularity, musical structure, dynamic attending theory

1. Introduction

A fundamental challenge for human cognitive systems lies in integrating discrete perceptual dimensions into unified representations. Music, as a quintessential multidimensional information carrier involving both pitch ("what") and rhythm ("when") hierarchies, provides an ideal model for exploring multidimensional information integration mechanisms (Krumhansl, 2000). Generally, musical pitch constructs a hierarchy of tonal stability through tonal rules (Krumhansl, 1990), while rhythm forms a temporal expectancy framework through metrical periodicity (Prince, Thompson, et al., 2009). This dual hierarchical structure creates organized sound patterns and offers a valuable opportunity to investigate human capacity for parallel processing of temporal and non-temporal information (Fitch, 2013; Prince, Thompson, et al., 2009). Previous research demonstrates that both structures significantly influence neural processing efficiency. For instance, in working memory tasks, pitch sequences conforming to tonal principles exhibit more pronounced chunking characteristics and are thus easier to store than atonal materials (Albouy et al., 2013; Bharucha & Krumhansl, 1983; Dowling, 1991; Lévêque et al., 2022; Schulze et al., 2012). Similarly, simple rhythms, compared to complex ones, possess more predictable temporal frame-

works that enhance reproduction accuracy and response speed (Essens & Povel, 1985; Martin et al., 2005; Sakai et al., 1999). However, these findings derive from dimension-segregated paradigms that essentially deconstruct music into single-dimensional auditory stimuli. This approach has limitations: it neglects the coupling between multidimensional structures in real musical cognition and cannot reveal whether pitch and rhythmic hierarchies are stored as independent modules or integrated through coordinated processing resources. Therefore, investigating the processing relationship between these two structures in working memory is crucial not only for understanding the nature of musical information storage but also for illuminating multidimensional information integration more broadly.

Theoretical frameworks for musical multidimensional processing primarily involve two hypotheses. Dynamic attending theory (Jones, 1976; Jones & Boltz, 1989) emphasizes the global dominance of temporal structure over cognitive resources: when rhythmic hierarchies establish stable periodic expectancies, attentional pulses synchronize with these oscillations, preferentially capturing pitch events that align with them within specific temporal windows. Based on this theory, musical working memory exhibits cross-dimensional gain effects—when pitch events coincide with rhythmic accents, working memory precision significantly improves (Large & Jones, 1999). In contrast, the dual-component model posits that pitch and rhythm processing are independent: pitch structure relies on rule-based symbolic encoding, whereas rhythm processing depends on temporal simulation and predictive mechanisms (Povel & Essens, 1985), with each relying on distinct neural pathways (Jerde et al., 2011; Schwartze & Kotz, 2013).

The core prediction of dynamic attending theory—that temporal structure globally modulates non-temporal processing—has received partial support under specific task conditions. For example, when listeners evaluate musical completeness holistically, non-isochronous rhythms weaken the influence of harmonic structure, but this cross-dimensional interference disappears during local chord judgments (Tillmann & Lebrun-Guillaud, 2006). This suggests that temporal dominance may be task-modulated: holistic processing promotes dimensional integration, whereas local processing relies on modular treatment. Similarly, Prince and Schmuckler et al. (2009) found that non-hierarchical temporal changes only interfere with pitch judgments in atonal contexts, becoming ineffective in tonal contexts, indicating that tonal structure may buffer interference from the temporal dimension—consistent with dynamic attending theory's assumption that "structural coordination enhances integration." However, the functional independence hypothesis of the dual-component model also receives empirical support. When participants judge tonal stability and rhythmic position separately, rhythmic hierarchies are influenced by tonal structure, yet tonal judgments remain independent of rhythmic context (Prince, Thompson, et al., 2009). This demonstrates that even when dimensional influences exist, their direction and strength are constrained by task demands. Furthermore, research on musical syntax processing shows that early detection of pitch and

rhythmic structures proceeds independently, with interactive effects emerging only at later integration stages, suggesting that dimension-specific processing precedes strategic integration (Sun et al., 2020; Zhang et al., 2019).

Despite these advances, knowledge gaps remain: most experimental tasks have focused on online perceptual judgments, leaving unclear how musical multidimensional information is processed during working memory storage. Moreover, the synergistic effects of structural hierarchy and task type have not been systematically examined. Is the interaction between pitch and rhythm a stable property of musical cognition or merely an epiphenomenon emerging from specific task contexts? This question can be directly tested through behavioral dissociation in selective attention paradigms. If the dual-component model holds, structural changes in task-irrelevant dimensions should not affect target dimension processing performance. Conversely, if dynamic attending theory dominates, the hierarchical nature of temporal structure should influence pitch processing efficiency across tasks.

The modulatory effect of musical training on multidimensional structure processing further complicates this issue. Although general listeners acquire basic musical structure processing abilities through long-term exposure to Western tonal environments (Koelsch et al., 2000; Koelsch et al., 2007), professional training significantly enhances this cognitive advantage. Musicians exhibit improved tonal structure processing in working memory tasks (Schulze et al., 2011) and more accurately synchronize with complex rhythms (Chen et al., 2008). They also demonstrate syntactic processing advantages for both pitch (Jentschke & Koelsch, 2009; Koelsch et al., 2002) and rhythmic structures (Sun et al., 2018). However, evaluations of musical training effects typically rely on deconstructed single dimensions (pitch or rhythm), whereas real musical experience inevitably involves real-time binding of multidimensional structures. Notably, when required to process pitch and rhythmic structures simultaneously, musicians' syntactic integration advantage disappears (Sun et al., 2020). This raises a critical question: does musical training confer processing advantages that manifest solely as dimension-specific ability enhancement, or does it also involve improved cross-dimensional integration capacity?

Based on these considerations, this study aimed to investigate the interactive mechanisms of pitch and rhythmic structures in musical working memory and the moderating role of musical training. We recruited musicians and nonmusicians to complete two independent experiments examining their maintenance of pitch and rhythm information. Experiment 1 focused on pitch maintenance in auditory working memory, requiring participants to judge pitch sameness while suppressing rhythmic variations. Experiment 2 reversed the task demands, examining rhythm maintenance in auditory working memory. Both experiments manipulated pitch structure (tonal vs. atonal) and rhythmic structure (simple vs. complex) to explore whether their processing is independent or interactive. If changes in pitch structure affect rhythmic structure effects, or vice versa (i.e., an interaction emerges), this would indicate that the two structures are

integrated during processing, thereby achieving structural gain. Conversely, if pitch and rhythmic structures produce independent main effects without interaction, this would suggest that under current task conditions, their processing is relatively independent (Prince, 2011; Prince & Pfordresher, 2012; Tillmann & Lebrun-Guillaud, 2006).

Based on previous research, we hypothesized that both groups would show superior working memory for structured information (tonal, simple rhythm) compared to unstructured conditions (atonal, complex rhythm), with musicians showing a larger advantage magnitude (Schulze et al., 2011). Additionally, considering that musical training enhances sensitivity to hierarchical pitch and rhythmic structures (Koelsch et al., 2002; Sun et al., 2018), we expected musicians to process pitch and rhythmic structures more interactively due to their enhanced information integration skills (Zhou et al., 2017). Finally, we predicted that result patterns would differ between the two experiments, reflecting task-specific cognitive processes' differential impact on musical structure integration.

2. Experiment 1: Pitch Maintenance

Experiment 1 examined how musicians and nonmusicians maintain pitch information through a same-different pitch judgment task. Working memory capacity limits the amount of material that can be retained: performance declines as sequences become longer (or contain more events) (Cowan, 2000). Given that previous studies have primarily used five- to seven-note sequences to investigate musical working memory (e.g., Schulze et al., 2012), this study selected five- and seven-note sequences to examine sequence length effects. Therefore, the experiment simultaneously manipulated melodic stimuli across three factors: pitch structure (tonal vs. atonal), rhythmic structure (simple vs. complex), and sequence length (five-note vs. seven-note). This design aimed to understand how musical training influences pitch maintenance and how different musical structures and sequence lengths affect performance in both groups.

2.1 Method

2.1.1 Participants We calculated the required sample size using G*Power 3.1 (Faul et al., 2009) based on effect sizes from previous research examining musical training and musical structure processing (Chen et al., 2008). To detect a large effect (f = 0.739) at a significance level of 0.05 with 80% statistical power, a minimum of 17 participants per group was needed. To ensure adequate statistical power and account for individual differences and experimental design requirements, we ultimately recruited 36 musicians and 36 nonmusicians to enhance result robustness and generalizability. Musicians had an average of 13.92 years (SD = 2.97) of instrumental training, practicing an average of 4.04 hours per day (SD = 0.91), with primary instruments including piano, violin, clarinet, and guitar. Nonmusicians had no musical training beyond school music classes. Before the formal experiment, all participants completed a nonverbal

intelligence test (Raven's Standard Progressive Matrices, SPM; Raven, 2000) and the Goldsmiths Musical Sophistication Index (Gold-MSI; Lin et al., 2021; Müllensiefen et al., 2014). As shown in Table 1 , musicians and nonmusicians did not differ significantly in age, gender, handedness, years of education, or nonverbal intelligence, but musicians scored significantly higher on the musical sophistication index. All experimental procedures were approved by the Academic Ethics and Morality Committee of Shanghai Normal University. All participants signed informed consent forms before the experiment and received compensation upon completion.

2.1.2 Stimuli Experimental stimuli consisted of monophonic melodies created for this study, each lasting four beats. To ensure diversity in melody length, we initially composed 20 melodies, including 10 five-note sequences and 10 seven-note sequences. We manipulated pitch and rhythmic structures based on previous research. Pitch structure was divided into tonal and atonal levels (e.g., Schulze et al., 2011; Schulze et al., 2012), while rhythmic structure was divided into simple and complex levels (e.g., Chen et al., 2008). Figure 1 [FIGURE:1] shows examples of these melodic stimuli.

Specifically, tonal melodies used notes from the C major scale (C4: 261.63 Hz, D4: 293.66 Hz, E4: 329.63 Hz, F4: 349.23 Hz, G4: 392 Hz, A4: 440 Hz, B4: 493.88 Hz, C5: 523 Hz), beginning with notes from the C major tonic chord and ending on the tonic (C) or dominant (G) to create a strong sense of tonality. Atonal melodies were based on the whole-tone scale starting from C (C4: 261.63 Hz, D4: 293.66 Hz, E4: 329.63 Hz, F#4: 369.99 Hz, G#4: 415.305 Hz, A#4: 466.164 Hz, C5: 523 Hz), which lacks a tonal center and cannot form traditional major or minor triads, making it difficult to generate a sense of tonality. We used Krumhansl and Schmuckler's key-finding algorithm (see Krumhansl, 1990) to analyze the strength of tonal centers established by melodic contexts. This algorithm calculates the Maximum Key-Profile Correlation (MKC)—the correlation coefficient with the best-matching tonal hierarchy—indicating the strongest degree of tonal establishment. We computed correlations between all melodic sequences in this study and the key profiles of 12 major and 12 minor keys. A two-way ANOVA with pitch structure (tonal vs. atonal) and sequence length (five-note vs. seven-note) as independent variables and MKC as the dependent variable revealed a significant main effect of pitch structure (0.76 vs. 0.48, $F(1,76) = 121.74$, $p < 0.001$, $\eta^2p = 0.61$), indicating higher MKC for tonal than atonal stimuli. Neither the main effect of sequence length nor the interaction between pitch structure and sequence length was significant (ps > 0.109). Additionally, we conducted acoustic analyses using MIRtoolbox (Lartillot et al., 2008), computing two tonality-related parameters: Chromagram Centroid and Key Clarity. Two-way ANOVA results showed significant main effects of pitch structure for both Chromagram Centroid (0.24 vs. 0.11, $F(1,76) = 9.32$, $p = 0.003$, $\eta^2p = 0.11$) and Key Clarity (0.74 vs. 0.45, $F(1,76) = 146.45$, $p < 0.001$, $\eta^2p = 0.66$), with significant differences between tonal and atonal conditions. Neither the main effect of sequence length nor the interaction was

significant (ps > 0.083). These results validated the differences in tonal characteristics between tonal and atonal melodies.

Simple rhythmic sequences consisted of basic note values (e.g., half notes, quarter notes, dotted quarter notes) without syncopation, ensuring stable meter. Complex rhythmic sequences contained syncopation, with accents falling on weak or off beats, creating rhythmic conflict and unstable temporal structure. According to Povel and Essens (1985), listeners identify an optimal internal clock interval that best aligns with a given temporal pattern. The appropriateness of this clock is determined by the C index, which quantifies the degree of counterevidence for maintaining a single clock interval. The C index is calculated as: $C = (W \times -ev) + (1 \times 0ev)$, where 0ev represents the number of clock ticks coinciding with non-accented events, -ev reflects clock ticks falling on silent positions (rests), and W is a weight constant (typically $W = 4$; see Povel & Essens, 1985). Lower C values indicate less counterevidence and thus stronger temporal regularity. In this study, we computed C values for all simple and complex rhythmic sequences using 200 ms as the basic temporal unit (the shortest interval in our materials), representing rhythmic patterns as multiples of 1, 2, 3, or 4 units. We then identified the clock with the lowest C value for each sequence as its optimal internal clock, using this as an index of relative beat strength. Results indicated that a 4-unit clock was optimal for both simple and complex rhythmic sequences. A two-way ANOVA on minimum C values revealed a significant main effect of rhythmic structure, with lower C values for simple than complex rhythms (2.45 vs. 4.05, $F(1,76) = 10.89$, $p = 0.001$, $\eta^2p = 0.13$); a significant main effect of sequence length, with lower C values for seven-note than five-note sequences (2.40 vs. 4.10, $F(1,76) = 12.30$, $p < 0.001$, $\eta^2p = 0.14$); and a non-significant interaction between beat type and sequence length ($p = 0.681$). These results confirmed that simple rhythmic sequences possessed stronger temporal regularity than complex ones.

Based on the four combinations of pitch and rhythmic structures, we generated 80 melodic stimuli: 20 tonal melodies with simple rhythm, 20 tonal melodies with complex rhythm, 20 atonal melodies with simple rhythm, and 20 atonal melodies with complex rhythm. Each group contained 10 five-note and 10 seven-note sequences, totaling 40 five-note and 40 seven-note melodies. The study thus comprised eight experimental conditions (pitch structure × rhythmic structure × sequence length), with 10 melodic stimuli per condition.

Each melodic stimulus lasted 3200 ms at a tempo of 75 beats per minute. We imported the created MIDI files into Cubase 5.1, used the YAMAHA S90ES grand piano timbre, and exported them as WAV files through Adobe Audition CS6. All stimuli were monophonic, with a sampling rate of 44.1 kHz and 16-bit depth, and were normalized to approximately 68 dB SPL using Adobe Audition CS6.

2.1.3 Procedure　This experiment employed a pitch recognition paradigm in which each melodic stimulus was presented twice per trial. Thus, based on the

80 original melodies, we created 160 trials, each comprising two sequentially presented melodies. In half of the trials (80), the second melody changed in pitch while rhythm remained unchanged; in the other half (80), the second melody changed in rhythm while pitch remained unchanged. Specifically, in pitch-change trials, one pitch in the melody was raised or lowered by a semitone or whole tone, with equal probability for direction, ensuring that pitch alterations did not change the tonal structure or original melodic contour. In rhythm-change trials, the order of note durations on the second or third beat was swapped, or a note duration was split into smaller units and recombined, ensuring that rhythmic adjustments did not alter the number of notes or the original rhythmic structure (simple or complex). To prevent ceiling effects, pitch or rhythm changes never occurred at the beginning or end of melodies but appeared randomly in the middle. Consequently, each experimental condition contained 20 trials, half with "same" pairs and half with "different" pairs. Additionally, to prevent participants from developing specific response strategies, we included 32 filler trials in which the two melodies were identical. Data from filler trials were excluded from statistical analysis.

The experiment was conducted in a quiet room, with stimuli presented via computer and played through headphones. Each trial began with the first melody (3200 ms), followed by a 2000 ms retention interval, then the second melody (3200 ms). After melody presentation, a response screen appeared. Participants were instructed to quickly and accurately judge whether the pitches of the two consecutively presented melodies were identical while ignoring rhythmic information. Response key assignments were balanced between participants' left and right hands. After each response, a 500 ms blank screen preceded the next trial.

Before the formal experiment, participants completed four practice trials with feedback to familiarize themselves with the procedure and task. The formal experiment comprised four blocks of 40 trials each. Stimulus presentation order was pseudorandomized, ensuring that "same" or "different" pairs did not appear consecutively more than three times, and that the first melody of any pair did not repeat consecutively across trials.

2.1.4 Statistical Analysis Based on signal detection theory, we computed detection sensitivity [d' = z(hit rate) − z(false alarm rate)] and response bias [c = −0.5 × (z(hit rate) + z(false alarm rate))] using standard formulas (Macmillan & Creelman, 2004). The d' value reflects participants' ability to discriminate between "same" and "different" trials, with higher values indicating greater sensitivity. The c value reflects response bias, with positive c values indicating a tendency to respond "different" and negative values indicating a tendency to respond "same." Hit rate was defined as the proportion of correct judgments on "different" trials, and false alarm rate as the proportion of incorrect judgments on "same" trials. For cases with no false alarms, corrected values of 0.05 were used for d' and c calculations, and for cases with perfect hit rates, cor-

---

rected values of 0.95 were used. Additionally, to examine decision speed in the memory task, we computed mean reaction times for correct responses in each condition, with reaction time analyses based only on correct trials. To investigate experimental condition effects, we conducted four-way repeated-measures ANOVAs with group as a between-subjects factor and pitch structure, rhythmic structure, and sequence length as within-subjects factors. We analyzed d', c, and reaction times separately to explore main effects and interactions of these factors on performance.

2.2 Results

2.2.1 Detection Sensitivity The ANOVA on detection sensitivity revealed a significant main effect of group, $F(1, 70) = 49.56$, $p < 0.001$, $\eta^2 p = 0.42$, with musicians showing higher detection sensitivity than nonmusicians (1.93 vs. 0.94, 95% CI = [0.71, 1.27]). The main effect of pitch structure was significant, $F(1, 70) = 64.03$, $p < 0.001$, $\eta^2 p = 0.48$, with higher sensitivity for tonal than atonal melodies (1.64 vs. 1.24, 95% CI = [0.29, 0.50]). The main effect of rhythmic structure was significant, $F(1, 70) = 24.36$, $p < 0.001$, $\eta^2 p = 0.26$, with higher sensitivity for simple than complex rhythms (1.59 vs. 1.28, 95% CI = [0.19, 0.43]). The main effect of sequence length was significant, $F(1, 70) = 10.70$, $p = 0.002$, $\eta^2 p = 0.13$, with higher sensitivity for five-note than seven-note sequences (1.53 vs. 1.34, 95% CI = [0.07, 0.30]). Additionally, the interaction between group and sequence length was significant, $F(1, 70) = 4.17$, $p = 0.045$, $\eta^2 p = 0.06$. Simple effects analysis indicated that musicians showed higher detection sensitivity for five-note than seven-note sequences (2.08 vs. 1.78, $t(35) = 4.06$, $p < 0.001$, Cohen's d = 0.68, 95% CI = [0.15, 0.45]), whereas nonmusicians' sensitivity was unaffected by sequence length (0.98 vs. 0.91, $t(35) = 0.81$, $p = 0.422$). Moreover, the three-way interaction between group, pitch structure, and rhythmic structure was marginally significant, $F(1, 70) = 3.97$, $p = 0.050$, $\eta^2 p = 0.05$, and the three-way interaction between pitch structure, rhythmic structure, and sequence length was significant, $F(1, 70) = 8.57$, $p = 0.005$, $\eta^2 p = 0.11$. These results confirm musicians' advantage in pitch detection sensitivity and demonstrate that sensitivity is influenced by pitch structure, rhythmic structure, and sequence length, with higher sensitivity observed for tonal melodies, simple rhythms, and shorter sequences.

To further explore the marginally significant interaction between group, pitch structure, and rhythmic structure, we conducted separate two-way ANOVAs for musicians and nonmusicians. For musicians, significant main effects emerged for pitch structure, $F(1, 35) = 54.50$, $p < 0.001$, $\eta^2 p = 0.61$, with higher sensitivity for tonal than atonal melodies (2.21 vs. 1.65, 95% CI = [0.40, 0.70]), and for rhythmic structure, $F(1, 35) = 24.32$, $p < 0.001$, $\eta^2 p = 0.41$, with higher sensitivity for simple than complex rhythms (2.09 vs. 1.77, 95% CI = [0.19, 0.46]). Crucially, the interaction between pitch and rhythmic structure was significant, $F(1, 35) = 6.32$, $p = 0.017$, $\eta^2 p = 0.15$. As shown in Figure 2

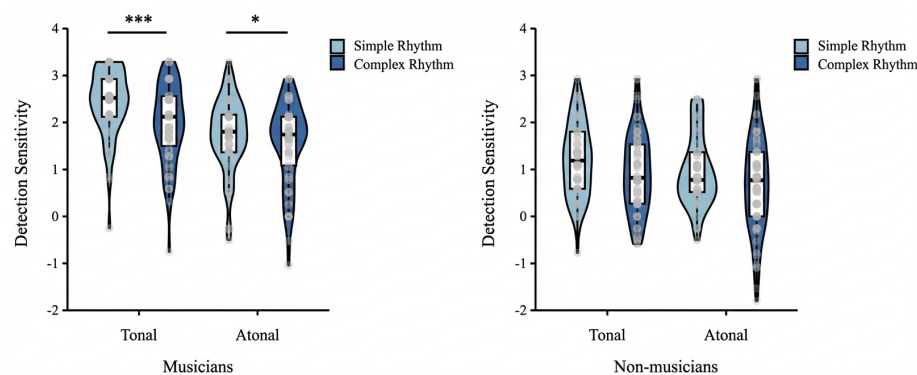, simple effects analysis revealed that musicians showed higher sensitivity for

Figure 1: Figure 2

simple than complex structures in both tonal and atonal conditions, but the effect was larger in the tonal condition (2.43 vs. 1.98, t(35) = 5.36, p < 0.001, Cohen's d = 0.89, 95% CI = [0.28, 0.63]) than in the atonal condition (1.75 vs. 1.56, t(35) = 2.30, p = 0.027, Cohen's d = 0.38, 95% CI = [0.02, 0.36]). Similarly, musicians showed higher sensitivity for tonal than atonal melodies in both simple and complex rhythmic conditions, but the effect was larger in the simple rhythm condition (2.43 vs. 1.75, t(35) = 8.64, p < 0.001, Cohen's d = 1.44, 95% CI = [0.52, 0.85]) than in the complex rhythm condition (1.98 vs. 1.56, t(35) = 4.11, p < 0.001, Cohen's d = 0.69, 95% CI = [0.21, 0.63]).

For nonmusicians, significant main effects emerged for pitch structure, F(1, 35) = 13.67, p < 0.001, ^{2}p = 0.28, with higher sensitivity for tonal than atonal melodies (1.06 vs. 0.82, 95% CI = [0.11, 0.37]), and for rhythmic structure, F(1, 35) = 7.62, p = 0.009, ^{2}p = 0.18, with higher sensitivity for simple than complex rhythms (1.09 vs. 0.80, 95% CI = [0.08, 0.51]). However, the interaction between pitch and rhythmic structure was not significant (p = 0.893). Thus, only musicians exhibited a significant interaction between pitch and rhythmic structures in detection sensitivity, suggesting that musicians achieve structural gain through integration of tonal and simple rhythmic information.

To further confirm the role of musical training experience, we computed the difference in rhythmic structure effects between tonal and atonal conditions as an index of pitch-rhythm structural interaction and correlated this with participants' Gold-MSI musical sophistication scores. As shown in Figure 3

A, a significant positive correlation emerged, Pearson's r(70) = 0.28, p = 0.019, Fisher's z = 0.28, 95% CI = [0.05, 0.48]. This indicates that participants with higher musical sophistication showed more pronounced interactive effects between pitch and rhythmic structures, reflecting the association between musical training and pitch-rhythm structural integration.

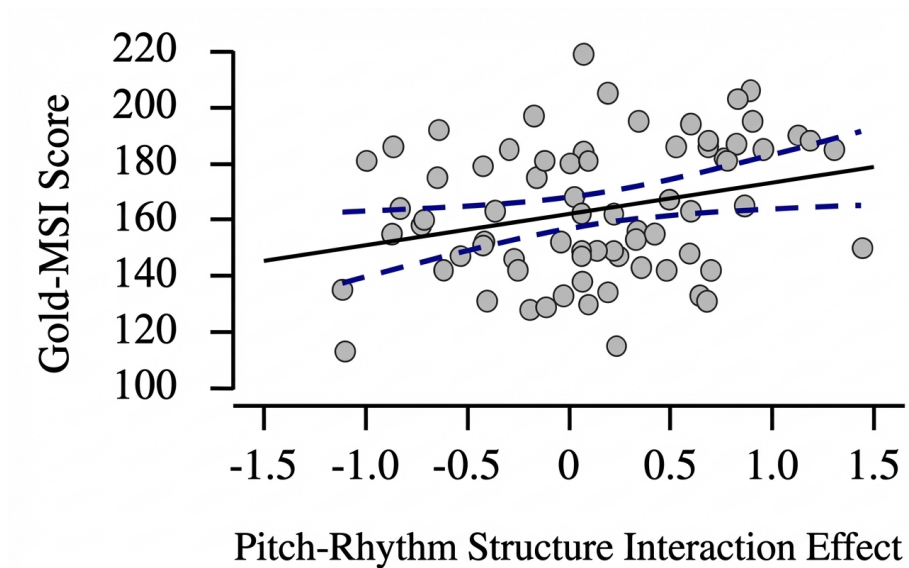To explore the interaction between pitch structure, rhythmic structure, and

Figure 2: Figure 3

sequence length, we conducted separate two-way ANOVAs for five-note and seven-note sequences. As shown in Figure 4
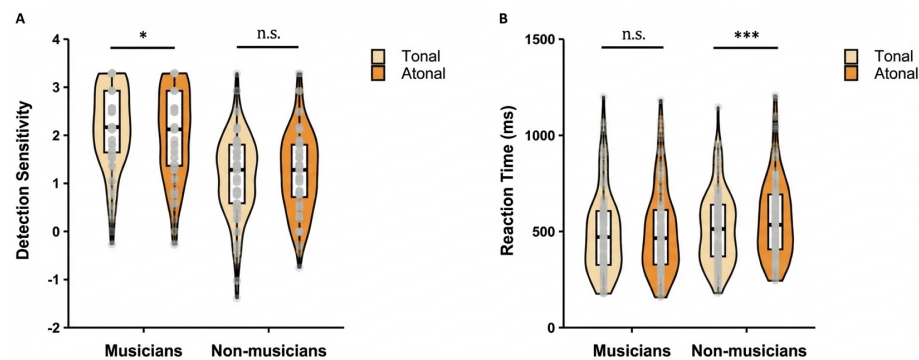

Figure 3: Figure 4

, for five-note sequences, a significant main effect of pitch structure emerged, $F(1, 71) = 122.66$, $p < 0.001$, $\eta^2p = 0.63$, with higher detection sensitivity for tonal than atonal melodies (1.79 vs. 1.27, 95% CI = [0.43, 0.62]), and a significant interaction between pitch and rhythmic structures, $F(1, 71) = 13.03$, $p < 0.001$, $\eta^2p = 0.15$. Simple effects analysis showed higher sensitivity for simple than complex rhythms in the tonal condition (1.91 vs. 1.67, $t(71) = 3.35$, $p = 0.001$, Cohen's d = 0.40, 95% CI = [0.10, 0.39]), but no effect in the atonal

condition (1.22 vs. 1.31, $t(71) = -0.86$, $p = 0.392$). For seven-note sequences, a significant main effect of pitch structure emerged, $F(1, 71) = 10.38$, $p = 0.002$, $\eta^2p = 0.13$, with higher sensitivity for tonal than atonal melodies (1.48 vs. 1.21, 95% CI = [0.10, 0.43]), and a significant main effect of rhythmic structure, $F(1, 71) = 44.51$, $p < 0.001$, $\eta^2p = 0.39$, with higher sensitivity for simple than complex rhythms (1.61 vs. 1.07, 95% CI = [0.38, 0.70]). However, the interaction between pitch and rhythmic structures was not significant, $F(1, 71) = 0.63$, $p = 0.429$. These results indicate that the interaction between pitch and rhythmic structures is modulated by sequence length, with integration of tonal and simple rhythms facilitating pitch detection sensitivity only under shorter sequence conditions.

2.2.2 Response Bias The ANOVA on response bias revealed a significant main effect of pitch structure, $F(1, 70) = 80.09$, $p < 0.001$, $\eta^2p = 0.53$, indicating greater response bias for tonal than atonal melodies (0.23 vs. 0.06, 95% CI = [0.13, 0.21]). The main effect of rhythmic structure was significant, $F(1, 70) = 122.49$, $p < 0.001$, $\eta^2p = 0.64$, indicating greater response bias for complex than simple rhythms (0.29 vs. 0.004, 95% CI = [0.24, 0.34]). Additionally, the main effect of sequence length was marginally significant, $F(1, 70) = 3.90$, $p = 0.052$, $\eta^2p = 0.05$, with seven-note sequences tending to elicit greater response bias than five-note sequences (0.17 vs. 0.12). The interaction between sequence length and group was significant, $F(1, 70) = 4.56$, $p = 0.036$, $\eta^2p = 0.06$. Simple effects analysis indicated that musicians showed greater response bias for seven-note than five-note sequences (0.17 vs. 0.07, $p = 0.005$, 95% CI = [0.007, 0.20]), whereas nonmusicians showed no such effect (0.18 vs. 0.18, $p = 0.911$). Furthermore, the three-way interaction between pitch structure, rhythmic structure, and sequence length was significant, $F(1, 70) = 5.12$, $p = 0.027$, $\eta^2p = 0.07$. No other main effects or interactions were significant (ps > 0.157). These results indicate that tonal melodies and complex rhythms elicited greater response bias, and only musicians showed increased response bias for longer sequences.

To further explore the interaction between pitch structure, rhythmic structure, and sequence length, we analyzed five-note and seven-note sequences separately. For five-note sequences, significant main effects emerged for pitch structure, $F(1, 71) = 114.89$, $p < 0.001$, $\eta^2p = 0.62$, with greater response bias for tonal than atonal melodies (0.25 vs. –0.01, 95% CI = [0.19, 0.33]), and for rhythmic structure, $F(1, 71) = 75.85$, $p < 0.001$, $\eta^2p = 0.52$, with greater response bias for complex than simple rhythms (0.26 vs. –0.01, 95% CI = [0.18, 0.37]). The interaction was not significant, $F(1, 71) = 2.88$, $p = 0.094$, $\eta^2p = 0.04$. For seven-note sequences, significant main effects emerged for pitch structure, $F(1, 71) = 7.04$, $p = 0.010$, $\eta^2p = 0.09$, with greater response bias for tonal than atonal melodies (0.22 vs. 0.13, 95% CI = [0.009, 0.16]), and for rhythmic structure, $F(1, 71) = 65.63$, $p < 0.001$, $\eta^2p = 0.48$, with greater response bias for complex than simple rhythms (0.33 vs. 0.02, 95% CI = [0.21, 0.40]). The interaction was not significant, $F(1, 71) = 2.35$, $p = 0.130$, $\eta^2p = 0.03$. Thus, for both five-note and seven-note sequences, pitch and rhythmic structures independently influenced

response bias in pitch detection.

2.2.3 Reaction Time   The ANOVA on reaction times revealed a significant main effect of group, $F(1, 70) = 15.50$, $p < 0.001$, $\eta^2 p = 0.18$, with musicians responding faster than nonmusicians (430 ms vs. 583 ms, 95% CI = [–233.10, –72.79]). The main effect of pitch structure was marginally significant, $F(1, 70) = 3.99$, $p = 0.050$, $\eta^2 p = 0.05$, with a tendency for faster responses to tonal than atonal melodies (498 ms vs. 514 ms, 95% CI = [–32.30, –0.03]). The main effect of rhythmic structure was not significant ($p = 0.068$). Additionally, the three-way interaction between group, rhythmic structure, and sequence length was significant, $F(1, 70) = 4.47$, $p = 0.038$, $\eta^2 p = 0.06$. To explore this interaction, we analyzed simple and complex rhythm conditions separately. For simple rhythms, a significant main effect of group emerged, $F(1, 70) = 14.78$, $p < 0.001$, $\eta^2 p = 0.17$, with musicians responding faster than nonmusicians (425 ms vs. 572 ms, 95% CI = [–226.22, –67.86]), and a significant interaction between group and sequence length, $F(1, 70) = 4.14$, $p = 0.046$, $\eta^2 p = 0.06$. Simple effects analysis showed that musicians responded faster to five-note than seven-note sequences (410 ms vs. 440 ms, $t(35) = –2.56$, $p = 0.015$, Cohen's d = –0.43, 95% CI = [–53.63, –6.43]), whereas nonmusicians showed no such effect (586 ms vs. 558 ms, $t(35) = 1.05$, $p = 0.299$). For complex rhythms, a significant main effect of group emerged, $F(1, 70) = 15.28$, $p < 0.001$, $\eta^2 p = 0.18$, with musicians responding faster than nonmusicians (435 ms vs. 593 ms, 95% CI = [–243.62, –74.06]), but neither the main effect of sequence length nor the group × sequence length interaction was significant (ps > 0.771). These results indicate that musicians responded significantly faster than nonmusicians, particularly under simple rhythm and shorter sequence conditions.

2.3 Discussion of Experiment 1

Experiment 1 results indicate that pitch and rhythmic structures influenced various behavioral indices of nonmusicians' pitch maintenance, but without interaction—pitch structure did not affect rhythmic structure effects, and vice versa—suggesting that nonmusicians processed the two structures relatively independently in pitch working memory. For musicians, pitch and rhythmic structures influenced pitch maintenance performance, and the effects of these structures interacted to impact detection sensitivity, indicating that musicians integrated pitch and rhythmic structures in pitch working memory. Moreover, this interactive effect correlated positively with participants' musical sophistication scores, suggesting that individuals with higher musical sophistication showed more pronounced interactive effects when processing pitch and rhythmic structures, further validating the role of musical training in pitch-rhythm structural interaction. Additionally, the interaction was more evident in five-note than seven-note sequences, suggesting that pitch-rhythm structural interaction may be modulated by task difficulty.

3. Experiment 2: Rhythm Maintenance

Experiment 2 examined musicians' and nonmusicians' working memory for rhythm through a rhythm recognition task. Similar to Experiment 1, we manipulated pitch structure, rhythmic structure, and sequence length.

3.1 Method

3.1.1 Participants The same participants as in Experiment 1.

3.1.2 Stimuli, Procedure, and Statistical Analysis Stimuli, procedure, and statistical analyses were identical to Experiment 1. The only difference was the task: this experiment examined rhythm maintenance, requiring participants to quickly and accurately judge whether the rhythms of two consecutively presented melodies were identical while ignoring pitch information. To control for order effects, experimental order was balanced across both groups.

3.2 Results

3.2.1 Detection Sensitivity The ANOVA on detection sensitivity revealed a significant main effect of group, $F(1, 70) = 41.99$, $p < 0.001$, $\eta^2 p = 0.38$, with musicians showing higher sensitivity than nonmusicians (2.08 vs. 1.18, 95% CI = [0.62, 1.17]). The main effect of rhythmic structure was significant, $F(1, 70) = 180.77$, $p < 0.001$, $\eta^2 p = 0.72$, with higher sensitivity for simple than complex rhythms (1.94 vs. 1.33, 95% CI = [0.51, 0.70]). The main effect of sequence length was significant, $F(1, 70) = 31.85$, $p < 0.001$, $\eta^2 p = 0.31$, with higher sensitivity for five-note than seven-note sequences (1.79 vs. 1.48, 95% CI = [0.20, 0.42]). The interaction between group and pitch structure was significant, $F(1, 70) = 6.98$, $p = 0.010$, $\eta^2 p = 0.09$. Simple effects analysis indicated that musicians showed higher sensitivity in the tonal than atonal condition (2.15 vs. 2.01, $t(35) = 2.18$, $p = 0.036$, Cohen's d = 0.36, 95% CI = [0.01, 0.28]), whereas nonmusicians showed no such effect (1.15 vs. 1.22, $t(35) = -1.50$, $p = 0.142$) (see Figure 5 [FIGURE:5]A). Additionally, the interaction between pitch structure and sequence length was significant, $F(1, 70) = 7.18$, $p = 0.009$, $\eta^2 p = 0.09$, indicating that pitch structure influenced performance in the seven-note condition (1.55 vs. 1.41, $t(71) = 2.29$, $p = 0.025$, Cohen's d = 0.27, 95% CI = [0.02, 0.27]) but not in the five-note condition (1.75 vs. 1.82, $t(71) = -1.33$, $p = 0.188$). Furthermore, a significant three-way interaction emerged between group, rhythmic structure, and sequence length, $F(1, 70) = 11.14$, $p = 0.001$, $\eta^2 p = 0.14$. These results demonstrate musicians' advantage in rhythm detection sensitivity; rhythmic structure and sequence length significantly affect rhythm detection sensitivity, while pitch structure effects are limited to musicians or longer sequences.

To further understand the interaction between group, rhythmic structure, and sequence length, we conducted separate two-way ANOVAs for musicians and nonmusicians. For musicians, significant main effects emerged for rhythmic

structure, $F(1, 35) = 241.33$, $p < 0.001$, $\eta^2_p = 0.87$, indicating higher sensitivity for simple than complex rhythms (2.44 vs. 1.71, 95% CI = [0.64, 0.83]), and for sequence length, $F(1, 35) = 21.31$, $p < 0.001$, $\eta^2_p = 0.38$, indicating higher sensitivity for five-note than seven-note sequences (2.28 vs. 1.88, 95% CI = [0.22, 0.57]). The interaction between rhythmic structure and sequence length was significant, $F(1, 35) = 6.37$, $p = 0.016$, $\eta^2_p = 0.15$, indicating that in the complex rhythm condition, sensitivity was higher for five-note than seven-note sequences (2.01 vs. 1.42, $t(35) = 5.07$, $p < 0.001$, Cohen's d = 0.85, 95% CI = [0.35, 0.82]), but no such effect appeared in the simple rhythm condition (2.54 vs. 2.34, $t(35) = 1.75$, $p = 0.089$). For nonmusicians, significant main effects emerged for rhythmic structure, $F(1, 35) = 39.48$, $p < 0.001$, $\eta^2_p = 0.53$, indicating higher sensitivity for simple than complex rhythms (1.43 vs. 0.94, 95% CI = [0.33, 0.64]), and for sequence length, $F(1, 35) = 10.66$, $p < 0.001$, $\eta^2_p = 0.23$, indicating higher sensitivity for five-note than seven-note sequences (1.30 vs. 1.07, 95% CI = [0.09, 0.36]). The interaction between rhythmic structure and sequence length was significant, $F(1, 35) = 4.77$, $p = 0.036$, $\eta^2_p = 0.12$, indicating that in the simple rhythm condition, sensitivity was higher for five-note than seven-note sequences (1.61 vs. 1.24, $t(35) = 4.04$, $p < 0.001$, Cohen's d = 0.67, 95% CI = [0.18, 0.55]), but no such effect appeared in the complex rhythm condition (0.98 vs. 0.90, $t(35) = 0.81$, $p = 0.423$). Thus, musicians' rhythm detection sensitivity was influenced by sequence length only in the complex rhythm condition, whereas nonmusicians showed this influence only in the simple rhythm condition.

3.2.2 Response Bias The ANOVA on response bias revealed a significant main effect of group, $F(1, 70) = 27.34$, $p < 0.001$, $\eta^2_p = 0.28$, indicating greater response bias for nonmusicians than musicians (0.54 vs. 0.26, 95% CI = [0.17, 0.39]), with nonmusicians more likely to judge trials as "different." The main effect of rhythmic structure was significant, $F(1, 70) = 24.41$, $p < 0.001$, $\eta^2_p = 0.26$, with greater response bias for complex than simple rhythms (0.47 vs. 0.33, 95% CI = [0.08, 0.20]). Additionally, the main effect of sequence length was significant, $F(1, 70) = 13.03$, $p < 0.001$, $\eta^2_p = 0.16$, with greater response bias for seven-note than five-note sequences (0.45 vs. 0.35, 95% CI = [0.04, 0.15]). No other main effects or interactions were significant (ps > 0.067). These results indicate that nonmusicians, complex rhythms, and longer sequences all elicited greater response bias.

3.2.3 Reaction Time The ANOVA on reaction times revealed a significant main effect of pitch structure, $F(1, 70) = 15.50$, $p < 0.001$, $\eta^2_p = 0.18$, with faster responses to tonal than atonal melodies (506 ms vs. 532 ms, 95% CI = [–42.37, –10.87]). The main effect of rhythmic structure was significant, $F(1, 70) = 56.90$, $p < 0.001$, $\eta^2_p = 0.45$, with faster responses to simple than complex rhythms (485 ms vs. 553 ms, 95% CI = [–86.39, –50.47]). The interaction between group and pitch structure was significant, $F(1, 70) = 6.90$, $p = 0.011$, $\eta^2_p = 0.09$, indicating that nonmusicians responded faster to tonal than atonal

melodies (519 ms vs. 565 ms, t(35) = –4.01, p < 0.001, Cohen's d = –0.67, 95% CI = [–70.14, –22.96]), whereas musicians showed no difference between tonal and atonal melodies (493 ms vs. 500 ms, t(35) = –0.69, p = 0.498) (see Figure 5 [FIGURE:5]B). The interaction between pitch structure and sequence length was significant, F(1, 70) = 4.65, p = 0.034, η^2p = 0.06, indicating that tonal structure facilitated reaction times for seven-note sequences (495 ms vs. 536 ms, t(71) = –3.58, p < 0.001, Cohen's d = 0.42, 95% CI = [–65.12, –18.50]) but not for five-note sequences (517 ms vs. 529 ms, t(71) = –1.23, p = 0.221). Additionally, the interaction between rhythmic structure and sequence length was significant, F(1, 70) = 9.92, p = 0.002, η^2p = 0.12, with a larger rhythmic structure effect for five-note sequences (478 ms vs. 568 ms, t(71) = –7.96, p < 0.001, Cohen's d = –0.94, 95% CI = [–112.48, –67.42]) than for seven-note sequences (492 ms vs. 539 ms, t(71) = –4.15, p < 0.001, Cohen's d = –0.49, 95% CI = [–69.48, –24.36]). These results demonstrate that tonal melodies and simple rhythms accelerated rhythm judgment responses, although atonal melodies delayed nonmusicians' reaction times.

3.3 Discussion of Experiment 2

The main results of Experiment 2 indicate that neither musicians nor nonmusicians showed significant interactions between pitch and rhythmic structures during the rhythm maintenance task. This contrasts with Experiment 1, suggesting that pitch and rhythmic structure effects are relatively independent during rhythm maintenance. Compared to musicians, nonmusicians did not show sensitivity differences between tonal and atonal conditions, although tonal structure still provided some cognitive support for rhythm maintenance, benefiting nonmusicians' response speed. Furthermore, sequence length influenced musicians' sensitivity to complex rhythmic sequences but did not affect their judgments of simple rhythms. For nonmusicians, sequence length only influenced their sensitivity to simple rhythms. These results suggest that sequence length effects primarily manifest in moderately difficult rhythm recognition tasks. When tasks are too easy (e.g., musicians recognizing simple rhythms) or too difficult (e.g., nonmusicians recognizing complex rhythms), performance is more constrained by existing musical experience or processing capacity than by stimulus length itself.

4. General Discussion

The structured organization of pitch and rhythm and their cognitive processing constitute the foundation of human musical and linguistic abilities. This study investigated the relationship between pitch and rhythmic structure processing in auditory working memory and how task demands and musical training influence this processing mode. Specifically, we focused on whether performance was better under tonal than atonal conditions (pitch structure effect) and under simple than complex meter conditions (rhythmic structure effect), and whether these two effects influenced each other (manifesting as interactions) or oper-

ated independently. Additionally, we examined the contextual modulation of this processing relationship, analyzing differences between tasks and the impact of musical expertise. Results showed that in the pitch maintenance task, nonmusicians exhibited relatively independent processing of pitch and rhythmic structures, whereas musicians processed them interactively, with the interaction effect on sensitivity measures positively correlated with musical sophistication scores. However, in the rhythm maintenance task, both groups processed pitch and rhythmic structures independently. These findings indicate that musical training selectively modulates the interaction between pitch and rhythmic structures in auditory working memory.

Before examining the interaction patterns between pitch and rhythmic structures, we can compare their overall impact on working memory task performance. Across both memory tasks, rhythmic structure significantly influenced both types of memory, with simple, stable rhythms enhancing detection sensitivity and reducing response bias, even when task instructions required ignoring temporal information. This emphasizes the crucial role of rhythmic structure in musical working memory. In contrast, pitch structure did not provide similar cognitive support for rhythm working memory, as no sensitivity differences emerged between tonal and atonal conditions for rhythm change detection. In auditory working memory, rhythmic regularity may provide a clear temporal framework that helps listeners predict when musical events will occur, reducing cognitive load and facilitating efficient allocation of cognitive resources. This temporal regularity helps segment musical sequences into easily processable chunks, maintaining attention and promoting tracking of musical progression. Therefore, our results support dynamic attending theory, which posits that attention is fundamentally time-based and that endogenous brain oscillators may synchronize with external rhythmic signals, directing temporal attention to predicted time points and thereby enhancing predictive processing (Jones, 1976; Jones & Boltz, 1989). In contrast, although pitch is a primary factor affecting melodic complexity (Prince & Pfordresher, 2012), it lacks the temporal cues provided by rhythm, potentially making it less effective for organizing musical information. While pitch changes can highlight phrase boundaries (Zhang et al., 2016), they do not naturally form clear segments like rhythmic patterns (Yang et al., 2022), possibly resulting in lower structural organization during memory. Moreover, rhythm often involves bodily movement such as tapping or dancing, which engages the motor system and enhances memory through cross-modal integration (Chen et al., 2008). Pitch typically does not engage the motor system in the same way, offering fewer opportunities for multisensory memory enhancement. These factors may render rhythmic ("when") regularity more readily utilizable by cognitive systems in auditory working memory than pitch ("what") regularity.

This study confirmed that nonmusicians can process both pitch and rhythmic structures. However, pitch structure effects were not modulated by rhythmic structure, and vice versa, indicating that nonmusicians processed the two structures relatively independently. This finding aligns with the dual-pathway neu-

ral architecture hypothesis (Schwartze & Kotz, 2013), suggesting that temporal and non-temporal information processing may rely on distinct neural pathways. Prince (2011) also demonstrated that when nonmusicians focus on pitch or temporal information while ignoring the other dimension, the effects of tonality and meter on melodic pleasantness ratings are additive without interaction. Since our task explicitly required focusing on a single dimension while ignoring or suppressing the other, this independent structural processing may reflect participants' strategic approach. Nonmusicians may effectively concentrate cognitive resources on specific information dimensions through selective attention, thereby weakening integration between the two structures and demonstrating how processing strategies regulate auditory information processing.

Unlike nonmusicians, musicians showed interactive processing of pitch and rhythmic structures, although this interaction only emerged in the pitch maintenance task. This suggests that the interaction between temporal and non-temporal dimensions is modulated by both task demands and musical training experience. According to dynamic attending theory (Jones, 1976; Jones & Boltz, 1989), pitch and temporal structures can jointly form a combined accent structure that optimizes attentional resource allocation and facilitates processing of non-temporal information such as pitch. Our results not only support dynamic attending theory's predictions about pitch-time interactions but also reveal that such interactions may depend on task type and individual musical experience, enriching the theory's application in music cognition. Further simple effects analysis of the interaction showed that simple rhythms significantly enhanced tonal effects, and tonal melodies further enhanced rhythmic structure effects. This suggests that musical training may improve musicians' structural integration abilities, enabling them to efficiently process pitch and rhythmic structures simultaneously. Even when working memory tasks directed attention to a single information dimension, musicians could effectively utilize dual information to optimize task performance.

Nevertheless, musicians showed independent processing of structures in the rhythm maintenance task. This may be because the rhythm maintenance task was relatively easy, allowing musicians to complete it successfully through independent processing alone. In melodies, pitch changes generally outnumber duration changes, increasing cognitive load for pitch maintenance. Therefore, compared to the less demanding rhythm maintenance task, musicians' integration of pitch and rhythmic structures was more evident in the pitch maintenance task, suggesting that moderate task difficulty promotes interactive processing of the two structures. However, when cognitive load (sequence length) increased further, both musicians and nonmusicians showed a tendency toward independent processing to reduce cognitive demands. This indicates that musical training endows musicians with flexible strategic regulation abilities, allowing them to adaptively choose between interactive or independent processing strategies based on task demands, highlighting music training's significant impact on flexibility in complex auditory information processing. In contrast, nonmusicians consistently favored independent processing strategies across working memory

tasks, likely because both pitch and rhythm working memory tasks were comparably challenging for them. Thus, interactive processing may serve as a compensatory strategy, while independent processing may be the default approach when facing difficult tasks.

Furthermore, our findings can be interpreted within the framework of a two-stage processing model for pitch and temporal dimensions. According to Thompson et al. (2001), processing occurs in two stages: an early stage that independently encodes melodic features (e.g., pitch and duration), and a later stage that integrates these features into a coherent whole. EEG studies support this hypothesis, particularly in musical syntax processing (Sun et al., 2020; Zhang et al., 2019). Building on this model, Tillmann and Lebrun-Guillaud (2006) emphasized the task-dependent nature of pitch-time interactions, noting that independent processing dominates in tasks requiring local judgments, whereas integrative processing becomes more prominent in global evaluation tasks. In our working memory tasks, participants made "same" or "different" judgments based on changes in one dimension while suppressing interference from the other. This explicit suppression task design may have encouraged participants to focus more on local features of melodies, thereby strengthening the tendency toward independent rather than interactive processing of pitch and temporal dimensions. This aligns with Tillmann and Lebrun-Guillaud's (2006) view that tasks involving global processing are more likely to elicit feature integration effects. These findings suggest that the interaction between pitch and rhythmic structures is fundamentally dynamic—shaped jointly by task design, processing demands, and the degree to which musical dimensions are emphasized or suppressed in the task.

Beyond our primary findings, this study yielded several secondary results. Overall, musicians outperformed nonmusicians in sensitivity and reaction time and showed smaller response bias, consistent with previous research (e.g., Chen et al., 2008; Schulze et al., 2011; Schulze et al., 2012; Sun et al., 2018), indicating that musicians can effectively utilize pitch and rhythmic structures to improve auditory processing performance. Compared to musicians, nonmusicians did not show pitch structure effects on sensitivity measures in the rhythm task. This aligns with Schulze et al. (2011), who observed no behavioral differences between tonal and atonal conditions in nonmusicians. However, other studies (e.g., Albouy et al., 2013; Lévêque et al., 2022; Schulze et al., 2012) have observed that nonmusicians' behavioral sensitivity is affected by tonal structure. This inconsistency may stem from experimental design differences: our study combined pitch and rhythm changes, whereas previous studies typically used sequences with uniform timing and no rhythmic variation, which may have simplified the task and allowed participants to focus on pitch structure.

5. Conclusion

In summary, this study investigated the processing relationship between pitch and rhythmic structures in auditory working memory and how this relationship

is influenced by task demands and musical training. Results showed that non-musicians consistently exhibited independent processing of pitch and rhythmic structures across working memory tasks. Rhythmic regularity served as a temporal framework that helped nonmusicians improve task performance, whereas pitch structure did not provide similar support, suggesting that temporal regularity may be more readily utilized by cognitive systems in auditory working memory than non-temporal regularity. In contrast, musicians could flexibly modulate interactive or independent processing of pitch and rhythmic structures, reflecting that musical training endows them with the ability to adjust processing strategies according to task demands. Future research should further explore the interaction between pitch and rhythmic structures under different task demands and cognitive loads, and investigate the role of different types of musical training in this process.

References

Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, *1537*, 224–232. https://doi.org/10.1016/j.brainres.2013.09.003

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*(1), 63–102. https://doi.org/10.1016/0010-0277(83)90003-3

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, *20*(2), 226–239. https://doi.org/10.1162/jocn.2008.20018

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, *24*, 87–185. https://doi.org/10.1017/s0140525x01003922

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, *50*(4), 305–313. https://doi.org/10.3758/bf03212222

Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*(1), 1–7. https://doi.org/10.3758/BF03207132

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses.* Behavior Research Methods*, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, *7*, Article 68. https://doi.org/10.3389/fnsys.2013.00068

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage*, *47*(2), 735–744. https://doi.org/10.1016/j.neuroimage.2009.04.090

Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *NeuroImage*, *57*(4), 1572–1579. https://doi.org/10.1016/j.neuroimage.2011.05.061

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491. https://doi.org/10.1037/0033-295X.96.3.459

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355. https://doi.org/10.1037/0033-295X.83.5.323

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541. https://doi.org/10.1162/089892900562183

Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, *44*(3), 476–490. https://doi.org/10.1111/j.1469-8986.2007.00517.x

Koelsch, S., Schmidt, B. H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*, *39*(5), 657–663. https://doi.org/10.1111/1460-9568.00075

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195148367.001.0001

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*(1), 159–179. https://doi.org/10.1037/0033-2909.126.1.159

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval, in C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Springer-Verlag.

Lévêque, Y., Lalitte, P., Fornoni, L., Pralus, A., Albouy, P., Bouchet, P., … & Tillmann, B. (2022). Tonal structures benefit short-term memory for real music: Evidence from non-musicians and individuals with congenital amusia. *Brain and Cognition*, *161*, Article 105881. https://doi.org/10.1016/j.bandc.2022.105881

Lin, H. R., Kopiez, R., Müllensiefen, D., & Wolf, A. (2021). The Chinese version of the Gold-MSI: Adaptation and validation of an inventory for the measurement of musical sophistication in a Taiwanese sample. *Musicae Scientiae*, *25*(2), 226–251. https://doi.org/10.1177/1029864919871987

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). London: Lawrence Erlbaum Associates.

Martin, T., Egly, R., Houck, J. M., Bish, J. P., Barrera, B. D., Lee, D. C., & Tesche, C. D. (2005). Chronometric evidence for entrained attention. *Perception & Psychophysics*, *67*(1), 168–184. https://doi.org/10.3758/BF03195020

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, *9*(2), Article e89642. https://doi.org/10.1371/journal.pone.0089642

Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440. https://doi.org/10.2307/40285311

Prince J. B. (2011). The integration of stimulus dimensions in the perception of music. *Quarterly Journal of Experimental Psychology*, *64*(11), 2125–2152. https://doi.org/10.1080/17470218.2011.573080

Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, *141*(2), 184–198. https://doi.org/10.1016/j.actpsy.2012.07.013

Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Memory & Cognition*, *37*(3), 368–381. https://doi.org/10.3758/MC.37.3.368

Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1598–1617. https://doi.org/10.1037/a0016456

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., & Nielsen, M. (1999). Neural representation of a rhythm depends on its interval ratio. *Journal of Neuroscience*, *19*(22), 10074–10081. https://doi.org/10.1523/JNEUROSCI.19-22-10074.1999

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, *29*(3), 255–267. https://doi.org/10.1525/mp.2012.29.3.255

Schulze, K., Müller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, *33*(1), 189–196. https://doi.org/10.1111/j.1460-9568.2010.07470.x

Schwartze, M., & Kotz, S. A. (2013). A dual-pathway neural architecture for specific temporal prediction. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2587−2596. https://doi.org/10.1016/j.neubiorev.2013.08.005

Sun, L., Thompson, W. F., Liu, F., Zhou, L., & Jiang, C. (2020). The human brain processes hierarchical structures of meter and harmony differently: Evidence from musicians and nonmusicians. *Psychophysiology*, *57*(9) Article e13598. https://doi.org/10.1111/psyp.13598

Sun, L., Liu, F., Zhou, L., & Jiang, C. (2018). Musical training modulates the early but not the late stage of rhythmic syntactic processing. *Psychophysiology*, *55*(2), Article e12983. https://doi.org/10.1111/psyp.12983

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research*, *70*(5), 345–358. https://doi.org/10.1007/s00426-005-0222-0

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 128–140. https://doi.org/10.1037//0096-1523.27.1.128

Yang, X., Shen, X., Zhang, Q., Wang, C., Zhou, L., & Chen, Y. (2022). Music training is associated with better clause segmentation during spoken language processing. *Psychonomic Bulletin & Review*, *29*(4), 1472–1479. https://doi.org/10.3758/s13423-022-02076-2

Zhang, J., Che, X., & Yang, Y.(2019). Event-related brain potentials suggest a late interaction of pitch and time in music perception. *Neuropsychologia*, *132*, 107118. https://doi.org/10.1016/j.neuropsychologia.2019.107118

Zhang. J., Jiang, C., Zhou, L., & Yang, Y. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia*, *91*, 490–498. https://doi.org/10.1016/j.neuropsychologia.2016.09.013

Zhou, L., Zhao, H., Jiang, C. (2017). Neural plasticity to musical performance training: A meta-analysis study. *Advances in Psychological Science*, *25*(11), 1877–1887. https://doi.org/10.3724/SP.J.1042.2017.01877

Source: ChinaXiv — Machine translation. Verify with original.