Music is similar to language in terms of working memory interference

Anna Marie Fennell [1] · Jennifer A. Bugos [2] · Brennan R. Payne [3] · Elizabeth R. Schotter [4]

Abstract
Some researchers theorize that musicians' greater language ability is mediated by greater working memory because music and language share the same processing resources. Prior work using working memory sentence processing dual-task paradigms have shown that holding verbal information (e.g., words) in working memory interferes with sentence processing. In contrast, visuospatial stimuli are processed in a different working memory store and should not interfere with sentence processing. We tested whether music showed similar interference to sentence processing as opposed to noninterference like visuospatial stimuli. We also compared musicians to nonmusicians to investigate whether musical training improves verbal working memory. Findings revealed that musical stimuli produced similar working memory interference as linguistic stimuli, but visuospatial stimuli did not—suggesting that music and language rely on similar working memory resources (i.e., verbal skills) that are distinct from visuospatial skills. Musicians performed more accurately on the working memory tasks, particularly for the verbal and musical working memory stimuli, supporting an association between musicianship and greater verbal working memory capacity. Future research is necessary to evaluate the role of music training as a cognitive intervention or educational strategy to enhance reading fluency.

Keywords Working memory · Music · Language · Interference

Music and language are cognitively intertwined in many ways (e.g., shared syntactic resources: Patel, 2003, 2013; auditory working memory systems: Salamé & Baddeley, 1989). Although musical training is linked with advantages in language skills in correlational (Corrigall, & Trainor, 2011; Deguchi et al., 2012; Piro, & Ortiz, 2009; Thompson, Schellenberg, & Husain, 2004; Tierney & Kraus, 2013) and interventional research (Benz, Sellaro, Hommel, & Colzato, 2016; Bhide, Power, & Goswami, 2013; Bugos & Mostafa, 2011), it is unclear which underlying cognitive mechanisms mediate these relationships. Some researchers believe working memory mediates the benefits of musicianship on reading ability (George & Coch, 2011; Suárez, Elangovan, & Au, 2016; Slevc & Okada, 2015), especially since working memory is linked to improvements in language outcomes (Caplan, 2016; Hussey et al., 2017; Payne & Stine-Morrow, 2017).

We suggest that, if working memory is the link between musical training and language skills, then tasks that require working memory to process music must interfere with tasks that require working memory for language. A key paradigm that is used to demonstrate the role of working memory in language processing is a dual-task working memory and sentence reading paradigm (Fedorenko, Gibson, & Rohde, 2006). When readers had to maintain verbal items in memory while reading sentences, comprehension and working memory performance both suffered, especially when both domains consumed more working memory resources (i.e., the sentences were syntactically complex and the working memory stimuli were similar to words in the sentence).

We adapted this dual-task paradigm to investigate whether music interferes with working memory for language processing like linguistic stimuli do. Working memory is traditionally described as having three components: the central executive, which is responsible for maintaining and switching attention, and two systems that are responsible for storing, rehearsing, and manipulating perceptual representations: the phonological loop (e.g., verbal information, speech sounds, tonal stimuli;

✉ Anna Marie Fennell
    amf2286@tc.columbia.edu

[1] Department of Counseling and Clinical Psychology, Teachers College, Columbia University, 525 W 120th St., New York, NY 10027, USA

[2] School of Music, University of South Florida, Tampa, FL, USA

[3] Department of Psychology, University of Utah, Salt Lake City, UT, USA

[4] Department of Psychology , University of South Florida , FL Tampa, USA

Salamé & Baddeley, 1989), and the visuospatial sketchpad (e.g., visual images; Baddeley, 1992). The distinction between these two working memory resources is supported through previous research (Shah & Miyake, 1996), which established that spatial skill is distinct from language skill because the two are not correlated with one another. Therefore, while we would expect verbal information to cause working memory interference with sentence processing, we would not expect visuospatial information to produce the same kind of interference.

Shared syntactic resources in music and language

A primary hypothesis as to why music and language share cognitive resources is that music and language are initially perceived independently, but share limited neural resources for syntactic integration and processing (Patel, 2003, 2013). This resource between language and music is demonstrated by a different dual-task paradigm in which participants performed a self-paced reading task and a musical chord coincided with the presentation of each word or short phrase in the sentence (Slevc, Rosenberg, & Patel, 2009). Readers took longer to read syntactically complex sentences than syntactically simple sentences, and this slowdown was amplified when it was paired with musical sequences that violated harmonic structures (i.e., when an out-of-key chord is presented, C major to B major), compared with reading with music that followed harmonic structure. The increase in processing difficulty when linguistic violations and musical structure violations are combined supports the idea that linguistic and musical information are processed using the same limited neural resources for syntax.

While music influences processing speed in a concurrent dual task, it is unclear whether it is drawing on working memory. If music and language share this limited resource for parsing syntax, then we should be able to find a similar interference for music as we do in linguistic dual-task scenarios (Fedorenko et al., 2006). We tested this by extending the working memory and sentence processing dual-task paradigm to include different kinds of working memory stimuli (e.g., language, music, visuospatial dots).

Working memory for language, visuospatial information, and music

It makes sense that music would be related to verbal skills because they are both sound based. If so, we theorize that music should interfere with verbal working memory, but not visual stimuli. In this case, perhaps interference functions through more abstract processes rather than perceptual similarity. Verbal working memory could be this commonality because it encompasses auditory speech, tonal stimuli, and written language.

While the Baddeley model predicts that musical skill is not related to visuospatial skill, it is possible that musicians develop improved visuospatial skill by learning to read complex scores at high speeds. Still, there are mixed results in terms of musicians' advantage on visuospatial tasks—some find an advantage (e.g., Jakobson, Lewycky, Kilgour, & Stoesz, 2008; Sluming, Brooks, Howard, Downes, & Roberts, 2007; Sluming et al., 2002), and others do not (e.g., Cohen, Evans, Horowitz, & Wolfe, 2011; Hansen, Wallentin, & Vuust, 2013; Ho, Cheung, & Chan, 2003; Talamini, Altoè, Carretti, & Grassi, 2017; Talamini, Carretti, & Grassi, 2016). Therefore, it is an open question in terms of musical training and improved visuospatial skill.

Musical training and experience

If music and language share working memory resources, a corollary prediction would be that musicians would exhibit less working memory interference than nonmusicians because their musical training would provide benefits to the verbal working memory system. Because past work by Slevc et al. (2009), among others, has tested only nonmusicians, there is scant research on this hypothesis. Research supports that musical training is associated with increased working memory (George & Coch, 2011; Suárez et al., 2016), but it is unclear whether that increased capacity affects the sharing of working memory resources across domains. In other words, would musicians perform better than nonmusicians on a dual-task working memory and sentence comprehension paradigm, and would they show different patterns of interference than nonmusicians?

MacDonald and Christiansen (2002) theorized that individual differences in sentence comprehension stem from an interaction between biological factors and language experience. Musicians tend to outperform nonmusicians on several general cognitive abilities such as long-term memory (Talamini et al., 2017), auditory working memory (Suárez, Elangovan, & Au, 2016; Talamini et al., 2017), and mathematics (Cheek, & Smith, 1999; Vaughn, 2000), as well as reading and linguistic skills such as detecting prosody within a sentence (Deguchi et al., 2012; Thompson et al., 2004), vocabulary (Piro & Ortiz, 2009), reading comprehension (Corrigall & Trainor, 2011), and word decoding (Tierney & Kraus, 2013). Such findings have led to the suggestion that music practice is linked with processes for language (Slevc & Okada, 2015) and music-based interventions may have generalizable effects that transfer over to other domains (Benz et al., 2016; Bugos & Mostafa, 2011).

If musical training is generalizable to other abilities such as language (i.e., far transfer), then we should expect improved

performance among musicians across both music and verbal working memory tasks. Such findings would suggest that extensive experience using the cognitive resources required for musical training can transfer to performance in a completely different domain such as language. More specifically, we hypothesize that musicians will use similar processes to accommodate challenging language and music (i.e., complex syntax and musical phrases with nonharmonic intervals) due to shared processing networks between language and music. But, if musical training only affects skills specific to music (i.e., near transfer), then we hypothesize improvements restricted to music-based tasks.

The current study

Following the dual-task logic laid out by Fedorenko et al. (2006), we manipulated the difficulty of the language task through the syntactic complexity of the sentences. We then compared performance to working memory stimuli that should not be processed by the same working memory system (e.g., visuospatial layouts; Wechsler, 1997). The key comparison is our test of whether musical stimuli pattern with linguistic stimuli or visuospatial stimuli. So, by using a language-only dual-task to demonstrate interference and a language-visuospatial task to demonstrate noninterference, we will have two boundary conditions to compare against a language-music task and observe whether music interferes with language tasks because it is processed in the verbal working memory system. We also manipulated the working memory load and whether the information was structured for the musical and visuospatial stimuli. To assess the relationship between musical training and working memory resources, we compared musicians to nonmusicians.

We hypothesized that (1) working memory accuracy would decrease for challenging stimuli (e.g., sentences with complex syntax, working memory stimuli with high load and/or unstructured harmony or geometry). We also hypothesized that (2) musicians would perform more accurately than nonmusicians because of an increased working memory capacity. However, (3) there may be different patterns for musicians in terms of interference depending on whether musical training provides an across-the-board increase to working memory capacity—in this case (3a) the increase should be equivalent across types of working memory stimuli. Or alternatively, if it is specific to the verbal system of working memory (i.e., the phonological loop), (3b) the increase in accuracy should be more pronounced for nouns and musical phrases relative to visuospatial layouts. Furthermore, if the processing of harmonic structure is analogous to the processing of syntactic structure, then (4a) structured musical phrases should lead to better working memory accuracy than nonstructured phrases, but (4b) possibly only for musicians because those with

musical training have more experience processing harmony and musical syntax.

Method

Participants

We conducted an a priori power analysis for a linear mixed-effects regression using the PANGEA program (Westfall, 2016) to determine our needed sample size with 10 items per condition, which is more than prior work (Fedorenko et al., 2006; Slevc et al., 2009). This analysis indicated that, in all blocks, we would be able to detect a standardized effect size of d =.25 or larger for the critical within-participants interaction between musical and linguistic variables—that is, the 2 (harmonic structure) × 2 (syntactic complexity) interaction in Block 2—with power equal to .89 with 64 participants (32 musicians and 32 nonmusicians). Due to COVID-19, we were not able to collect our full sample, but when we reran this power analysis with our smallest group-level sample size (22 nonmusicians), we found that our power decreased only to .77.

Sixty participants were recruited from the University of South Florida and the surrounding community, either through the psychology department's subject pool and compensated with course credit, or through flyers and word-of-mouth and compensated with financial incentives ($20/hour for musicians; $10/hour for nonmusicians). We followed all APA guidelines with respect to the treatment of human subjects. We excluded participants due to technical errors in the experimental script (n = 7) or misunderstanding instructions (n = 1). The analyses reported below include 22 nonmusicians and 30 musicians.

All participants were native English speakers, had normal or corrected-to-normal vision, no significant hearing loss, and no history of learning, reading, or musical disabilities. Criteria for musician status included a minimum of 5 years of music training, currently practicing an instrument (min. 5 hours per week), ability to read music, and experience performing music.[1] Musicians were instrumentalists (i.e., nonvocalists) and did not have perfect pitch. Criteria for nonmusicians included no experience performing music, less than 5 years of musical training, and currently not practicing an instrument.

To collect quantitative measures of musical background, we also collected a measure of aural music aptitude (Advanced Measures of Music Audiation [AMMA]; Gordon, 1989), and a self-report measure of musical training,

---

[1] Our criteria are slightly more inclusive than other researchers in terms of years of musical training, (i.e., 5 as opposed to 6, as suggested by Zhang, Susino, McPherson, & Schubert, 2020), but are more stringent in terms of practice schedule (i.e., we required 5 hours per week as opposed to the proposed 1 hour per week).

Table 1 Descriptive statistics (means with standard deviations in parentheses) of participants in the musician and nonmusician groups

Participant group	N	Age	Gender (% female)	MUSE: Hours of practice per day	MUSE: Years of musical experience	AMMA Score
Musician	30	20.6 (2.8)	0.47	4.53 (1.8)	11.6 (3.7)	34.2 (6.2)
Nonmusician	22	19.7 (1.6)	0.59	0.36 (0.9)	0.8 (1.4)	26.5 (5.0)

experience, and musical engagement (Musical USE [MUSE] questionnaire; Chin, & Rickard, 2012). The AMMA is an aural music aptitude test yielding a tonal, rhythm, and composite score (i.e., sum of the two sub-scores) that takes approximately 20 minutes to complete. The participant hears two short phrases of music and then selects an answer in a multiple-choice online form for whether they were the same, had a change in pitch, a change in rhythm, or whether they did not know (there are never any phrases that differ in both pitch and rhythm). The participant gets a point for every correctly answered question, loses a point for every incorrectly answered question, and neither loses nor gains a point if they do not know the answer. The MUSE questionnaire is an index of quality and quantity of music production and music reception that takes approximately 10 minutes to complete. It includes a music engagement profile with indices of music training, music instrument playing, and music listening, and five styles of music engagement (Cognitive & Emotional Regulation, Engaged Production, Social Connection, Physical Exercise, and Dance). Welch's t test for unequal independent groups revealed that, on average, the group of musicians reported practicing music for more hours per day, $t(45.26) = 11.06$, $p < .001$, and for more years, $t(39.16) = 14.81$, $p < .001$, and scored higher on the AMMA test composite score, $t(49.55) = 4.91$, $p < .001$, than the group of nonmusicians (see Table 1).

Materials and design

Participants each experienced 186 dual-task working memory-sentence processing trials in which they remembered a working memory stimulus and read a sentence. There were three blocks, one where the working memory stimuli were nouns ($n = 20$ trials), another where the working memory stimuli were musical phrases ($n = 80$ trials), and another where the working memory stimuli were visuospatial layouts of dots ($n = 80$ trials). The nouns block only manipulated sentence structure (simple vs. complex) in a single-factor, two-condition within-subjects design. The music and dots blocks were both 2 (sentence structure: simple vs. complex) × 2 (memory load: low vs. high) × 2 (memory item type: structured vs. unstructured) full-factorial within-subjects design. Therefore, each block contained 10 trials per condition, and started with two practice trials.

The sentences were taken from different studies manipulating syntactic complexity by either comparing subject versus object extraction (Fedorenko et al., 2006; e.g., "The violinist [who flattered the cellist/who the cellist flattered] played a piece from the symphony."), garden-paths (Schotter, Tran, & Rayner, 2014; additional stimuli that were written for this study), reduced relative clauses (Rayner, Carlson, & Frazier, 1983), or whether the critical participial verb had the same surface form as the verb's simple-past form (Levy, Bicknell, Slattery, & Rayner, 2009). Of the 360 sentences used in the experiment, only 84 were not taken from published sources. Sentences from different sources were presented evenly across conditions and blocks and the order of trials was fully randomized within blocks. Sentences were fully counterbalanced across conditions so that no participant saw the same item more than once, and sentences from each source were evenly distributed across conditions and blocks. There was a question after each sentence that probed for syntactic understanding ($n = 138$; e.g., "Did the violinist flatter the cellist?" YES/NO) or for semantic understanding ($n = 42$; e.g., "While Sam cried the bread in the toaster caught on fire." "Was the bread okay?" YES/NO). Our counterbalancing scheme consisted of 16 lists in which 8 lists saw a nouns block, then a music block, then a dots block, and the other eight lists saw a nouns block, then a dots block, then a music block.

The memory stimuli in the nouns block were similar to the semantic information of the sentence (e.g., occupations: Teacher–Lawyer–Doctor). After the sentence, the words

Fig. 1 Example stimuli from the music block (presented auditorily), manipulated for number of notes (few/many) and harmonic structure (structured/unstructured)
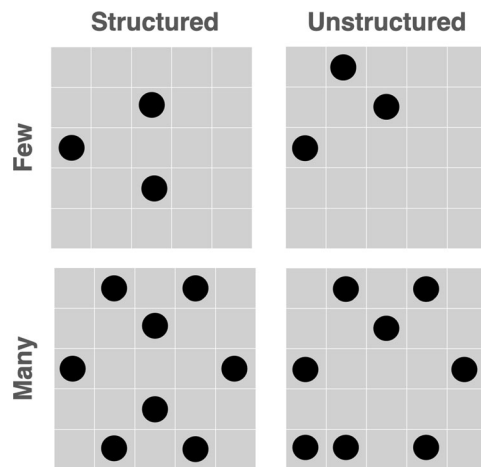
Fig. 2 Example stimuli from the dots block, manipulated for number of dots (few/many) and visual structure (structured/unstructured)

could be in the same or a different order (e.g., Lawyer–Doctor–Teacher). The memory stimuli in the music block were musical phrases, which were either structured or unstructured, and had either few or many notes (see Fig. 1). The melodies were composed in MuseScore with a piano timbre and were modeled from Halpern and Bower (1982). The structured musical phrases originated from *American Folk Songs for Children* (Seeger, 1948). The few-notes condition had five notes (avg. duration = 2.74 s), while the many-note conditions had 10 notes (avg. duration = 5 s). The unstructured phrases were edited from the structured phrases to sound similar, but violate harmony, resulting in dissonant intervals, lack of internal structure, and/or a hard-to-locate tonic. To adjust the phrases for the unstructured and different conditions, one to five notes within each phrase were altered. This was done by adjusting the original notes up or down by one half or whole step. For these adjustments, the few-note conditions were altered by one or two notes, and the many-note conditions were altered by three to five notes. The different variations did not change the structuredness, but lowered or raised select notes of the phrases. The first and last notes of each phrase remained unchanged.

The memory stimuli in the dots block were a layout of dots within a 5 × 5 grid (see Fig. 2). The matrix was 166 px × 166 px presented in the center of the screen, allowing for four degrees of visual angle from the center of the matrix to the perimeter. Structured dot matrices were based on regular (i.e., equiangular/equilateral) geometric shapes (e.g., squares or triangles) or straight lines, either in the center or the corner of the matrix, whereas the unstructured dot matrices were not vertically symmetrical and did not have any distinct regular geometric shape. The few conditions had three or four dots, and the many conditions had eight or nine dots. Every dot layout for the many conditions was built from its few condition

counterparts. The variation for each different condition was consistent within each pair (e.g., shifting the figure up one unit, rotating the figure 90 degrees).

Procedure

All participants provided informed consent after study procedures were explained to them and the voluntary nature of participation was emphasized. No identifying information (e.g., names) was obtained, and the only information connected to their data files was an arbitrary code. After providing informed consent, participants took the AMMA assessment, and then proceeded to the main experiment. On each trial, participants were presented with a memory element (three nouns, a melody, or a dot matrix) followed by a sentence and then a comprehension question. After they answered the comprehension question, they had to judge whether a second memory element was the same as the first one (see Figs. 3, 4 and 5). The nouns block was always presented first, and the order of the music block and the dots block were counterbalanced across participants. Between the music and the dots blocks, the participants completed the music background questionnaire (MUSE). Data collection took approximately 2 hours total per participant and occurred one participant at a time.

The experiment was programmed in SR Research Experiment Builder (SR Research Ltd., 2011). Participants were seated approximately 60 cm away from an HP p1230 CRT monitor (1,024 × 768 resolution, 150-Hz refresh rate), connected to sound-canceling headphones, and responses to comprehension questions and memory questions were recorded by buttons on a response pad. The memory stimulus in the nouns block and the dots block was presented in the center of the screen for 1,800 ms. While the memory stimulus in the music block was being played, a treble clef image was presented on the screen. After the initial working memory stimulus was presented (e.g., nouns, music, dots), and then they had the option of taking time to rehearse the stimuli by looking at a crosshair in the center of the screen (i.e., "crosshair" time). Then, the participant triggered a sentence to appear by looking at a black box on the left side of the screen; they read the sentence as many times as needed, and pressed a button to indicate that they were finished reading. Afterward, they answered a yes/no comprehension question about the sentence. Finally, a working memory test array was presented, where participants responded whether the initial working memory stimulus and test array matched. The participant responded whether it was the same or different on a response pad.
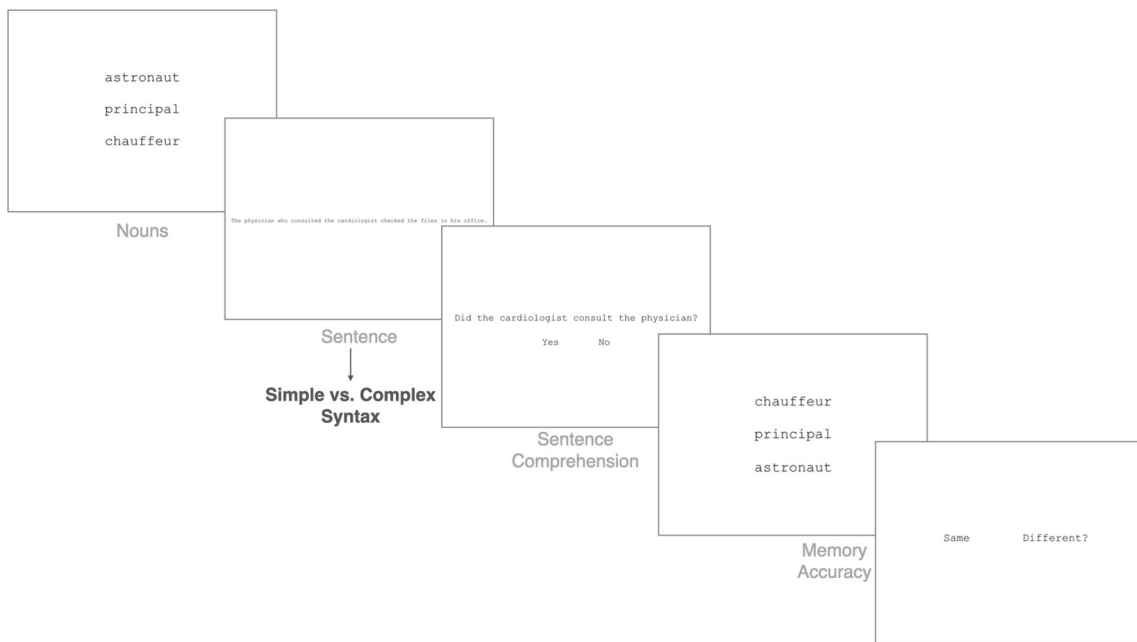
Fig. 3 Procedure for the nouns block. The participant saw a list of three nouns (occupations that were similar to the entities in the sentence), then the participant read a sentence (manipulated for complexity: simple/complex), answered a comprehension question about the sentence, then the nouns were presented again (either in the same or different order), and the participant responded whether they were the same or different. Each participant was specifically instructed that the same/different question referred to the *order* of the nouns
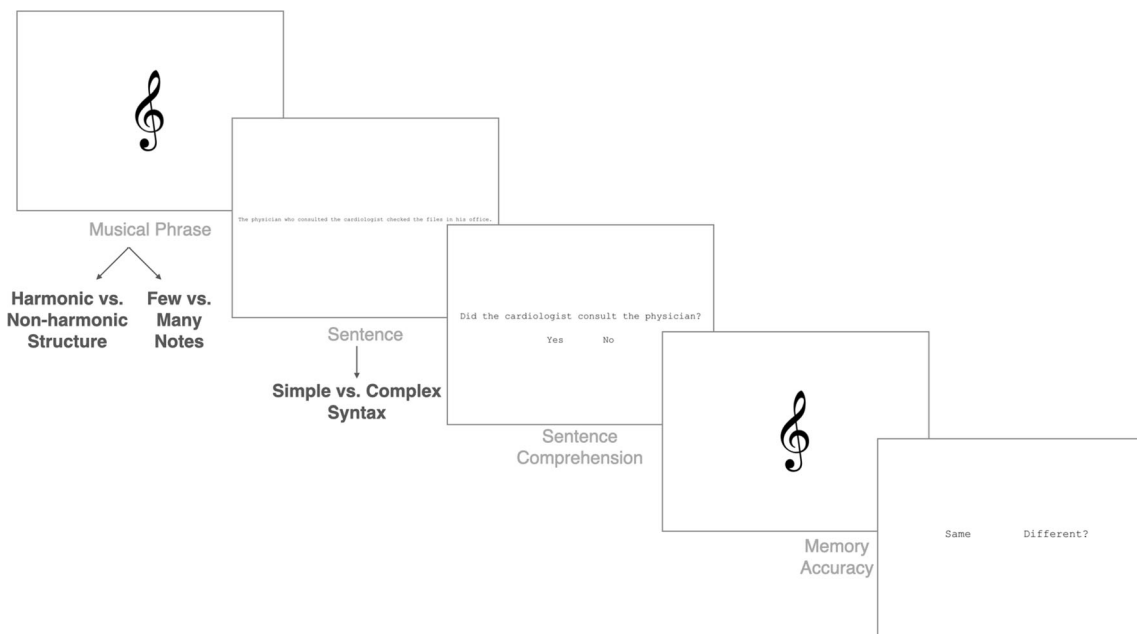

Fig. 4 Procedure for the music block. The participant heard a musical phrase auditorily (manipulated for structure: harmonic vs. nonharmonic and memory load: few vs. many notes), then the participant read a sentence (manipulated for complexity: simple/complex), answered a comprehension question about the sentence, then a melody is played again (either the same melody or a different melody) and the participant responded whether it is was the same or different
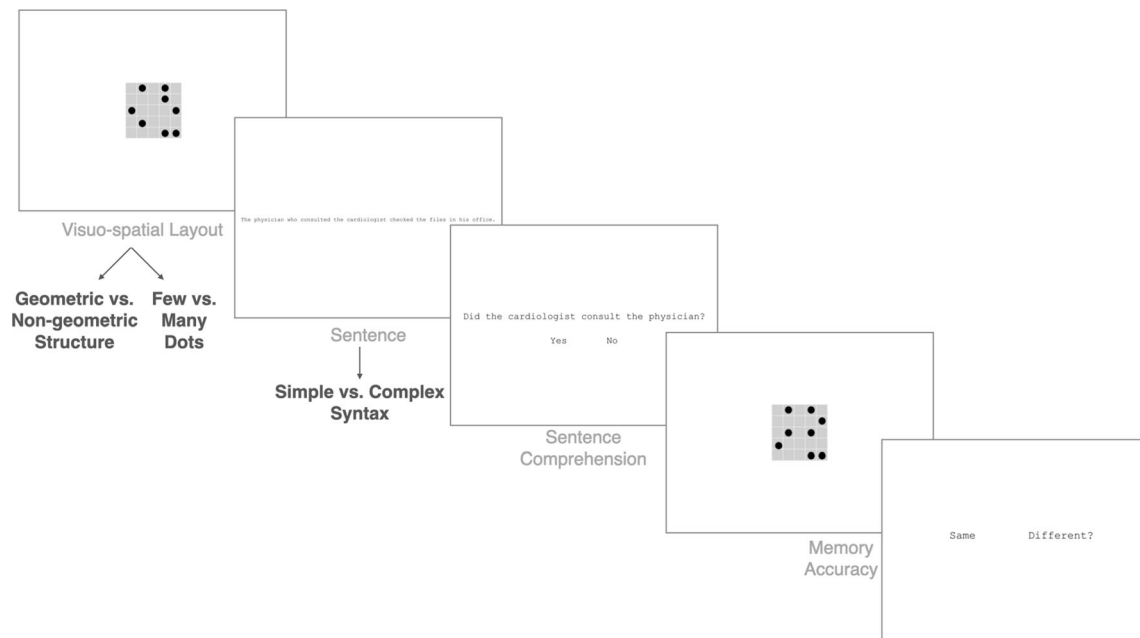
Fig. 5 Procedure for the dots block. The participant saw a dot matrix (manipulated for structure: geometric/nongeometric and working memory load: few vs. many dots), then the participant read a sentence (manipulated for complexity: simple/complex), answered a comprehension question about the sentence, then a second matrix was presented (either the same matrix or a different matrix), and the participant responded whether it is was the same or different

Results

Data were analyzed with (generalized) linear mixed-effects models (G)LMMs using the glmer() function from the lme4 package (Version 1.1- 12; Bates, Maechler, Bolker, & Walker, 2015) within the R Environment for Statistical Computing (Version 3.3.2; R Development Core Team, 2016). Model structures are specified in each section, below.

General memory accuracy

We conducted a logistic regression with memory accuracy as the dependent measure (see Table 2 and Fig. 6). The model included three factors: (1) block, which consisted of two contrasts (one that compared between the music and noun blocks, and the second that compared between the music and dot blocks); (2) sentence complexity, which compared simple to complex sentences with a sum-to-zero contrast; and (3) musicianship status, which compared nonmusicians to musicians with a sum-to-zero contrast. The model also included the two-way and three-way interactions between each of these factors. Therefore, the model produced 11 contrasts: one comparing dots to music, one comparing music to nouns, one for the effect of sentence complexity, one for the effect of musicianship status, two for the interaction between block and complexity, two for the interaction between musicianship status and complexity, and two for the three-way interaction of block, status, and complexity. We included crossed random effects of items and participants (intercepts only because some participants and items exhibited ceiling effects in some conditions and the model could not estimate variances around random slopes).

Accuracy was significantly higher in the dots block than in the music block ($p < .001$), but there was no difference between the music and nouns blocks ($p = .35$). Musicians significantly outperformed nonmusicians ($p < .001$). There was no main effect of sentence complexity ($p = .67$), no two-way interactions between complexity and either of the contrasts between blocks (both $ps > .69$), and no three-way interactions (all $ps > .7$). This is consistent with previous studies (see Fedorenko et al., 2006) in which we only see complexity effects for sentence comprehension, not working memory accuracy (see below).

There was a significant interaction between sentence complexity and musicianship status ($p < .01$). Follow-up analyses performed separately for each of the participant groups revealed that while the musicians have higher

Table 2 Results of the lmer model for memory accuracy ~ block × complexity × status

Predictors	Memory accuracy			
	Odds ratios	CI	Statistic	p
(Intercept)	4.02***	[3.42, 4.74]	16.72	**<.001**
Music-Dots	0.51***	[0.41, 0.63]	−6.23	**<.001**
Nouns-Music	1.17	[0.84, 1.62]	0.93	.351
Complexity	1.05	[0.83, 1.34]	0.43	.665
Musicianship status	2.21***	[1.72, 2.85]	6.16	**<.001**
Music-Dols × Complexity	0.94	[0.62, 1.44]	−0.27	.787
Nouns-Music × Complexity	1.04	[0.54, 2.00]	0.11	.912
Music-Dots*Status	1.61***	[1.28, 2.02]	4.13	**<.001**
Nouns-Music*Status	1.02	[0.72, 1.44]	0.11	.916
Complexity × Status	1.47**	[1.14, 1.90]	2.93	**.003**
Music-Dots × Complexity × Status	0.92	[0.58, 1.44]	−0.38	.708
Nouns-Music × Complexity × Status	1.92	[0.96, 3.83]	1.84	.065
Random effects				
σ^2	3.29			
$\tau_{00\ item}$	0.63			
$\tau_{00\ subject}$	0.15			
ICC	0.19			
$N_{subject}$	51			
N_{item}	360			
Observations	9,179			
Marginal R^2/Conditional R^2	0.057/0.238			

Note. *p < .05. **p < .01. ***p < .001

Boldface indicates statistical significance at the alpha = .05 level or below

accuracy in the simple conditions (estimate = −0.90, SE = 0.14, z value = −6.48, p < .001), the nonmusicians have higher accuracy in the complex conditions (estimate = −0.62, SE = 0.13, z value = −4.85, p < .001). There was no three-way interaction between sentence

complexity, musicianship status, and the contrast between the music and noun blocks (p = .07). There was no two-way interaction between musicianship status and the contrast between the noun and music blocks (p = .80), but there was a significant two-way interaction
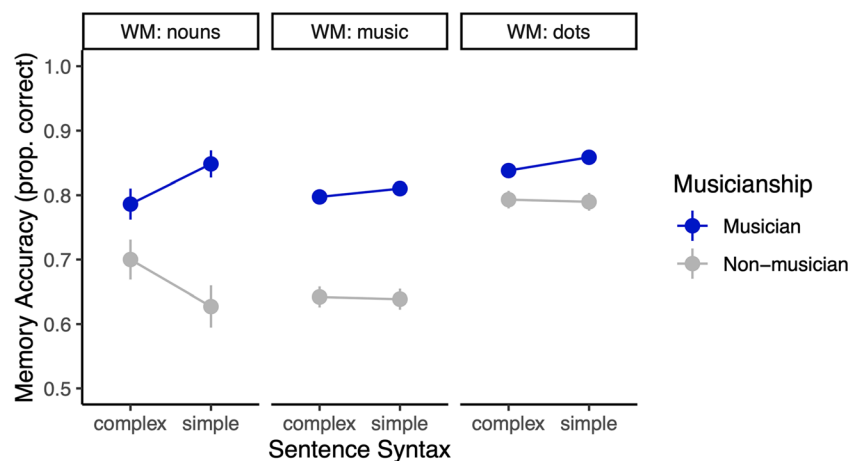

Fig. 6 Memory item accuracy across blocks (nouns/music/dots) as a function of sentence complexity (simple/complex) and musicianship status. Error bars represent 1 ±SEM
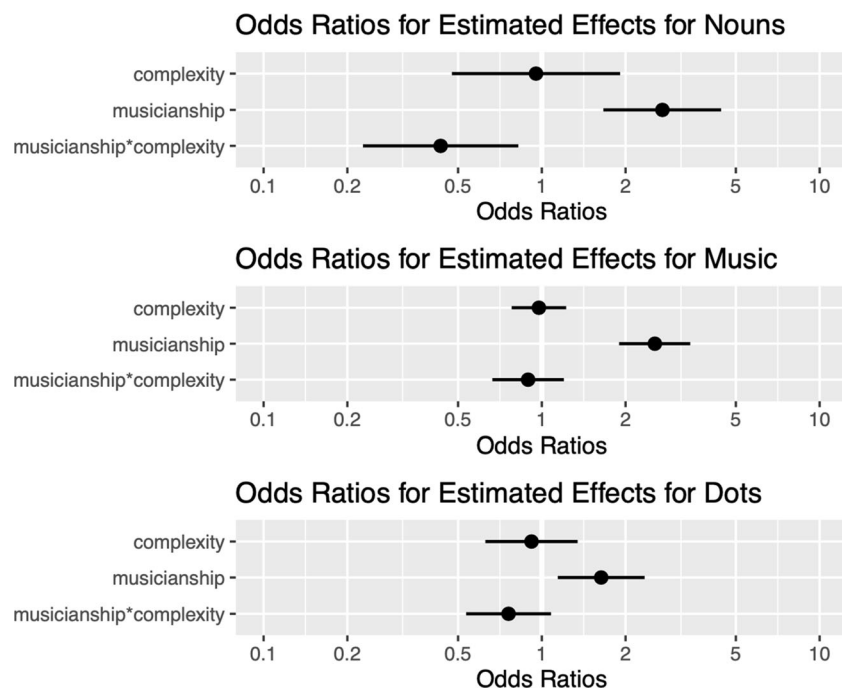
Fig. 7 Estimates of effect size and variability for the effects of complexity, musicianship status, and their interaction on memory accuracy within the noun, music, and dots block

between musicianship status and the contrast between the music and dot blocks ($p < .001$) with a larger advantage for musicians over nonmusicians in the music block compared with the dot block. Follow-up analyses run for the difference in memory accuracy as a function of musicianship status separately in each block revealed that musicians were more accurate than nonmusicians for all types of working memory stimuli (all ps < .001), although the effect sizes did vary between blocks (see below).

We visualized the model estimates of effect size and variability for the effects of complexity, musicianship status, and their interaction from the separate models for the nouns block, music block, and dots block using the sjplot() package in R (see Fig. 7). The model effect estimates are represented as odds ratios, the exponent of the log-odds coefficients from the GLMM models. An odds ratio of 1 indicates a null effect of the fixed effect of interest (i.e., conditions in which the participant is equally likely to be accurate as to be inaccurate), and odds ratios further away from 1 represent effect sizes that are larger, with those below 1 representing that accuracy decreases and those above 1 representing that accuracy increases for complex sentences or musicians. As reported above, there is an effect of musicianship status (i.e., musicians perform better than nonmusicians on memory tasks), which is larger for the nouns and

music blocks compared with the dots block, indicated by the estimates being further above 1.

Sentence accuracy

To ensure that the effects in the working memory task were not due to different strategies for prioritizing the two tasks,[2] we analyzed the sentence accuracy with the same model structure as the model for general memory accuracy (see Table 3 and Fig. 8). There was significantly higher accuracy in the simple conditions compared with the complex conditions ($p < .001$), replicating past research (Fedorenko et al., 2006; Gordon, Hendrick, & Levine, 2002), and musicians significantly outperformed nonmusicians ($p < .01$). However, there were no main effects of block (both ps > .08), nor were

_____

[2] We also tested whether the time spent on the screen before revealing the sentence (i.e., crosshair time) or time spent processing the sentence influenced memory accuracy. There was no main effect of crosshair time, nor any interactions with block. There was an interaction with musicianship, but this cannot explain the group differences in accuracy; nonmusicians who spent longer were more accurate in the memory task, while musicians always outperformed nonmusicians and showed no relationship between crosshair time and memory accuracy. There was no main effect of sentence processing time. Although there was an interaction between sentence processing time and block, this cannot explain the working memory accuracy patterns either. There was no relationship between sentence processing time and memory accuracy in the nouns block (all ps > .09). These data are available on the Open Science Framework.

Table 3 Results of the lmer model for sentence accuracy ~ block × complexity × status

Predictors	Sentence accuracy			
	Odds ratios	CI	Statistic	p
(Intercept)	5.50***	[3.97, 7.60]	10.28	**<.001**
Music-Dots	1.14	[0.78, 1.67]	0.69	.489
Nouns-Music	0.59	[0.33, 1.07]	−1.74	.082
Complexity	6.06***	[3.92, 9.36]	8.11	**<.001**
Musicianship status	2.20**	[1.33, 3.63]	3.08	**.002**
Music-Dots × Complexity	1.65	[0.78, 3.52]	1.31	.192
Nouns-Music × Complexity	0.72	[0.22, 2.35]	−0.55	.585
Music-Dots × Status	1.16	[0.89, 1.51]	1.13	.260
Nouns-Music × Status	0.83	[0.56, 1.24]	−0.90	.366
Complexity × Status	1.05	[0.79, 1.42]	0.35	.723
Music-Dots × Complexity × Status	1.31	[0.77, 2.21]	0.99	.321
Nouns-Music × Complexity × Status	1.95	[0.87, 4.37]	1.63	.103
Random effects				
σ^2	3.29			
$\tau_{00\ item}$	2.47			
$\tau_{00\ subject}$	0.74			
ICC	0.49			
$N_{subject}$	51			
N_{item}	360			
Observations	9,179			
Marginal R^2/Conditional R^2	0.137/0.563			

Note. *p < .05. **p < .01. ***p < .001.

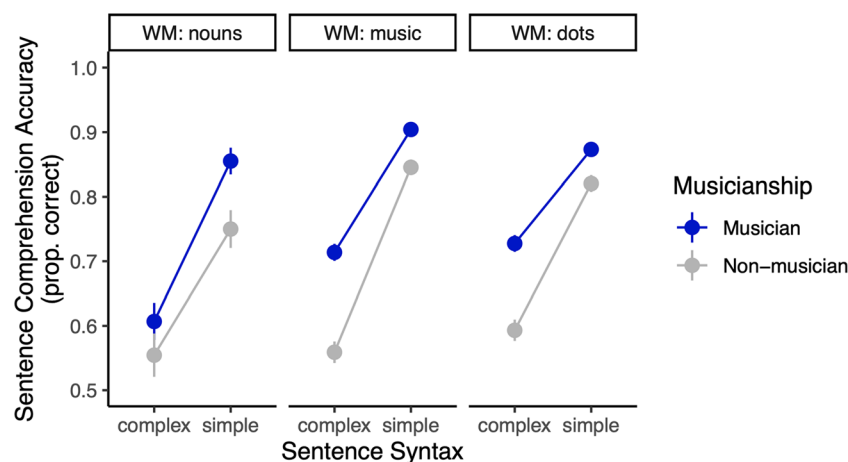Boldface indicates statistical significance at the alpha = .05 level or below

Fig. 8 Sentence comprehension accuracy across blocks (nouns/music/dots) as a function of sentence complexity (simple/complex), and musicianship status. Error bars represent ±1 *SEM*

there any two-way or three-way interactions between any of the factors (all *p*s > .10). These patterns suggest that all participants were sensitive to the manipulation in the sentence processing task equally across the different blocks. Furthermore, these data suggest that the musicians' better performance in the memory accuracy task could not have been due to a decrease in effort allocated to the sentence processing task relative to the nonmusicians.

Effects of memory structure and load

We constructed two linear mixed-effects models: one for memory accuracy in the music block (see Fig. 9), and another for memory accuracy in the dots block (see Fig. 10 and Table 4). Both models included four factors: (1) sentence complexity, which compared simple with complex sentences with a sum-to-zero contrast; (2) structure, which compared unstructured with structured working memory items with a sum-to-zero contrast; (3) load, which compared many with

Fig. 9 Memory item accuracy in the music block as a function of sentence complexity (simple/complex), harmonic structure (structured/unstructured), load (few/many), and musicianship status. Error bars represent ±1 *SEM*
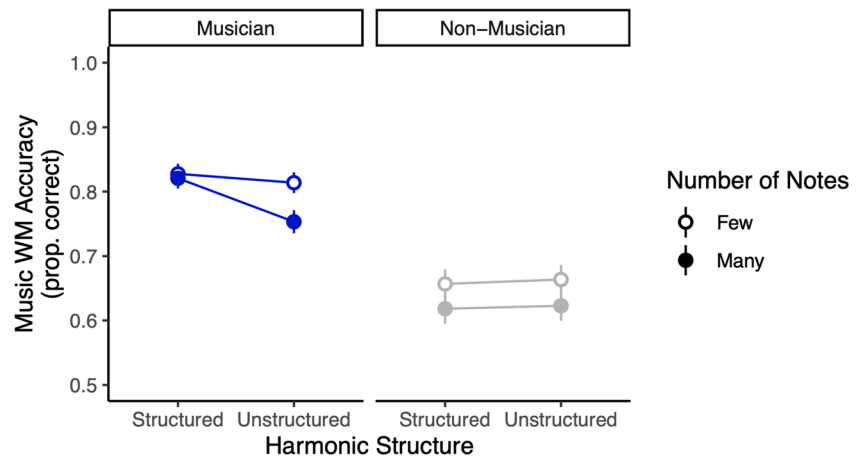

Fig. 10 Memory item accuracy in the dots block as a function of sentence complexity (simple/complex), visual structure (structured/unstructured), load (few/many), and musicianship status. Error bars represent ±1 *SEM*
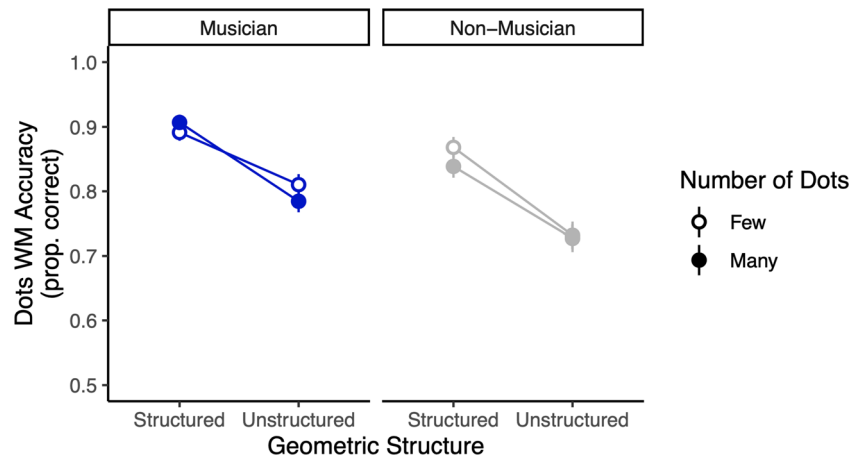
Table 4 Results of the lmer models for memory accuracy ~ structure × load × status within the music block, and memory accuracy ~ structure × load × status within the dots block

Predictors	Music memory accuracy				Dots memory accuracy			
	Odds ratios	CI	Statistic	p	Odds ratios	CI	Statistic	p
(Intercept)	3.02***	[2.54, 3.60]	12.55	**<.001**	7.21***	[5.57, 9.33]	15.03	**<.001**
Structure	0.89	[0.77, 1.03]	−1.52	.128	0.39***	[0.33, 0.47]	−9.88	**<.001**
Load	0.81**	[0.70, 0.94]	−2.75	**.006**	0.93	[0.78, 1.12]	−0.77	.443
Musicianship Status	2.57***	[1.91, 3.47]	6.21	**<.001**	1.67**	[1.16, 2.41]	2.73	**.006**
Soucturc × Load	0.83	[0.61, 1.11]	−1.26	.206	0.99	[0.68, 1.42]	−0.08	.938
Structurc × Status	0.74*	[0.55, 1.00]	−1.99	**.047**	0.95	[0.66, 1.37]	−0.27	.786
Load × Status	0.97	[0.72, 1.30]	−0.23	.815	1.14	[0.79, 1.64]	0.68	.494
Structure × Load × Status	0.73	[0.40, 1.32]	−1.05	.296	0.54	[0.26, 1.11]	−1.67	.095
Random effects								
σ^2	3.29				3.29			
τ_{00}	0.30 item				1.19 item			
	0.21 subject				033 subject			
ICC	0.14				032			
N	51 subject				51 subject			
	160 item				160 item			
Observations	4079				4,080			
Marginal R^2/Conditional R^2	0.060/0.188				0.057/0.355			

Note. *p < .05. **p < .01. ***p < .001

Boldface indicates statistical significance at the alpha = .05 level or below

few working memory items with a sum-to-zero contrast; and (4) musicianship status, which compared nonmusicians with musicians with a sum-to-zero contrast, as well as the interactions between these factors. Therefore, the models produced seven contrasts: one for the effect of structure, one for the effect of load, one for the effect of musicianship status, one for the interaction between structure and load, one for the interaction between structure and musicianship status, one for the interaction between load and musicianship status, and one for the three-way interaction between structure, load, and musicianship status. The models also contained crossed random effects of items and participants (intercepts only).

Working memory accuracy in the music block There was no significant main effect of structure ($p = .21$). There was a significant effect of load ($p < .01$) in which accuracy was higher when there were fewer notes. There was a significant effect of musicianship status ($p < .001$), in which musicians outperformed nonmusicians. There was a significant interaction between structure and musicianship status ($p < .05$), in which musicians have higher memory accuracy for structured musical phrases whereas nonmusicians show no difference between structured and unstructured musical phrases (see Fig. 9). There were no significant interactions between structure and load, or load and status, nor was there a significant three-way interaction of structure, load, and status (all $ps > .1$).

Working memory accuracy in the dots block There was a significant effect of structure ($p < .001$), in which accuracy was higher in the structured conditions. There was a significant effect of musicianship status ($p < .05$), in which musicians outperformed nonmusicians. There was no effect of load, and no two-way interactions between structure, load, or status (all $ps > .4$). There was not a significant three-way interaction between structure, load, and musicianship status ($p = .09$).

Discussion

The primary goals of this study were to investigate whether music relied on similar working memory systems compared with language or visuospatial stimuli. We found that music interfered with sentence processing similarly to language, indicated by higher memory accuracy in the dots block and the greater effect sizes for the nouns and music blocks. We also found that musical training was associated with greater performance. Musicians had a greater verbal, musical, and visuospatial memory than nonmusicians, and took advantage of the harmonic structure within musical melodies. These findings suggest that musical training could provide a benefit to working memory, specifically through verbal working memory by organizing information through syntactic structure (i.e., chunking) and utilizing the language system.

Our data support the theory that music taxes the working memory system in a similar way to language, whereas visuospatial stimuli do not show the same pattern—which is consistent with prior work showing that verbal and visuospatial working memory are distinct systems (Baddeley, 1992; Shah & Miyake, 1996). This finding is particularly interesting considering how the working memory tasks in previous studies were founded on similarity-based interference (e.g., Fedorenko et al., 2006); the fact that we see interference for music on verbal memory, even though they have very little surface similarity, challenges this phenomenon. It could be that because music taxes language, which is a different domain, similarity-based interference is more abstract, and not necessarily based on perceptual relatedness alone.

Our data support the theory that musical training is associated with a generalizable improvement in working memory because musicians were better than nonmusicians across not only linguistic and musical items (representing verbal working memory) but also dots (representing the visuospatial working memory). However, the difference in performance between musicians and nonmusicians was larger in the verbal and musical tasks, compared with visuospatial tasks. The presence of any advantage in the visuospatial task is somewhat surprising given the past literature not showing such advantages (see Cohen et al., 2011; Hansen et al., 2013; Ho et al., 2003; Talamini et al., 2017; Talamini et al., 2016). While musical training seems to benefit the working memory system as a whole, it seems to provide a benefit more dramatically to verbal skill (i.e., language and music) rather than visuospatial skill. This is consistent with experimental findings that demonstrate benefits of music training on verbal memory performance (Bugos, 2010; Ferreri, Aucouturier, Muthalib, Bigand, & Bugaiska, 2013; Franklin et al., 2008; Ho et al., 2003).

We anticipated a general pattern that musical and visuospatial stimuli would be more challenging (i.e., lower accuracy) when it was unstructured. And, we predicted that musicians would have more experience to adapt to the lack of structure (i.e., the potential deficit would be smaller in the musician group compared with the nonmusician group). While nonmusicians were more accurate when the musical phrases had fewer notes, regardless of harmonic structure, they were more accurate when the visuospatial layouts were geometrically structured, regardless of how many dots there were. Of interest was whether the musicians would show similar patterns of working memory interference, structure, and load. For tonal memory, musicians were able to remember melodies more when they were harmonically structured, and this structure helped them compensate for when there were many notes in the musical phrase. This supports the theory that musical training may be associated with an increased sensitivity to syntactic structure, particularly within music (e.g., harmony; Patel, 2003, 2013), and the observed advantage may be due to chunking strategies rather than a general increase in capacity.

For visuospatial memory, in contrast, the musicians showed the same pattern as the nonmusicians in that they were sensitive to the geometric structure, regardless of the number of dots. This suggests that the benefit of musicianship to working memory is specific to verbal processes and does not change processes associated with visuospatial working memory.

Limitations and future directions

Although we aimed to make the memory tasks as analogous as possible, it could be argued that the difference in performance across blocks is due to the variance in task difficulty rather than the differences in how the stimuli are processed (i.e., in the verbal versus visuospatial working memory system). We believe this alternative explanation is unlikely because there were similar effects of the working memory tasks on sentence comprehension accuracy and because musicians did not show differences in working memory accuracy across tasks. However, this is an important avenue for future research; to test this, future studies should test this by norming each task for difficulty by running a single-task version of the working memory blocks without an intervening sentence processing task and ensure they are equated on accuracy level.

Conclusion

Our study provides evidence from a dual-task interference paradigm that music taxes the verbal working memory system in a similar way to language, and that those with musical experience are less affected by this interference, perhaps due to greater verbal working memory. Musicians displayed less interference for linguistic and musical stimuli compared with visuospatial stimuli, which supports the distinction between these component parts of the working memory system. Musicians benefitted from harmonic structure, suggesting that musical training affords the ability to chunk musical information based on harmony. Because musical phrases produced similar patterns of interference as written text, this suggests that verbal working memory is based on abstract mental representations rather than perceptual representations. Furthermore, this suggests that musical training could confer benefits to cognitive systems in a domain-general way that extend beyond processing music, such as reading and processing language.

References

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music makes the world go round: The impact of musical training on nonmusical cognitive functions—A review. *Frontiers in Psychology, 6*, 2023.

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education, 7*(2), 113–123.

Bugos, J. A. (2010). The benefits of music instruction on processing speed, verbal fluency, and cognitive control in aging. *Music Education Research International, 4*, 1–9.

Bugos, J. A., & Mostafa, W. (2011). Musical training enhances information processing speed. *Bulletin of the Council for Research in Music Education, 187*, 7–18.

Caplan, D. (2016). Working memory and sentence comprehension. In G. Hickock & S. M. Small (Eds.), *Neurobiology of language* (pp. 633–645). New York, NY: Academic Press.

Cheek, J. M., & Smith, L. R. (1999). Music training and mathematics achievement. *Adolescence, 34*(136), 759.

Chin, T., & Rickard, N. S. (2012). The music USE (MUSE) questionnaire: An instrument to measure engagement in music. *Music Perception: An Interdisciplinary Journal, 29*(4), 429–446.

Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic Bulletin & Review, 18*(3), 586–591.

Corrigall, K. A., & Trainor, L. J. (2011). Associations between length of music training and reading skills in children. *Music Perception, 29*(2), 147–155.

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and nonmusicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research, 1455*, 75–89.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541–553.

Ferreri, L., Aucouturier, J. J., Muthalib, M., Bigand, E., & Bugaiska, A. (2013). Music improves verbal memory encoding while decreasing prefrontal cortex activity: An fNIRS study. *Frontiers in Human Neuroscience, 7*, 779. https://doi.org/10.3389/fnhum.2013.00779

Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music, 36*(3), 353–365.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia, 49*(5), 1083–1094.

Gordon, E. (1989). *Advanced measures of music audiation.* Chicago, IL: Gia Publications.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science, 13*(5), 425–430.

Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *The American Journal of Psychology, 95*(1), 31–50.

Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and nonmusicians. *Psychology of Music, 41*(6), 779–793.

Hussey, E. K., Harbison, J., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., & Novick, J. M. (2017). Memory and language improvements following cognitive control training. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 23.

Ho, Y. C., Cheung M. C., & Chan, A. A. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology 17,* 439–450. https://doi.org/10.1037/0894-4105.17.3.439

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception: An Interdisciplinary Journal, 26(1)*, 41–55.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America, 106*(50), 21086–21090.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*(1), 35–54.

Payne, B. R., & Stine-Morrow, E. A. (2017). The effects of home-based cognitive training on verbal working memory and language comprehension in older adulthood. *Frontiers in Aging Neuroscience, 9*, 256.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D. (2013). Sharing and nonsharing of brain resources for language and music. In M. A. Arbib (Ed.), *Language, music, and the brain* (Strüngmann Forum Reports), Vol. 10, J. Lupp, Series ed. Cambridge, MA: MIT Press.

Piro, J. M., & Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychology of Music, 37*(3), 325–347.

R Development Core Team (2016). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358–374.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A, 41*(1), 107–122.

Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during

reading. *Psychological Science, 25*(6), 1218–1226. https://doi.org/10.1177/0956797614531148

Seeger, R. C. (1948). *American folk songs for children.* Garden City, NY: Doubleday & Company, Inc.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of experimental psychology: General, 125*(1), 4.

Slevc, L. R., & Okada, B. M., (2015). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review, 22*(3), 637–652.

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374–381.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage, 17*(3), 1613–1622.

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *Journal of Neuroscience, 27*(14), 3799–3806.

SR Research Experiment Builder 1.10.165 [Computer software]. (2011). Mississauga, Ontario, Canada: SR Research Ltd.

Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults. *Australian Journal of Psychology, 68*(1), 38–46.

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLOS ONE, 12*(10), e0186773.

Talamini, F., Carretti, B., & Grassi, M. (2016). The working memory of musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal, 34*(2), 183–191.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help?. *Emotion, 4*(1), 46.

Tierney, A., & Kraus, N. (2013). Music training for the development of reading skills. *Progress in Brain Research*, *207,* 209–241.

Vaughn, K. (2000). Music and mathematics: Modest support for the oft-claimed relationship. *Journal of Aesthetic Education, 34*(3/4), 149–166.

Wechsler, D. (1997). *WAIS-III, Wechsler adult intelligence scale: Administration and scoring manual.* San Antonio, TX: Psychological Corporation.

Westfall, J. (2016). *PANGEA: Power ANalysis for GEneral Anova designs* (Working paper). Retrieved from http://jakewestfall.org/publications/pangea.pdf

Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music, 48*(3), 389–409.