



Published in final edited form as:

Psychol Music. 2017 January ; 45(1): 5–21. doi:10.1177/0305735616638942.

Visual-spatial sequence learning and memory in trained musicians

Esperanza M Anaya¹, David B Pisoni^{2,3}, and William G Kronenberger^{3,4}

¹Department of Behavioral Medicine, Northwestern University, Downers Grove, IL, USA

²Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

³DeVault Otologic Research Laboratory, Department of Otolaryngology–HNS, Indiana University School of Medicine, Indianapolis, IN, USA

⁴Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA

Abstract

Previous research has shown that musicians have enhanced visual-spatial abilities and sensory-motor skills. As a result of their long-term musical training and their experience-dependent activities, musicians may learn to associate sensory information with fine motor movements. Playing a musical instrument requires musicians to rapidly translate musical symbols into specific sensory-motor actions while also simultaneously monitoring the auditory signals produced by their instrument. In this study, we assessed the visual-spatial sequence learning and memory abilities of long-term musicians. We recruited 24 highly trained musicians and 24 nonmusicians, individuals with little or no musical training experience. Participants completed a visual-spatial sequence learning task as well as receptive vocabulary, nonverbal reasoning, and short-term memory tasks. Results revealed that musicians have enhanced visual-spatial sequence learning abilities relative to nonmusicians. Musicians also performed better than nonmusicians on the vocabulary and nonverbal reasoning measures. Additional analyses revealed that the large group difference observed on the visuospatial sequencing task between musicians and nonmusicians remained even after controlling for vocabulary, nonverbal reasoning, and short-term memory abilities. Musicians' improved visuospatial sequence learning may stem from basic underlying differences in visual-spatial and sensory-motor skills resulting from long-term experience and activities associated with playing a musical instrument.

Keywords

auditory perception/cognition; memory; recall; sensori-motor skills; spatial abilities

Musical training involves the acquisition and use of complex sensory and motor skills. Playing a musical instrument requires musicians to translate musical symbols into specific motor actions while also simultaneously monitoring the auditory signals produced by their

instrument. Musicians develop strong associations between visual musical symbols, motor commands, auditory signals, and temporal patterns. Musical training and experience have been found to be associated with enhanced motor (Zafranas, 2004) and visual-motor skills (Brochard, Dufour, & Despres, 2004; Patston, Corballis, Hogg, & Tippett, 2006), as well as enhanced memory for auditory-motor sequences (Tierney, Bergeson, & Pisoni, 2008). Moreover, the visual and auditory signals that musicians regularly process, as well as their fine motor movements, are all temporal and sequential in nature. Thus, years of formal training and experience with sequential information and a coupling of sensory and motor systems could affect the development of musicians' basic sequence learning abilities.

The current study set out to assess musicians' visual-spatial abilities and, specifically, to explore their visual-spatial sequence learning and memory skills. In the present study, we used an implicit visual-spatial sequence memory and learning task that made use of a finite-state "grammar" to generate visual-spatial sequences (Conway, Bauernschmidt, Huang, & Pisoni, 2010). Implicit statistical learning tasks involve the unconscious acquisition of information (Seger, 1994). These behavioral tasks are used to assess how participants make use of transitional probabilities, which involve the probability of event A given event B (Conway & Christiansen, 2006). A large number of studies have been carried out on implicit statistical learning in nonmusicians (Aslin, Saffran, & Newport, 1998; Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). In the nonmusician literature, the ability to detect and use statistical probabilities encoded in temporal patterns has been shown in tactile perception (Conway & Christiansen, 2005), motor sequencing abilities (Hunt & Aslin, 2001), perception of non-linguistic patterns, such as tones (Saffran, Johnson, Aslin, & Newport, 1999), as well as speech perception (Conway, Pisoni, et al., 2011; Saffran, Newport, & Aslin, 1996).

Impairments in speech perception have been associated with poor implicit statistical learning (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). Conway, Pisoni, et al. (2011) found that congenitally deaf children who use cochlear implants exhibit deficits in implicit statistical learning of visual-spatial sequences when compared to typically developing peers. Moreover, these deficient sequencing skills were associated with poor speech perception. The authors argue that the deficits in visual-spatial sequencing abilities found in these children may stem from the period of auditory deprivation that they experienced early in life, specifically the period of time prior to implantation with the cochlear implant. As was stated earlier, sound is temporal and sequential in nature. Conway and colleagues argue that a period of early auditory deprivation (i.e., a lack of processing temporal and sequential information) can lead to disturbances in domain general sequencing abilities (Conway, Karpicke, et al., 2011; Conway, Pisoni, & Kronenberger, 2009).

Theories of embodied cognition argue for associations between the body, sensory-systems, and cognition, where sensory-motor developmental experiences can affect cognitive processes (Conway et al., 2009; Smith & Gasser, 2005). Congenitally deaf children who use cochlear implants and long-term trained musicians each have unique sensory histories. While deaf children will have experienced some period of auditory deprivation, long-term musicianship will have likely involved the active use of complex sensory and motor skills. The contrast in sensory experiences and cognitive abilities for these groups can be seen in

their speech perception abilities. While some deaf children display poor speech perception abilities (Nicholas & Geers, 2007; Niparko et al., 2010), a growing body of literature shows that musicians have improved speech perception skills (Fuller, Galvin, Maat, Free, & Baskent, 2014; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011; Soncini & Costa, 2006; Strait & Kraus, 2011). Given musicians' long-term experience with sensory-motor tasks and relations between speech perception and implicit statistical sequencing abilities, it is very likely that implicit statistical learning abilities may be altered in this population. While only a handful of studies have examined visual and/or auditory implicit statistical learning in musicians (Francois & Schon, 2011; Loui, Wessel, & Hudson Kam, 2010; Romano Bergstrom, Howard, & Howard, 2012), these earlier studies have yielded mixed results, with some research suggesting that musicians have enhanced implicit sequence learning abilities where other research has shown no differences in these abilities between musicians and nonmusicians.

There have also been studies showing how nonmusicians and musicians use implicit memory to acquire knowledge of language and music structure (for review see Ettliger, Margulis, & Wong, 2011; Loui, 2012). However, little is currently known about musicians' implicit sequence learning abilities and how they make use of transitional probabilities encoded in these patterns. Loui et al. (2010) conducted the first investigation of auditory sequence learning in musicians. The authors developed a new musical scale and generated novel melodies according to two underlying grammars (Grammar 1 and Grammar 2). Each melody contained eight pure tone notes. Musician and nonmusician participants listened to 400 melodies during a 30-minute passive exposure phase. Half of the participants in each group listened to melodies generated from Grammar 1 whereas the remaining participants listened to melodies that were generated from Grammar 2. During the testing phase, participants heard two melodies, one generated from each grammar, and had to identify which melody sounded more familiar. Results showed no differences between groups in melody recognition. These results suggest that formally trained musicians and nonmusicians show similarities in implicit learning of musical structure.

Additional research by Francois and Schön (2011) also failed to show any behavioral differences between musicians and nonmusicians in implicit learning. In their study, musicians and nonmusicians listened to 5.5 minutes of a continuous stream of an artificial sung language. The language consisted of 11 syllables, where each syllable was associated with a specific tone. Syllables were combined to form five trisyllabic sung words. Transitional probabilities of syllable pairings varied and could be found within and across word boundaries. After listening to this artificial language, participants were presented with a linguistic and a musical test. For the linguistic test, participants heard two words and had to select which word most closely resembled those in the sung language. In the music test, participants heard two brief melodies that were played on a piano and participants again selected the stimulus that best matched the language. EEG recordings were also gathered from participants during both testing phases.

The behavioral results showed no performance differences in accuracy between the musician and nonmusician participants for either of the linguistic and musical tests.

Electrophysiological data compared participants' reactions to the presentation of familiar

and unfamiliar stimuli during the testing phases. Group differences were found in the ERP data. Musicians had larger N1 responses than nonmusicians for the unfamiliar stimuli. This result was found in both the linguistic and musical testing phases. While no differences were found in the behavioral data, Francois and Schön concluded that the results of their study showed that musicians were able to learn the linguistic and the musical structure of the artificial language better than the nonmusician participants.

Unlike Loui et al. (2010) and Francois and Schön (2011), Romano Bergstrom et al. (2012) found behavioral differences between musicians and nonmusicians, with musicians showing better implicit statistical learning in comparison to nonmusician controls. In their study, participants completed the Alternating Serial Reaction Time Task (Nissen & Bullemer, 1987). In the task, participants saw four circles on a computer screen where each became individually illuminated and participants were required to make a corresponding finger press on a keyboard. Once the participant made his response, another circle became illuminated and again the participant made a corresponding button press. Embedded within the stimulus sequences were three-item high frequency and low frequency patterns, which alternated between random items. It was found that musicians had faster reaction times than controls when responding to high frequency patterns in comparison to low frequency patterns. The authors also reported differences in accuracy between groups; however, it was not clear how accuracy was measured in the task.

Overall, the nature of musicians' implicit statistical learning remains unclear. Loui et al. (2010) and Francois and Schön (2011) both failed to show any behavioral differences between musicians and nonmusicians on statistical learning tasks. In contrast, Romano Bergstrom et al. (2012) did find group differences with musicians showing enhanced statistical learning abilities. The mixed results from these studies may be the result of methodological differences. Loui et al. (2010) as well as Francois and Schön (2011) tested participants in one modality, in which participants were presented with auditory stimuli and then asked to discriminate between later presented auditory stimuli. In contrast, Romano Bergstrom et al. used a task that required participants to rely on sensory-motor abilities, specifically visual and fine-motor skills. The present study was carried out to assess relations between long-term formal musical experience and implicit visual-spatial sequence learning skills. Unlike the Romano Bergstrom et al. (2012) study which utilized a restricted set of pre-generated triplet patterns, participants in the current study were presented with visual-spatial sequences that contained transitional probabilities generated by an underlying finite-state grammar. We examined how well participants were able to learn the sequences and make use of the statistical regularities within the visualspatial sequences. Additionally, we explored whether individual differences in nonverbal intelligence, vocabulary knowledge, and short-term memory capacity could account for participant's ability to learn visual-spatial sequences. We hypothesized that long-term formal musical training and experience would be associated with enhanced visual-spatial abilities and that highly trained musicians would perform better than nonmusicians, individuals with little or no musical training experience, on the implicit visual-spatial sequence learning and memory tasks.

Method

Participants

In total, 48 participants between the ages of 18 and 30 years were recruited for this study. All participants completed the informed consent that was approved by Indiana University's Institutional Review Board. All participants were monolingual speakers of American English. Twenty-four participants (13 females and 11 males) were highly trained musicians who were recruited from Indiana University's Jacob School of Music. All musicians played either piano or organ. The musicians began their musical training at or before the age of nine ($M = 5.79$ years of age, $SD = 1.86$), had on average 17.33 years of musical training ($SD = 3.67$), and continued to practice their instrument regularly (hours of practice per week, $M = 18.12$, $SD = 11.85$). The 24 participants (14 females and 10 males) who were classified as nonmusicians had little or no musical training experience (experience playing in years, $M = 2.06$, $SD = 1.99$). All nonmusicians reported no longer playing a music instrument at time of testing. Nonmusician participants were recruited through the use of flyers that were posted around the Indiana University Bloomington campus.

Materials and procedures

All tasks, with the exception of Matrix Reasoning and the Peabody Picture Vocabulary Test, were conducted in a sound-attenuated IAC booth in the Speech Research Laboratory at Indiana University Bloomington. For the auditory tasks, stimuli were presented over high-quality headphones (Beyerdynamic DT109). All participants completed a pure-tone hearing test. With the exception of three musicians, all participants exhibited normal hearing (≤ 20 decibels hearing level [dB HL] pure tone thresholds from 250 to 8000 Hz). Two of the three musicians with atypical hearing had high-frequency hearing loss at 8000 Hz that was localized to one ear, 35 dB HL left ear for one participant and 65 dB HL right ear for the other participant, but exhibited otherwise normal hearing for the remaining frequencies. The third musician had atypical hearing in the right ear with 40 dB HL at 4000 Hz and 90 dB HL at 8000 Hz, but exhibited normal hearing for the remaining frequencies. Because these three musicians had hearing loss restricted to high frequencies only, these participants were retained in the study.

Measures

Matrix reasoning.—The Matrix Reasoning subtest of the WASI II (Wechsler, 2011) was administered to obtain a normed baseline measure of global nonverbal intelligence. The Matrix Reasoning task is commonly used to assess nonverbal abstract problem solving abilities. Participants were shown an array of visual images with one missing square. The participant was required to complete, or fill in, the missing portion of the abstract patterns by selecting an image that best fit the array from five picture options. The task was terminated when participants were unable to identify the correct pattern in four consecutive trials. A T score was calculated for each participant based on his/her raw score (the number of correctly completed patterns).

Peabody Picture Vocabulary Test IV (PPVT).—The PPVT was administered to all participants to assess age-appropriate receptive vocabulary levels. In this task, the examiner

said a word out loud while participants viewed four pictures. Participants were required to select the image that best depicted the stimulus word. PPVT is a well-known standardized measure of vocabulary knowledge for ages 2.6–90. Stimulus words are divided into numbered sections with each section containing 12 words. Words within each section become more difficult as the test progresses. A baseline vocabulary level is established when a participant scores one error or less in a section. Testing ceases when a participant has made eight or more errors in a section. A raw score of each participant's vocabulary level was calculated by taking the number of the highest word correctly answered and subtracting the number of errors made. The raw score was then converted into a standard score, which was obtained from the norms in the PPVT manual (Dunn & Dunn, 2007).

Digit span.—A modified version of the WMS III forward digit span subtest was administered to all participants in order to measure short-term auditory-verbal memory (Wechsler, 1997). Lists of spoken digits were played through headphones at a rate of 1 second per digit. List lengths started at two digits per list and increased by one digit with the longest list length containing 10 digits. The task was terminated when participants failed to correctly repeat back both of the lists at the same length. Participants were required to repeat the digits in the exact order in which the stimuli were presented. Responses were scored by recording the longest sequence length that a participant could correctly recall.

Visual-spatial sequence learning and memory.—A sequence learning and memory task based on Milton and Bradley's Simon memory game was used to assess procedural memory, specifically, participants' ability to implicitly learn sequences of visually-presented patterns (Cleary, Pisoni, & Geers, 2001; Karpicke & Pisoni, 2004; Pisoni & Cleary, 2004). In the current version of the task, participants were shown four black squares on a touch-screen monitor. The squares were individually illuminated to form a sequence. Participants were asked to observe a sequence. Once a sequence ended, all four black squares reappeared on the screen and participants were required to reproduce the sequence by pressing the appropriate locations on the touch screen in the order that was shown on the display. Participants were asked to use their dominant hand when making their responses. No feedback was given. The visual-spatial sequences used in this task were generated according to a set of finite-state "grammars" which specified the order of the sequence elements (Conway et al., 2011; Karpicke & Pisoni, 2004). The grammars specified the probability of a stimulus appearing in a specific location given the preceding location. A display of the visual-spatial sequence task is shown in Figure 1.

The sequence learning task consisted of two phases, a learning phase and a test phase. Each phase contained 32 sequences. The sequences were divided into four blocks with each block consisting of two exposures to set sizes of 5, 6, 7, and 8 items. Within a block, set sizes were randomly presented to each participant. During the initial learning phase, participants were exposed to grammatical sequences generated by only one grammar, Grammar T (trained), whereas in the testing phase participants were shown novel test sequences generated by two grammars, Grammar T and Grammar U (untrained). Half of the test sequences followed Grammar T, which was used in the learning phase, and the remaining 16 sequences were generated from Grammar U, which was an unfamiliar grammar that was different from

Grammar T. The artificial grammars used to generate the visual-spatial sequences are shown in Figure 2. Presentation of novel temporal sequences from the trained grammar used in the learning phase (Grammar T) and novel sequences from Grammar U were randomized during the testing phase. Participants were not told about the underlying grammars used to create the visual patterns and were also unaware of the separate phases because the task seamlessly transitioned from the learning phase to the testing phase without the participants' knowledge.

A Magic Touch™ touch-sensitive CRT monitor was used to display the visual-spatial sequences and record participants' responses. During sequence presentations, stimuli were displayed for 700 ms with an inter-stimulus-interval of 500 ms. The four response squares appeared on the screen 500 ms after the end of a sequence presentation. Sequence presentations started 3 s following a participant's final response (pressing the "continue" button on the screen). Participants' implicit statistical sequence learning was assessed by examining changes in their memory span for visual-spatial sequences from the trained grammar (Grammar T) compared to the visual-spatial sequences from the novel untrained grammar (Grammar U).

Results

An initial MANOVA was conducted in order to examine differences between musicians and nonmusicians in terms of age, years of education, nonverbal reasoning, vocabulary, and short-term memory. A significant group difference was found, Wilks' Lambda = .53, $F(1, 46) = 7.44$, $p < .001$. Subsequent ANOVAs revealed significant group differences in age, $F(1, 46) = 9.7$, $p = .003$, and years of education, $F(1, 46) = 13.38$, $p = .001$. Musicians ($M = 23.41$, $SD = 3.76$) were older than the nonmusicians ($M = 20.75$, $SD = 1.84$) and had more years of education (musicians: $M = 17.56$, $SD = 3.14$; nonmusicians: $M = 14.83$, $SD = 1.40$). Group differences were also found for nonverbal Matrix Reasoning, $F(1, 46) = 9.75$, $p = .003$, and PPVT, $F(1, 46) = 16.07$, $p < .001$. Musicians performed better than nonmusicians on both measures. Descriptive statistics for Matrix Reasoning, PPVT, and Digit Span are listed in Table 1. The difference in Digit Span approached significance, $F(1, 46) = 3.34$, $p = .074$. Given these group differences, age, years of education, nonverbal reasoning, vocabulary knowledge, and short-term memory capacity were entered as covariates in concluding, cumulative analyses.

Visual-spatial sequence learning and memory

A sequence was scored as correct if the participant was able to reproduce the entire sequence without error. Scores obtained during the learning and testing phase were calculated using a weighted span method where the total number of correct sequences was multiplied by the set size length, and then all scores were summed (Conway et al., 2010). For example, if a participant correctly reproduced 4 sequences at length 5, 4 at length 6, 3 at length 7, and 2 at length 8, then his score would be computed as $(4 \times 5) + (4 \times 6) + (3 \times 7) + (2 \times 8) = 81$. Percent correct scores were then calculated where 100% equaled a weighted score of 104 in the testing phase for both the trained and untrained sequences. Descriptive statistics from the visual-spatial sequence learning task are listed in Table 1.

In examining performance on the visual-spatial sequencing task, we first assessed performance in the learning phase. A 2 (group) \times 4 (block) ANCOVA, with age and years of education used as covariates, revealed a main effect of group, $F(1, 1526) = 74.29, p < .001$, and a main effect of block, $F(3, 1526) = 6.0, p < .001$. No interaction was found between group and block, $F < 1$. Musicians were better at reproducing sequences in the initial learning phase when compared to nonmusicians. We then used one-way ANOVAs to examine group differences across the four presentation blocks. Musicians performed significantly better than the nonmusicians on all four blocks: Block 1, $F(1, 44) = 8.62, p = .005$; Block 2, $F(1, 44) = 9.74, p = .003$; Block 3, $F(1, 44) = 8.02, p = .007$; Block 4, $F(1, 44) = 13.01, p = .001$. Figure 3 displays musicians' and nonmusicians' performance during the learning phase.

Analyses were then conducted to examine participants' performance in the testing phase. A 2 (group) \times 4 (block) \times 2 (grammar) ANCOVA, with age and years of education used as covariates, revealed a main effect of group, $F(1, 1518) = 107.42, p < .001$, a main effect of block, $F(3, 1518) = 48.65, p < .001$, and a main effect of grammar, $F(1, 1518) = 35.34, p < .001$. While no significant interactions were found, the group \times block interaction approached significance, $F(3, 1518) = 2.17, p = .089$. Subsequent univariate analyses revealed significant group differences for trained visual-spatial T sequences, $F(1, 44) = 13.41, p = .001$, as well as for untrained visuospatial U sequences, $F(1, 44) = 12.76, p = .001$. Musicians were significantly better than nonmusicians at reproducing sequences from the trained (T) and untrained (U) grammars. Figure 4 displays musicians' and nonmusicians' performance during the testing phase.

We then calculated an implicit visual-spatial sequence learning score for each participant. This was done in order to examine whether there were differences in performance between musicians and nonmusicians in implicitly learning sequences that were generated from the trained grammar when compared to sequences that were generated from the untrained grammar. The implicit learning score was calculated by subtracting participants' untrained scores from their trained scores for the testing phase. Musicians showed an 18.27% implicit learning score (79.2% accuracy for trained grammar sequences vs. 60.93% accuracy for untrained grammar sequences). Nonmusicians showed a 14.35% implicit learning score (50.68% accuracy for trained grammar sequences vs. 36.33% accuracy for untrained grammar sequences). A one-way ANCOVA with age and years of education entered as covariates revealed no difference between musicians and nonmusicians for implicit statistical learning, $F < 1$. In other words, musicians and nonmusicians showed no difference in their ability to implicitly learn sequences that were generated from a grammar that they had been trained on in comparison to sequences that had been generated from a novel, untrained grammar. Although no group differences were found in implicit statistical learning, significant differences were observed in overall sequence learning abilities. Musicians showed better visual-spatial sequence learning and memory skills when compared to nonmusicians.

Relations among measures

Partial correlations (controlling for age and years of education) were carried out to examine relations among the other dependent measures. A correlation matrix of these results collapsing over both groups is shown in Table 2. In examining the collapsed group correlations, visuospatial sequence learning was correlated with years of musical experience. Relations between visual-spatial sequence learning of the trained T and untrained U grammars and years of musical experience are displayed in Figure 5. Years of playing a musical instrument was correlated with both the trained visual-spatial T ($r = .505$, $p < .001$) and the untrained visual-spatial U ($r = .501$, $p < .001$) scores. The more musical experience an individual had, the better his or her visual-spatial sequencing abilities. Short-term memory capacity was also correlated with both the trained visual-spatial T ($r = .322$, $p = .029$) and the untrained visual-spatial U ($r = .302$, $p = .042$) scores. Individuals with better short-term memory capacity exhibited better performance in reproducing visual-spatial sequences.

Effects of covariates

While the Visual-Spatial Sequence Learning Task was designed to assess visual-spatial sequence learning and memory, it could be argued that successful performance on this task was constrained by short-term memory abilities. During the task, the set size lengths of the to-be-learned sequences were between five and eight items. The group difference found between musicians and nonmusicians on the Visual-Spatial Sequence Learning Task could be the result of more basic elementary differences in short-term memory capacity. As stated earlier, the group difference in short-term memory capacity between musicians and nonmusicians approached significance ($p = .074$). It is possible that musicians' enhanced visual-spatial sequence learning abilities could be attributed to their better short-term memory capacity. Furthermore, group differences were also found in nonverbal reasoning and receptive vocabulary, where musicians exhibited better performance in both. These measures have been used as proxy variables of general intelligence. It is possible that the difference found in visual-spatial sequencing between musicians and nonmusicians may also stem from group differences in general intelligence. While it is unclear precisely how general intelligence would support sequencing abilities (Tierney, Bergeson, & Pisoni, 2009), there has been some criticism of this earlier research for the failure to control for differences in full-scale intelligence in examining differences between musicians and nonmusicians (Schellenberg, 2008).

To deal with these issues, we conducted an ANCOVA to assess the influence of short-term memory capacity, nonverbal reasoning, and vocabulary knowledge on visual-spatial sequence learning. Visual-spatial sequence learning performance was collapsed across all conditions (learning phase and testing phase) to create a composite visual sequence score for each participant. The ANCOVA (covarying age, years of education, Matrix Reasoning, PPVT, and Digit Span), revealed that the group differences found between musicians and nonmusicians still remained even after including short-term memory, nonverbal reasoning, and vocabulary knowledge as covariates, $F(1, 41) = 6.31$, $p = .016$. These results suggest that differences in short-term memory capacity, nonverbal reasoning abilities, and vocabulary

knowledge cannot account for the performance differences obtained between musicians and nonmusicians on the Visual-Spatial Sequence Learning and Memory Task.

In order to further examine the unique effect of musical training on visual-spatial sequencing skills, separate analyses were also conducted treating nonverbal reasoning ability, vocabulary, and short-term memory capacity as dependent variables. Individual univariate analyses were conducted on each variable, and the remaining dependent variables were entered as covariates, including visual-spatial sequencing abilities. Results revealed that previously significant group differences in nonverbal reasoning and vocabulary as well as the marginal significance in short-term memory became non-significant after controlling for the remaining variables (nonverbal $p = .603$; vocabulary $p = .092$; short-term memory $p = .706$). In summary, our results showed that there was a group difference in visual-spatial sequencing skills even after controlling for the other dependent variables. However, the same result was not found using ANCOVAs when nonverbal reasoning, vocabulary, and short-term memory were treated as dependent variables. Overall, this result suggests that there is a unique association between musical training and visual-spatial sequencing skills.

Discussion

In the current study, we set out to examine long-term musicians' implicit learning and memory of sequential visual-spatial patterns. While we did not find any differences in performance between musicians (individuals with long-term formal musical training) and nonmusicians (individuals with little or no musical training experience) in implicit statistical learning of the trained grammar versus the novel untrained grammar, the present results demonstrated robust group differences in overall visual-spatial sequence learning abilities. Experienced musicians were better able to learn and reproduce visual-spatial sequences compared to nonmusicians. Differences in performance were observed early on in the Visual-Spatial Sequence Learning Task where musicians outperformed nonmusicians in the initial blocks of the learning phase. The differences observed in visual-spatial sequence learning and memory between musicians and nonmusicians could not be explained by differences in age, years of education, nonverbal reasoning, vocabulary knowledge, or short-term memory capacity.

Recently, several researchers have suggested that musicians may have stronger and more robust auditory skills that are the result of their long-term experience attending to and processing complex auditory signals (Parbery-Clark et al., 2011; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Strait, Parbery-Clark, Hittner, & Kraus, 2012). Musical training and experience with sound patterns has been linked to better subcortical encoding of auditory signals (Musacchia, Sams, Skoe, & Kraus, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007) and more robust speech perception in noise (Fuller et al., 2014; Parbery-Clark et al., 2009, 2011; Soncini & Costa, 2006; Strait & Kraus, 2011). However, results of the current study on visual-spatial sequence learning and memory challenge these earlier conclusions that musicians' enhanced cognitive abilities are limited to the auditory modality, by showing strong associations between long-term musical experience and enhanced visual-spatial sequence learning abilities. The present results cannot be explained by simple differences in audibility or differences in subcortical auditory encoding abilities, because the stimuli used

in the Visual-Spatial Sequence Learning Task were all visual patterns. Furthermore, to reduce the possibility that participants might use verbal coding or verbal labels, and thus rely on automatized rapid phonological coding skills, we modified the earlier sequence memory task used by Cleary et al. (2001) and Conway et al. (2010) and presented participants with all black stimulus squares. This methodology differs from earlier versions of the sequence learning and memory task used by Conway et al. (2010), which used four different colored squares (e.g., red, blue, yellow, and green). Thus, the current findings, coupled with this previous literature, suggest that long-term musical experience may lead to broad cognitive enhancements in both the auditory and visual modalities.

The differences observed in visual-spatial sequence learning and memory found between musicians and nonmusicians in the current study may reflect underlying differences in visuospatial skills. In the Visual-Spatial Sequence Learning Task, participants were presented with visual-spatial sequences on a touch screen monitor and had to reproduce the sequences by pressing the appropriate locations on the display in the order that was shown. As noted earlier, musicians have extensive long-term experience rapidly translating visual symbols into motor actions, such as reading sheet music and then making corresponding finger movements on their instruments. Similarly, the Visual-Spatial Sequence Learning Task used in this study required participants to monitor and encode sequential visual patterns and then make corresponding hand and fine finger movements to execute their responses. Research conducted by Brochard and colleagues (2004) found that musicians have shorter reaction times during visual imagery tasks, suggesting that musicians are faster than nonmusicians at associating visual stimuli with specific sensory-motor movements and actions.

The findings obtained in the present study taken together with the recent study of Pau, Jahn, Sakreida, Domin, and Lotze (2013) suggest that musicians with long-term formal training and experience have qualitatively different sensory-motor abilities based on their unique enriched developmental histories. The present results are also supported by converging neuroimaging research showing anatomical changes in musicians' cortical motor areas (Elbert, Pantev, Wiendbrunch, Rockstroh, & Taub, 1995; Pantev, Engelien, Candia, & Elbert, 2001; Schlaug, Jancke, Huang, & Steinmetz, 1995). The study by Pau and colleagues (2013) revealed associations between visual sequence learning and functional cortical changes in musicians. Using fMRI, Pau et al. examined changes in the neural processing activities of pianists and nonmusicians as they learned and replicated finger sequences on a piano. Participants were shown visual images of hands with each hand display highlighting a specific to be remembered finger. Each finger had an associated tone. Participants were shown a row of eight hand displays that indicated the to-be-remembered finger sequence. Both the left and the right hand were used in the finger sequence learning task. Participants completed 12 training sequences and 40 test sequences. During the study phase, participants saw the hand displays but were told not to move their fingers. Following the study phase, participants completed a test phase where they reproduced the learned finger sequence. Behavioral results showed that the pianists learned the novel finger sequences better than the naïve nonmusicians. Additional neuroimaging results revealed that musicians recruited more motor areas during the study phase in comparison to the naïve nonmusicians. Pau and colleagues suggested that the musicians' increased activation of motor areas during the study

phase was the result of their long-term musical training experience and activities that involved the repeated association of visual and sensory-motor skills.

While the Pau et al. (2013) study showed that musicians were able to learn visually presented finger sequences better than nonmusicians, this result is not surprising given that the musicians were pianists who already had long-term experience learning complex finger sequences on a keyboard. The visual-spatial sequence learning task used in the current study differed in several important ways from the task Pau et al. used. Our Visual-Spatial Sequence Learning Task consisted of the visual presentation of black squares that were fixed to specific visual-spatial locations on the screen and participants used their dominant hand to make their responses. While our study included highly experienced pianists and organists who were undergraduate and graduate students at the Indiana University School of Music, the visual-spatial sequence learning task that our participants completed did not directly reflect the same skills needed to play their instruments (e.g., learning finger sequences). Furthermore, our visuospatial sequences contained transitional probabilities generated by an underlying grammar. The presentation of one stimulus square influenced the probability of the subsequent stimulus square that was illuminated in a specific location. The present study revealed robust associations between long-term formal musical training and experience and enhanced visual-spatial sequence learning and memory of arbitrary unfamiliar temporal patterns.

Group differences were not found in visual implicit statistical learning between musicians and nonmusicians. Both musicians and nonmusicians were able to learn and reproduce sequences from the trained grammar better than sequences from the untrained grammar. The average learning score reflecting the difference between performance on the trained versus untrained grammar was comparable for both groups. Other studies that examined auditory statistical learning in musicians also failed to find behavioral differences in implicit learning abilities between musicians and nonmusicians (Francois & Schön, 2011; Loui et al., 2010). However, the findings from the present study cannot be used to conclusively establish that musicians have comparable visual-spatial implicit statistical learning skills relative to nonmusicians. Four musicians reached ceiling and scored 100% correct for trained grammar sequences during the testing phase. This result limited their potential learning score because sequence learning and memory was assessed as a difference score between trained and untrained conditions. While the group difference was not significant, it is possible that the addition of larger set sizes may have revealed significant group differences in implicit visual-spatial sequence learning abilities using this methodology.

In addition to the visual-motor component of the Visual-Spatial Sequence Learning Task, musicians may have made use of the statistical regularities encoded within the visual-spatial sequences to complete the task. Using a similar version of the sequence task with randomly generated patterns that contained no statistical regularities, Tierney and colleagues (2008) found no differences in performance in sequence memory between musicians and several control groups of nonmusicians when the visual-spatial sequences were randomly presented colored stimulus squares. Furthermore, additional research found that musicians respond faster than nonmusicians to high-frequency patterns in comparison to low-frequency patterns (Romano Bergstrom et al., 2012). These earlier results taken together with the findings of

the current study suggest that trained musicians are able to utilize statistical regularities encoded within temporal patterns.

While we suspect that musicians' superior visual-spatial sequencing abilities largely stems from their unique developmental experiences, this study is correlational in nature and we are unable to draw direct causal inferences regarding musical training and sequencing skills. While some research has shown causal relations between musical training and enhanced cognitive abilities (Norton et al., 2005; Schellenberg, 2004), other research found innate differences in pitch recognition between musicians and nonmusicians (Drayna, Manichaikul, de Lange, Snieder, & Spector, 2001; Ukkola, Onkamo, Raijas, Karma, & Jarvela, 2009). Future research should make use of longitudinal studies to help parse out innate and experiential differences between musicians and nonmusicians. Additional research should continue to examine the sequence learning abilities of musicians, both in the auditory and visual modalities, and should also explore the implicit sequencing abilities of vocalists, who have different sensory-motor experiences relative to musicians who play a musical instrument, like a piano keyboard.

Conclusions

The present study examined associations between long-term musical training and implicit visual sequence learning and memory. No differences were found in visual-spatial implicit statistical learning between musicians (individuals with long-term formal musical training) and nonmusicians (individuals with little or no musical training experience). However, robust differences were found in overall visual-spatial sequence learning abilities. Musicians displayed enhanced visual-spatial sequence learning and they consistently outperformed nonmusicians in both phases of the Visual-Spatial Sequence Learning Task. The differences in visual-spatial sequence learning and memory found between musicians and nonmusicians could not be accounted for by differences in age, years of education, nonverbal reasoning, verbal knowledge, or short-term memory capacity. We suggest that musicians' enhanced visual-spatial sequence learning may reflect underlying differences in visual-spatial and motor skills that may result from highly enriched early experience associated with formal training, experience, and activities playing a keyboard instrument for many years. Musicians' long-term experience and formal training involve an inseparable coupling of visual-spatial and sensory-motor systems used in the encoding, storage, retrieval, and processing of temporal sequential patterns regardless of the sensory modality of the input signals.

Acknowledgments

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by NIH-NIDCD: 1R01 DC000064 and T32 DC00012.

References

- Aslin RN, Saffran JR, & Newport EL (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324.

- Brochard R, Dufour A, & Despres O (2004). Effects of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, 54(2), 103–109. [PubMed: 14980450]
- Cleary M, Pisoni DB, & Geers AE (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear & Hearing*, 22, 395–411. [PubMed: 11605947]
- Conway CM, Bauernschmidt A, Huang SS, & Pisoni DB (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371. [PubMed: 19922909]
- Conway CM, & Christiansen MH (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 24–39.
- Conway CM, & Christiansen MH (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science*, 17(10), 905–912. [PubMed: 17100792]
- Conway CM, Karpicke J, Anaya EM, Henning SC, Kronenberger WG, & Pisoni DB (2011). Nonverbal conditioning deaf children following cochlear implantation: Motor sequencing disturbances mediate language delays. *Developmental Neuropsychology*, 36(2), 237–254. [PubMed: 21347923]
- Conway CM, Pisoni DB, Anaya EM, Karpicke J, & Henning SC (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*, 14(1), 69–82. [PubMed: 21159089]
- Conway CM, Pisoni DB, & Kronenberger WG (2009). The importance of sound for cognitive sequencing abilities: The auditory scaffolding hypothesis. *Current Directions in Psychological Science*, 18(5), 275–276. [PubMed: 20725604]
- Drayna D, Manichaikul A, de Lange M, Snieder H, & Spector T (2001). Genetic correlates of musical pitch recognition in humans. *Science*, 291(9), 1969–1972. [PubMed: 11239158]
- Dunn LM, & Dunn DM (2007). *Peabody picture vocabulary test* (4th ed). Minneapolis, MN: NCS Pearson.
- Elbert TC, Pantev C, Wiendbrunch C, Rockstroh B, & Taub B (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, 270(5234), 305–307. [PubMed: 7569982]
- Ettlinger M, Margulis EH, & Wong PCM (2011). Implicit memory in music and language. *Frontiers in Psychology*, 2, 1–10. [PubMed: 21713130]
- Fiser J, & Aslin RN (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458–467.
- Francois C, & Schön D (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex*, 21(10), 2357–2365. [PubMed: 21383236]
- Fuller CD, Galvin JJ, Maat B, Free RH, & Baskent D (2014) The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, 8, 179. [PubMed: 25071428]
- Hunt RH, & Aslin RN (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual listeners. *Journal of Experimental Psychology: General*, 130(4), 658–680. [PubMed: 11757874]
- Karpicke JD, & Pisoni DB (2004). Using immediate memory span to measure implicit learning. *Memory & Cognition*, 32(6), 956–964. [PubMed: 15673183]
- Kirkham NZ, Slemmer JA, & Johnson SP (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, 83(2), B35–B42. [PubMed: 11869728]
- Loui P (2012). Statistical learning – What can music tell us? In Rebuschat P & Williams J (Eds.), *Statistical learning and language acquisition* (pp. 433–462). Berlin, Germany: Mouton de Gruyter.
- Loui P, Wessel DL, & Hudson Kam CL (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, 27(5), 377–388. [PubMed: 20740059]
- Musacchia G, Sams M, Skoe E, & Kraus N (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40), 15894–15898.

- Nicholas JG, & Geers AE (2007). Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe-profound hearing loss. *Journal of Speech, Language, and Hearing Research*, 50(4), 1048–1062.
- Niparko JK, Tobey EA, Thai DJ, Eisenberg LS, Wang NY, Quittner AL, & Fink NE (2010). Spoken language development in children following cochlear implantation. *Journal of the American Medical Association*, 303(15), 1498–1506. [PubMed: 20407059]
- Nissen MJ, & Bullemer P (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1–32.
- Norton A, Winner E, Cronin K, Overy K, Lee DJ, & Schlaug G (2005). Are there pre-existing neural, cognitive, or motoric markers for musical abilities? *Brain and Cognition*, 59, 124–134. [PubMed: 16054741]
- Pantev C, Engelien A, Candia V, & Elbert TC (2001). Representational cortex in musicians: Plastic alterations in response to musical practice. *Annals of the New York Academy of Sciences*, 930, 300–314. [PubMed: 11458837]
- Parbery-Clark A, Skoe E, Lam C, & Kraus N (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, 30, 653–661. [PubMed: 19734788]
- Parbery-Clark A, Strait D, Anderson S, Hittner E, & Kraus N (2011). Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS ONE*, 6(5), e18082. [PubMed: 21589653]
- Patston LLM, Corballis MC, Hogg SL, & Tippett LJ (2006). The neglect of musicians: Line bisection reveals an opposite bias. *Psychological Science*, 17, 1029–1031. [PubMed: 17201783]
- Pau S, Jahn G, Sakreida K, Domin M, & Lotze M (2013). Encoding and recall of finger sequences in experienced pianists compared with musically naïve controls: A combined behavioral and functional imaging study. *Neuroimage*, 64(0), 379–387. [PubMed: 22982586]
- Pisoni DB, & Cleary M (2004). Learning, memory, and cognitive processes in deaf children following cochlear implantation In Zeng FG, Popper AN & Fay RR (Eds.), *Springer handbook of auditory research: Auditory prosthesis* (pp. 377–426). New York, NY: Springer-Verlag.
- Romano Bergstrom JC, Howard JH, & Howard DV (2012). Enhanced implicit sequence learning in college-age video game players and musicians. *Applied Cognitive Psychology*, 26, 91–96.
- Saffran JR, Aslin RN, & Newport EL (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928. [PubMed: 8943209]
- Saffran JR, Johnson EK, Aslin RN, & Newport EL (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27–52. [PubMed: 10193055]
- Saffran JR, Newport EL, & Aslin RN (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606–621.
- Schellenberg EG (2004). Music lessons enhance IQ. *Psychological Science*, 15(8), 511–514. [PubMed: 15270994]
- Schellenberg EG (2008). Commentary on “Effects of early musical experience on auditory sequence memory” by Adam Tierney, Tonya Bergeson, and David Pisoni. *Empirical Musicology Review*, 3(4), 205–207.
- Schlaug G, Jancke L, Huang YX, & Steinmetz H (1995). In-vivo evidence of structural brain asymmetry in musicians. *Science*, 267(5198), 699–701. [PubMed: 7839149]
- Seger CA (1994). Implicit learning [Review]. *Psychological Bulletin*, 115(2), 163–196. [PubMed: 8165269]
- Smith L, & Gasser M (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life*, 11, 13–29. [PubMed: 15811218]
- Soncini F, & Costa MJ (2006). The effect of musical practice on speech recognition in quiet and noisy situations. *Pro-Fono Revista de Atualizacao Cientifica*, 18(2), 161–170. [PubMed: 16927621]
- Strait D, & Kraus N (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Auditory Cognitive Neuroscience*, 2, 1–10.
- Strait D, Kraus N, Parbery-Clark A, & Ashley R (2010). Muscial experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 261, 22–29. [PubMed: 20018234]

- Strait D, Parbery-Clark A, Hittner E, & Kraus N (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, 123, 191–201. [PubMed: 23102977]
- Tierney AT, Bergeson TR, & Pisoni DB (2008). Effects of early musical experience on auditory sequence memory. *Empirical Musicology Review*, 3(4), 178–186. [PubMed: 21394231]
- Tierney AT, Bergeson TR, & Pisoni DB (2009). General intelligence and modality-specific differences in performance: A response to Schellenberg (2008). *Empirical Musicology Review*, 4(1), 37–39.
- Ukkola LT, Onkamo P, Rajas P, Karma K, & Jarvela I (2009). Musical aptitude is associated with AVPR1A-Haplotypes. *PLoS ONE*, 4(5), e5534. [PubMed: 19461995]
- Wechsler D (1997). Wechsler memory scale third edition (WMS III): San Antonio, TX: Harcourt Assessment.
- Wechsler D (2011). Wechsler abbreviated scale of intelligence-second edition. Bloomington, MN: Pearson.
- Wong PCM, Skoe E, Russo NM, Dees T, & Kraus N (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420–422. [PubMed: 17351633]
- Zafranias N (2004). Piano keyboard training and the spatial-temporal development of young children attending kindergarten in Greece. *Early Child Development and Care*, 174, 199–211.

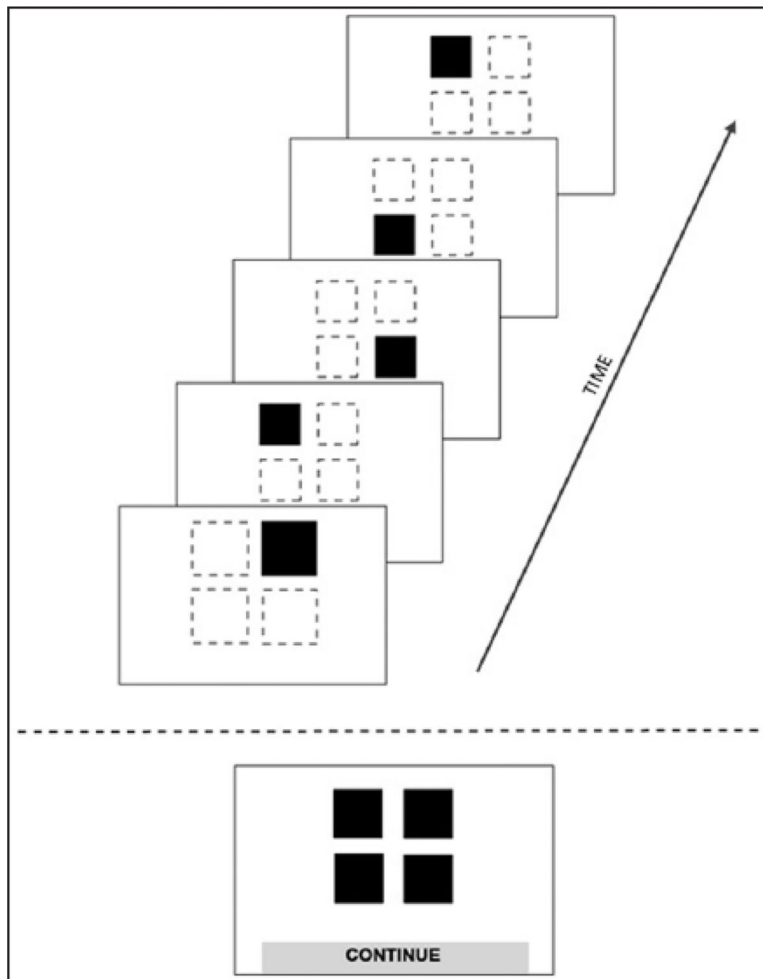


Figure 1. Display of Visual-Spatial Sequence Learning and Memory Task. Squares were individually illuminated to form a sequence. After the full sequence was shown, all four squares reappeared on the screen. Participants reproduced the sequence by touching the squares in the order that they were shown. After reproducing the sequence, participants pressed “continue” to advance to the next sequence.

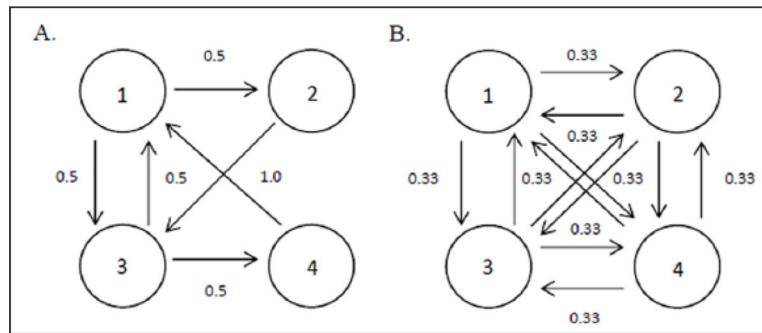


Figure 2. Artificial grammars used in the Visual-Spatial Sequence Learning and Memory Task. Grammar T is displayed in Panel A. Grammar U is displayed in Panel B. Each numbered circle represents the location of a stimulus square from the task. The arrows linking the circles, along with the numbers next to the arrows, indicate the probability of the presentation of one square being followed by the presentation of another square.

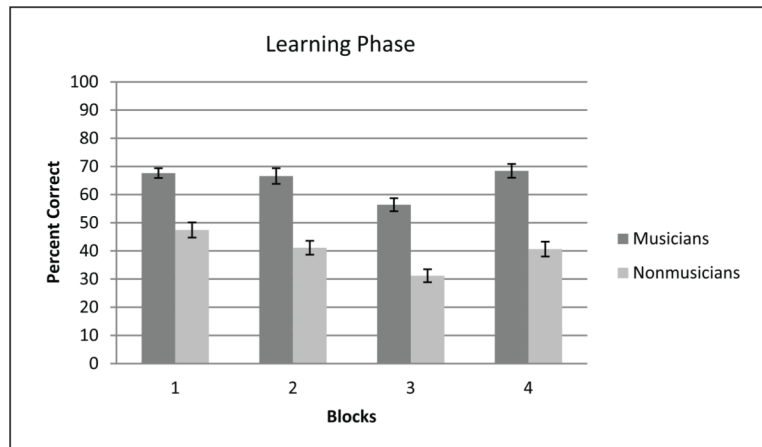


Figure 3. Group differences in visual-spatial sequence learning across blocks in the learning phase. Percentage of correctly reproduced visual sequences is shown on the y axis. Blocks are shown on the x axis. Musicians are indicated in dark gray and nonmusicians are shown in light gray (standard error bars are included).

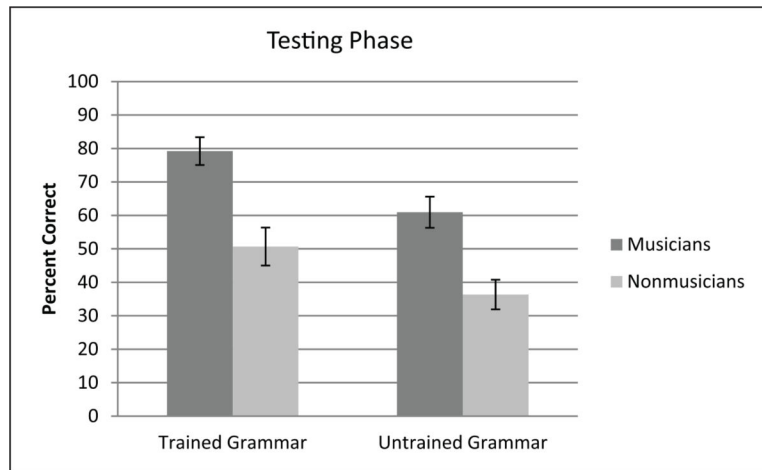


Figure 4. Group differences in visual-spatial sequence learning between musicians and nonmusicians for the trained and untrained grammar during the testing phase. Percentage of correctly reproduced visual sequences is shown on the y axis. The trained grammar (Grammar T) and untrained grammar (Grammar U) are shown on the x axis. Musicians are indicated in dark gray and nonmusicians are shown in light gray (standard error bars are included).

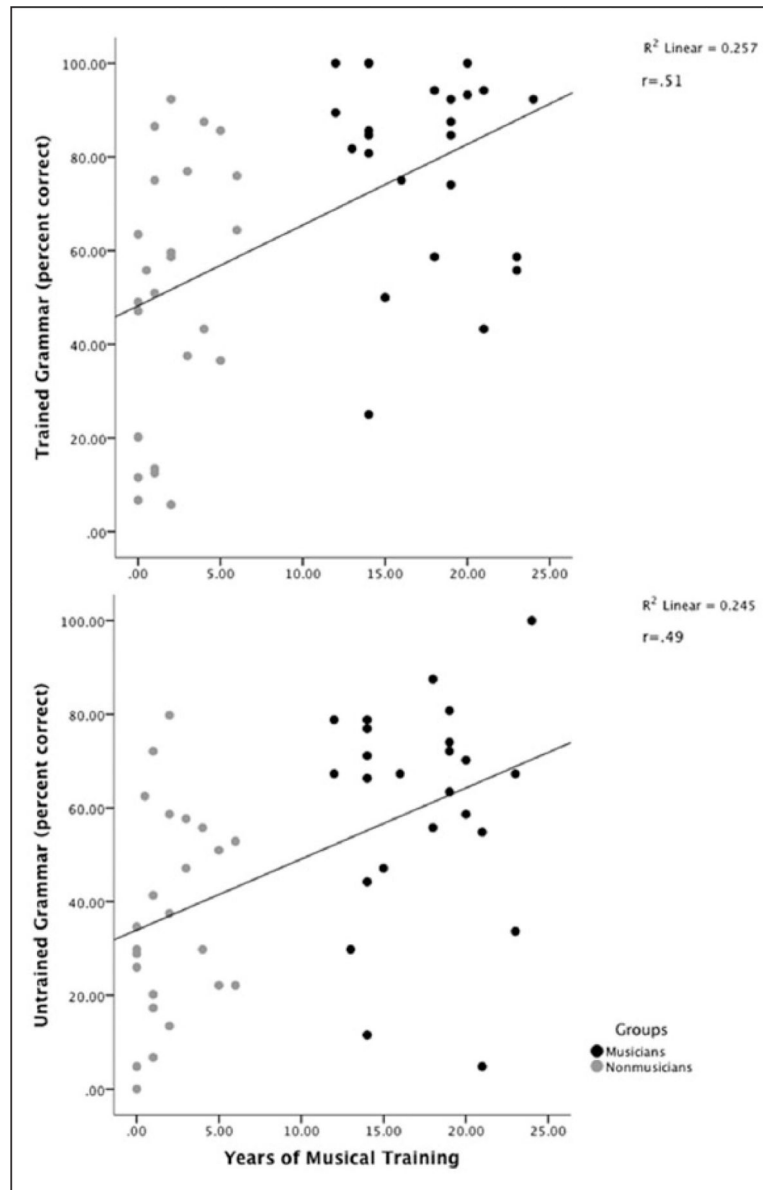


Figure 5.

Relations between visual-spatial sequence learning of the trained and untrained grammar as a function of years of musical experience. The trained grammar is shown in the top panel and the untrained grammar is shown in the bottom panel. Percentage of correctly reproduced visual sequences is shown in the y axis. The number of years of musical training is shown in the x axis. Musicians are indicated in black and nonmusicians in gray.

Table 1.

Musicians vs. nonmusicians group results as shown by univariate analyses.

Measure	Musicians		Nonmusicians		<i>F</i>	<i>P</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Vis-Sp LP	64.76	15.43	40.08	19.08	17.03	< .001
Vis-Sp T	79.2	20.34	50.68	27.76	13.41	.001
Vis-Sp U	60.93	22.85	36.33	21.67	12.76	.001
Vis-Sp L	18.27	16.01	14.35	19.17	.31	.577
Matrix	56.79	4.22	50.58	8.77	9.75	.003
PPVT	120.58	10.37	106.12	14.29	16.07	< .001
Digit span	8.29	1.39	7.54	1.44	3.34	.074

Note. Vis-Sp LP: Visual-Spatial Task Learning Phase sequence percent correct; Vis-Sp T: Visual-Spatial Task trained grammar percent correct; Vis-Sp U: Visual-Spatial Task untrained grammar percent correct; Vis-Sp L: Visual-Spatial Task implicit statistical learning score; Matrix: Matrix Reasoning; PPVT: Peabody Picture Vocabulary Test.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Partial correlations (controlling for age and years of education) between behavioral measures and demographic information (groups collapsed).

	YrPI	Matrix	PPVT	Digit	Vis-Sp T	Vis-Sp U	Vis-Sp L
YrPI	–	–	–	–	–	–	–
Mat	.392**	–	–	–	–	–	–
PPVT	.571***	.589***	–	–	–	–	–
Digit	.139	.07	.229	–	–	–	–
Vis-Sp T	.505***	.416**	.492**	.322*	–	–	–
Vis-Sp U	.501***	.419**	.421**	.302*	.786***	–	–
Vis-Sp L	.081	.056	.175	.077	.456**	–.192	–

Note.

*
 $p < .05$

**
 $p < .01$

 $p < .001$ (two-tailed)

YrPI: Years of musical experience; Mat: Matrix Reasoning; PPVT: Peabody Picture Vocabulary Test; Digit: Digit Span; Vis-Sp T: Visual-Spatial Task trained grammar; Vis-Sp U: Visual-Spatial Task untrained grammar; Vis-Sp L: Visual-Spatial Task implicit learning.