New Music System Reveals Spectral Contribution to Statistical Learning

Psyche Loui

Northeastern University

Abstract

Knowledge of speech and music depends upon the ability to perceive relationships between sounds in order to form a stable mental representation of grammatical structure. Although evidence exists for the learning of musical scale structure from the statistical properties of sound events, little research has been able to observe how specific acoustic features contribute to statistical learning independent of the effects of long-term exposure. Here, using a new musical system, we show that spectral content is an important cue for acquiring musical scale structure. Tone sequences in a novel musical scale were presented to participants in three different timbres that contained spectral cues that were either congruent with the scale structure, incongruent with the scale structure, or did not contain spectral cues (neutral). Participants completed probe-tone ratings before and after a half-hour period of exposure to melodies in the artificial grammar, using timbres that were either congruent, incongruent, or neutral, or a no-exposure control condition. Learning outcome was assessed by correlating pre/post-exposure ratings and the statistical structure of tones within the exposure period. Timbre condition significantly affected learning outcome: participants did best at learning with timbres that were congruent with the scale. Results show that spectral content is a useful cue for statistical learning, and suggest that the frequency structure of sounds might be acquired through exposure to spectral distribution in natural sounds.

Keywords: statistical, learning, sound, timbre, music, language, speech, acoustics

Significance Statement

In the study of how the human brain acquires musical structure, the effects of learning are often confounded with the long-term effects of exposure through one's culture. Here we circumvent this challenge using a new musical scale, and we test the causal role of musical timbre, specifically the spectral distribution of sounds, on the learning of musical scale structure. We show that timbres with spectral energy that is consistent with the mathematical ratios of the scale, are more effective at helping participants learn the structure of the musical scale. Results show the influence of sound acoustics, traditionally a low-level perceptual feature, in the traditionally high-level cognitive capacity of statistical learning.

Implicit learning is a human capacity that is crucial for successful interactions within one's environment, including via language and music. Abundant evidence has shown that humans can learn about the statistical distribution of sounds from passive exposure to sound sequences (e.g. (Bigand, Perruchet, & Boyer, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). The role of implicitly acquired knowledge for musical structure has been robustly demonstrated in multiple behavioral paradigms (e.g. (Krumhansl, 1990; Loui & Wessel, 2007; Tillmann & McAdams, 2004) and others) and with electrophysiological and neuroimaging indices, even among people who have received no explicit musical training (e.g. (Koelsch, Gunter, Friederici, & Schroger, 2000; Landau & D'Esposito, 2006), among others). This robust evidence for musical knowledge without explicit instruction suggests that music can be a valuable model system that provides a window into how the human mind implicitly acquires knowledge from exposure.

In assessing implicitly acquired musical knowledge, a well-replicated behavioral technique is the probe-tone method (Krumhansl & Shepard, 1979), which has been described as a functional hearing test of musical knowledge (Russo, 2009). The probe-tone method involves presenting a musical context (such as a tone sequence) followed by a single tone (i.e. probe tone) to human listeners, who then rate how well the probe tone fits the context. Probe-tone profiles in Western musical scales reflect the frequency of pitch classes in the common-practice Western tonal system, even among untrained listeners (Krumhansl, 1991). This correspondence suggests that some aspect of the implicit knowledge that is reflected in these probe-tone profiles could be acquired from exposure to music in the Western tonal system in the listeners' environment. To test the contribution of exposure to implicit knowledge using the probe-tone method, Castellano, Bharucha, and Krumhansl (1984) showed that after listening to North Indian rags, Western listeners made probe-tone ratings that were consistent with the distribution of tones they encountered during exposure, suggesting that the tonal hierarchy could be conveyed by the statistical distribution of tones. Additionally, Krumhansl et al. (2000) obtained probe-tone profiles for North Sami yoiks (a vocal musical tradition of Northern Scandinavia) from Western, Finnish, and Sami listeners, and showed that Finnish

listeners' probe-tone ratings reflected some familiarity of both Western and Yoik scale structures, again suggesting that probe-tone profiles reflect sensitivity to statistical structures of musical sounds in one's environment.

While these cross-cultural methods provide powerful evidence that knowledge of scale structure is not limited to Western tonal systems, these results could not disentangle between long-term musical knowledge (i.e. knowledge that is acquired from birth and accumulated over the lifespan) and short-term statistical learning that might be operating on a moment-by-moment basis. More generally, the question of how rapidly humans could acquire new knowledge about scale structure is difficult to address using conventional musical systems of any culture. This is because conventional musical systems evolved through complex cultural evolution over time (Cross, 2001; Mithen, 2007); thus they are already overlearned throughout the lifespan as a result of exposure within that culture (to music as well as other sounds such as speech) (Patel & Daniele, 2003). Thus, conventional musical systems cannot capture learning de novo in a way that is not intertwined with culture and cultural evolution.

To circumvent this challenge, several studies have turned to test learning of novel musical systems (e.g. (Creel & Newport, 2002); Leung and Dean (2018)). Loui et al (Loui, 2012; Loui & Wessel, 2008; Loui, Wessel, & Hudson Kam, 2010; Loui, Wu, Wessel, & Knight, 2009) developed a musical system that uses the Bohlen-Pierce (BP) scale, which differs from existing musical scales in important ways. While other musical scales are based on the octave, which is a doubling (2:1 ratio) in frequency, the BP scale is based on the 3:1 frequency ratio (tritave). The equal-tempered Western chromatic scale is based on 12 logarithmically-even divisions of the octave; this enables the selection of certain tones, such as steps 0 (starting point), 4, and 7 along the chromatic scale, that approximate the 3:4:5 integer ratio in frequency, low-integer ratios that lead to consonant sounds that form stable chords in traditional tonal harmony. In contrast, the BP scale divides the tritave into 13 logarithmically-even steps, resulting in steps 0, 6, and 10 being stable tones that approximate a 3:5:7 integer ratio, thus forming a chord in the BP scale (Krumhansl, 1987; Mathews, Pierce, Reeves, & Roberts, 1988). Learning of the statistical structure of the music, which

was operationalized as the increase in sensitivity to event-probabilities of different pitches, was captured by comparing probe-tone ratings before and after exposure to tone sequences in the BP scale (Loui et al, 2010). Statistical sensitivity was assessed by the correlation between probe-tone ratings and the event distribution of tones in the exposure set, and the increase in sensitivity to statistical structure was assessed by the change in correlation between pre-exposure and post-exposure probe-tone ratings.

Importantly, the tone sequences in the previous studies were pure tones ranging in fundamental frequency from 220 Hz to 1320 Hz, with a single acoustic frequency presented for each acoustic event in time. Thus, there was a one-to-one correspondence between spectral information (i.e. acoustic frequency content) and pitch class information. Although pure tones in this range provide a clear percept of pitch, they are not representative of real-world acoustic input because they lack additional acoustic energy along the frequency spectrum (Sethares, 2004). This additional spectral energy provides crucial cues to the listener on many aspects of sound object recognition (Bregman, 1990), including the identity of a musical instrument based on its timbre (Wessel, 1979), the identity of a human speaker based on their voice (Belin, Fecteau, & Bédard, 2004), and the identity of phonemes in speech (Smith, 1951). Spectral information is important for identifying tone sequences even among nonhuman animals such as starlings, who rely on the shape of energy distributed along the frequency spectrum (rather than pitch information per se) to recognize sound patterns (Bregman, Patel, & Gentner, 2016). For periodic sounds, the shape of the spectral distribution is dependent on multiple factors including the frequency of the weighted average of the harmonics (i.e. the spectral centroid), and the spacing between individual frequency components (i.e. spectral fine structure), all of which contribute to the percepts of timbre (Caclin, McAdams, Smith, & Winsberg, 2005; Sethares, 2004; Wessel, 1979).

Despite the importance of spectral information, little is known about the role of spectral information in the statistical learning of scale structure. How does spectral information influence the learning of musical scales? One possibility is that spectral information is orthogonal and unrelated to the event structure of sound sequences. In that case, probe-tone ratings should not differ between different timbres. Another

possibility is that spectral information, as determined here by the spacing between individual harmonics, can be a cue towards the statistical structure of sounds. This latter hypothesis is implicit in one argument for the naturalness of harmony in Western music, which states that the statistical structure of naturally-occurring periodic signals, such as speech, can predict perceived pitch and other musical phenomena such as the relative stability of consonant musical intervals in the chromatic scale (Schwartz, Howe, & Purves, 2003). Support for this comes from findings of similarities between the periodicity of speech sounds and perceived pitch (Schwartz & Purves, 2004) and also in covariations between speech and musical pitch patterns across cultures (Han, Sundararajan, Bowling, Lake, & Purves, 2011). While these findings provide correlative evidence for a relationship between environmental sounds and musical structures, a causal test of this relationship is challenging with naturally-occurring speech sounds and musical structures, because of the aforementioned inherent difficulty in teasing apart listeners' knowledge from their auditory environment. The BP scale again circumvents this challenge: since most listeners have no experience with this new musical scale, it offers an optimal test for the relationship between spectral information and statistical learning.

The hypothesis that spectral information provides cues for the statistical structure of sound sequences makes a strong prediction: that tone sequences in a timbre that is congruent with the musical scale should help in learning the musical scale. Thus, as the BP scale is based on the 3:1 frequency ratio, timbres with harmonics that are spaced apart in 3:1 frequency ratios are congruent with the scale structure, whereas timbres with harmonics that are spaced apart in 2:1 frequency ratios are incongruent with the BP scale structure (despite being congruent with the Western scale structure). By generating Shepard tones, which are complex tones with a fixed spectral centroid but that vary in specific placement of their harmonically related partials (Shepard, 1964), it is possible to manipulate the spacing between harmonics to be congruent or incongruent with the BP scale (see Figure 1), thus directly testing the effect of spectral information on statistical learning.

Here we test the role of spectral content on statistical learning, by comparing learning of the BP scale between timbres that are spectrally congruent and incongruent with the scale. Learning is quantified by improvements in the accuracy of probe-tone ratings as a result of exposure to tone sequences in the BP scale. By comparing learning among participants who heard pure tones (no harmonic partials), tritave-based Shepard tones (complex tones where partials were related to the fundamental in 3:1 multiples in frequency), and octave-based Shepard tones (complex tones with partials related to the fundamental in 2:1 multiples of frequency), and a no-exposure control, we test the role of spectral content in our acquisition of musical knowledge, which informs our ability to learn from sound input more generally.

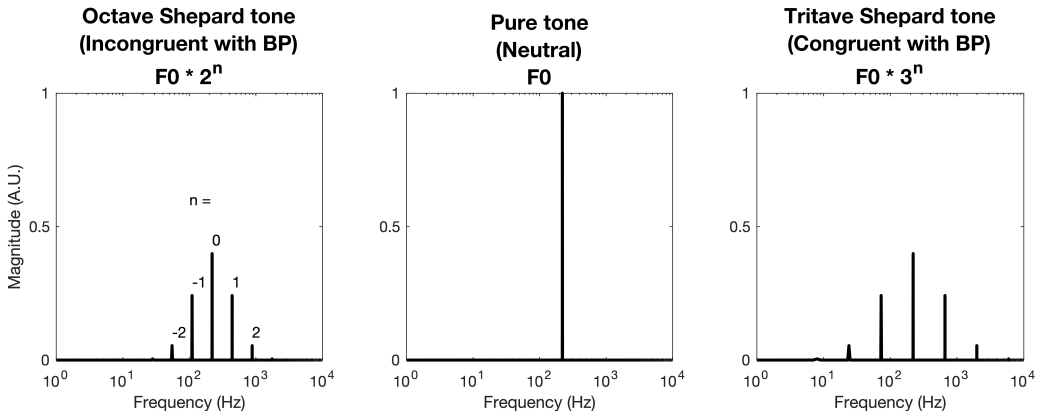

Figure 1. Frequency-domain representations of Tritave-based Shepard tones, Octave-based Shepard tones, and pure tones.

Results

Sensitivity to statistical structure of the scale was quantified by the correlation between probe-tone ratings and exposure. All participant groups except for the no-exposure control group showed significant statistical learning, as quantified by a significant improvement in correlation between exposure and ratings over time (post-exposure vs. pre-exposure) ($F(1,91) = 45.32$, $p < .001$). There was also a significant effect of group, indicating that participants performed differently depending on the timbre ($F(3,91) = 8.8936$, $p < .001$). Furthermore, different timbres resulted in different levels of learning, as supported by a significant group by time interaction ($F(1,91) = 2.6833$, $p = .05$). A direct comparison between ratings in the two

Shepard tone conditions revealed that post-exposure ratings in the Tritave Shepard tone condition were more highly correlated with the exposure profiles than post-exposure ratings in the Octave Shepard tone condition (see figure 3), $t(45) = 2.1120$, $p = 0.04$), suggesting better learning in the Tritave Shepard tone condition.

For all exposure conditions, both pre- and post-exposure ratings were significantly correlated with the event distribution of exposure tones (Figure 2), with the ratings being significantly higher in correlation post-exposure than pre-exposure (Tritave Shepard tones condition: pre-exposure average $r = 0.5382$, two-tailed t-test against chance level of zero: $t(23) = 8.75$, $p < 0.001$; post-exposure average $r = 0.79$, $t(23) = 35.6$, $p < 0.001$. T-test comparing pre-exposure and post-exposure correlations: $t(23) = 3.81$, $p < 0.001$. Octave Shepard tones condition: pre-exposure average $r = 0.381$, $t(23) = 8.70$, $p < 0.001$; post-exposure average $r = 0.61$, $t(23) = 13.14$, $p < 0.001$; two-tailed t-test comparing pre- and post-exposure ratings: $t(23) = 3.81$, $p < 0.001$). Pure tone condition: pre-exposure average $r = 0.4583$, $t(23) = 9.05$, $p < 0.001$; post-exposure average $r = 0.6543$, $p < 0.001$, $t(23) = 15.8$, $p < 0.001$, two-tailed t-test comparing pre- and post-exposure ratings: $t(23) = 3.92$, $p < 0.001$. Only the no-exposure control condition did not show a significant increase in ratings, as expected: pre ratings average $r = 0.2592$, $t(23) = 3.67$, $p = 0.0013$, post ratings average $r = 0.3212$, $t(23) = 5.27$, $p < 0.001$, two-tailed t-test comparing pre- and post-exposure ratings: $t(23) = 1.0358$, $p = 0.3110$. Since a probe tone context had to be presented in order to obtain ratings, participants relied on the context to make ratings before exposure. To disentangle the effect of the context from the effect of exposure on probe-tone ratings, partial correlations were obtained by partialling out the event distribution of the context from the relationship between the event distribution of the exposure and the ratings. For all conditions, partial correlations of pre-exposure ratings were now at chance levels, indicating no knowledge of the scale before exposure (Tritave Shepard tones: pre-exposure partial correlation: $r_{ratings*exposure|context} = 0.027$, $t(23) = 0.42$, n.s.; Octave Shepard tones: pre-exposure partial correlation: $r_{ratings*exposure|context} = -0.009$, $t(23) = 0.14$, n.s.; pure tones: $r_{ratings*exposure|context} = 0.0968$, $t(23) = 1.79$, n.s.; no-exposure: $r_{ratings*exposure|context} = 0.0295$, $t(23) = 0.54$, n.s.), whereas post-exposure partial correlations were

significantly above chance (Tritave: post-exposure $r_{ratings*exposure|context} = 0.38$, $t(23) = 7.17$, $p < 0.001$; Octave: post-exposure $r_{ratings*exposure|context} = 0.27$, $t(23) = 4.86$, $p < 0.001$; pure tones $r_{ratings*exposure|context} = 0.3319$, $t(23) = 4.70$, $p < .001$; no-exposure control $r_{ratings*exposure|context} = 0.1102$, $t(23) = 1.7151$, n.s.). Two-tailed paired t-tests comparing pre- and post-exposure partial correlations showed improvement in post- vs. pre-exposure partial correlations only in the Tritave Shepard tone and pure tone conditions, not in the Octave Shepard tone and the no-exposure conditions (Tritave: $t(23) = 3.8864$, $p < 0.001$; pure tone: $t(23) = 3.1115$, $p < .01$; Octave: $t(23) = -1.6938$, n.s., no-exposure control: $t(23) = 1.51$, n.s.). The partial correlations also showed an overall significant increase over time (pre vs. post-exposure): $F(1,91) = 27.02$, $p < 0.001$, and a significant overall effect of group (Octave vs. Tritave vs. pure tone vs. no-exposure), $F(3,91) = 9.89$, $p < 0.001$ as well as a marginally significant interaction between group and time ($F(1,91) = 2.30$, $p = .08$). Only the Tritave Shepard tone condition led to better learning than the no-exposure condition, as indexed by the significantly higher post-exposure partial correlations compared to the no-exposure condition ($t(45) = 2.17$, $p = .0356$). Together, these results show better learning in the Tritave Shepard tone condition, the timbre that was congruent with the BP scale (Figure 3).

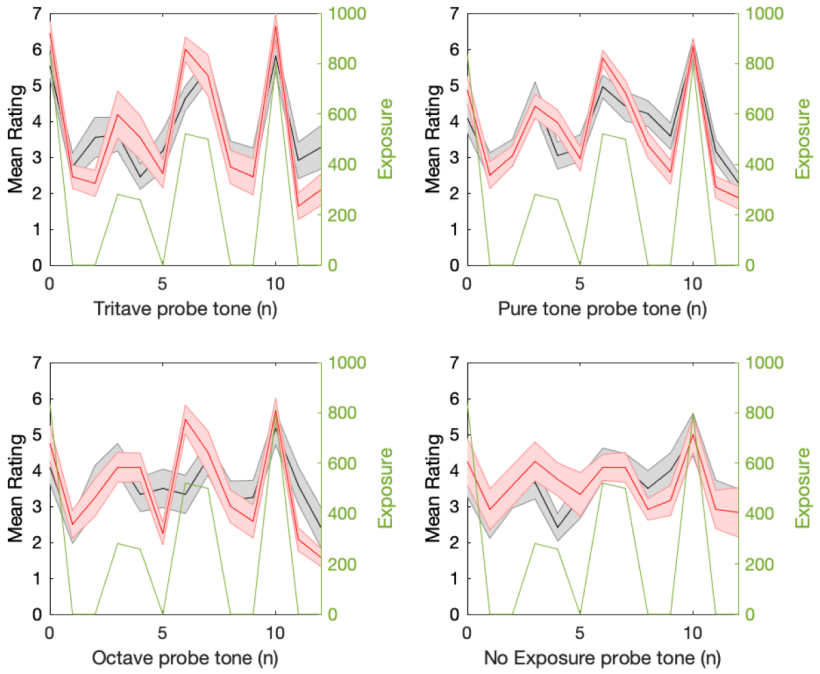

Figure 2. Probe-tone ratings in the four experimental conditions. X-axis represents the probe tone in steps along the BP scale. Pre-exposure ratings are in black and post-exposure ratings are in red. The exposure profile is shown in green. The red ratings are more highly correlated with the green exposure profile than the black ratings, suggesting that learning occurred as a result of exposure.

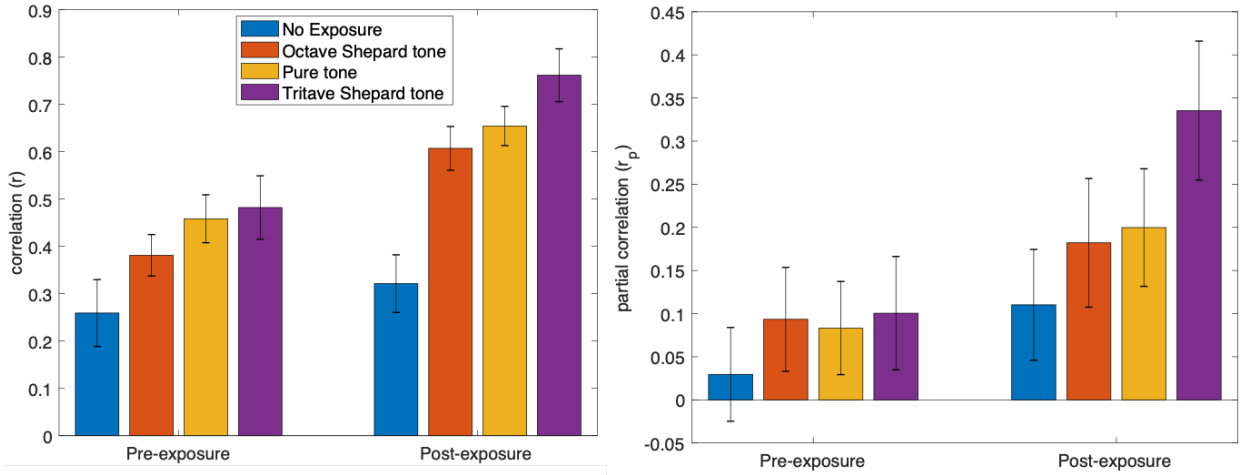

Figure 3. Correlations and partial correlations between probe-tone ratings and exposure frequencies.

Discussion

Musical scales around the world are built around whole-integer mathematical ratios in frequency, which are perceived as consonant, but there are cultural and training-related differences in preference for consonance (McDermott, Lehr, & Oxenham, 2010; McDermott, Schultz, Undurraga, & Godoy, 2016), and we do not know why consonant intervals are relatively important in musical scales across many cultures. Some theories posit that musical scale structures reflect statistical properties of periodic sounds in the environment, such as speech sounds (Bowling & Purves, 2015). Support for this comes from the association between speech sounds and consonant pitch intervals, but until now this has been mostly correlational evidence. Here, using a new musical system to which participants are exposed for the first time, we test the causal relationship between sound spectrum and learning of musical structure. Participants were better at learning the BP scale when they heard melodies in the BP scale presented in a timbre that was consistent with BP scale structure. Results show a relationship between timbre — specifically the spacing between adjacent harmonics — and the learning of scale structure, thus providing the first support for the role of sound spectrum in the statistical learning of music.

Learning of scale structure was quantified using the exposure-related change in correlation between subjective ratings from the probe-tone method and the distribution of event probabilities of exposure. Probe-tone methodology has shown sensitivity to musical scale structure (Krumhansl, 1990). Here, by comparing probe-tone profiles before and after exposure to tone sequences in a new musical system, we can capture new musical knowledge as it emerges for the first time as a result of ongoing statistical learning from short-term exposure. Results converge with existing literature on statistical learning to demonstrate that the human brain is a flexible learner that adapts rapidly to the frequencies and probabilities of sounds in the environment (Daikoku, Yatomi, & Yumoto, 2017; Jonaitis & Saffran, 2009; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2009; Saffran et al., 1999).

While other studies have conceptualized statistical learning as the sensitivity to transitional probabilities, which are first-order probabilities of an event given its previous event, here we conceptualize

the sensitivity to scale structure as zero-order probability, or the distribution of frequency of occurrence across different pitches along the musical scale. This sensitivity to scale structure underlies musical tonality and is best captured behaviorally using the probe-tone method. Previous evidence from electrophysiological recordings (Loui et al., 2009) also support the idea that musical harmonies in the BP scale can be rapidly acquired via statistical learning, thus converging with the present results.

Pre-exposure probe-tone ratings showed significant correlation with the scale structure, however these dropped to chance levels after partialling out the effect of the tone sequence used to obtain the probe-tone ratings. This suggests that contextual information played a role in these ratings. After partialling out the effect of the immediate context, post-exposure probe-tone ratings were significantly more accurate (i.e. more highly correlated with the exposure scale structure) for the Tritave Shepard tone condition than for the Octave Shepard tone condition. Since the Tritave Shepard tones are congruent whereas Octave Shepard tones are incongruent with the tritave-based tuning system, more accurate ratings for congruent timbres than for incongruent timbres suggests that the timbre, specifically the spectral content, affected participants' sensitivity to frequencies of tones in their input. While the study of timbre is often manipulated by varying musical instruments (Marozeau, de Cheveigne, McAdams, & Winsberg, 2003; Menon et al., 2002; Shahin, Roberts, Chau, Trainor, & Miller, 2008), here spectral content was varied systematically by changing the arrangement of frequency components in pitches, resulting in Shepard tones that vary in the spread of energy across the frequency spectrum while keeping other acoustic features constant. This more controlled method ensures that spectral content can be manipulated independently of temporal content and spectrotemporal flux, which are two of the other features that contribute to the overall percept of timbre (McAdams, 2013).

Although the present study taps into a fundamental aspect of musical ability, participants were unselected for musical training, as previous studies on BP scale learning had shown that statistical learning of the new musical system did not interact with musical training (Loui et al., 2010). Indeed, the results show robust learning as indicated by increased correlations over time, as well as sensitivity to context as indicated

by a decrease in correlation scores when effects of context were partialled out. Rather than reflecting a music-specific ability, performance on the rating tasks here may reflect more domain-general learning abilities that also underlie the input-based acquisition of other materials such as speech and language, and environmental sounds more generally. Thus, results converge with findings from the domain of speech and language acquisition, where timbre is found to play a role in mother-infant communication (Piazza, Iordan, & Lew-Williams, 2017). In this broader context, our results contribute to a growing body of evidence on the importance of timbre, specifically spectral information, as a crucial source of input in forming our schemas for speech, music, and the auditory environment more generally. In sum, the present work at the intersection of music theory, music technology, and cognitive psychology may shed light on psychological and biological theories of learning. By reimagining the conventions of musical systems, we can start to answer novel questions about the extent to which our minds derive structure from exposure to sounds within our environment.

Materials and Methods

Participants. Ninety-six undergraduates at the University of California at Berkeley participated in this experiment in return for course credit. All participants reported having normal hearing and were unselected for musical training. Each participant was randomly assigned to an exposure condition. After providing written informed consent, participants were tested in a sound-attenuated room while facing a Dell desktop PC running Max/MSP software (Zicarelli, 1998). Stimuli were presented at approximately 70 dB through Sennheiser HD 280 headphones. All data and code used in analysis are available on https://osf.io/pjkq2/.

Procedure. The experiment was conducted in three phases: 1) pre-exposure probe-tone ratings test, 2) exposure phase, and 3) post-exposure probe-tone ratings test.

1. Pre-exposure probe-tone ratings test: Thirteen trials were conducted in this phase. In each trial, participants were presented with a melody in the Bohlen-Pierce scale, followed by a tone (Krumhansl, 1991). Participants' task was to rate how well the last tone (i.e. the probe tone) fit the preceding melody, on a scale of 1 (least fitting) to 7 (best fitting).

2. Exposure phase: Participants were presented with 400 melodies in the BP scale. Each melody was 4 seconds long, consisting of 8 tones of 500 milliseconds each, including rise and fall times of 5 ms each, with a 500 ms silent gap between successive melodies, resulting in an exposure phase that lasted 30 minutes. The pitches of the melodies were determined by a finite state grammar previously described in Loui et al. (2010). The melodies were presented in one of four possible timbre conditions as described below: congruent (Tritave Shepard tones), incongruent (Octave Shepard tones), or neutral (Pure tones), or the no-exposure control condition.

Congruent Condition. **Tritave-based Shepard tones** were computer-generated complex tones with five partials centering around the target frequency, where the partials were related to the fundamental in 3:1 ratios (tritave) in frequency. Since the BP scale is based on the 3:1 frequency ratio (instead of 2:1 frequency ratio, i.e. the octave), the timbre of tones in this condition is congruent with a tritave-based musical system.

Incongruent Condition. **Octave Shepard complex tones** were computer-generated complex tones with five partials centering around the target frequency, where the partials were related to the fundamental in 2:1 ratios in frequency. Thus these tones were incongruous with the BP scale, but consistent with Western and other musical scales that are based on the 2:1 frequency ratio.

Neutral Condition. **Pure tones** were computer-generated with fundamental frequency only, i.e. no partials; thus these tones provided a clear percept of pitch, but did not provide any spectral cues as to the tuning system.

Control Condition. In a **no-exposure control condition**, participants made probe-tone ratings twice, using the same procedures as the Neutral (pure tone) condition. They were not given exposure to any auditory stimuli between the pre- and post-exposure ratings; instead, they were asked to sit quietly for 30 minutes between the two probe tone tests, as this was the duration of the exposure condition.

3. Post-exposure probe-tone ratings test: Probe-tone ratings were conducted again after exposure, using the same methods as phase 1.

References

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129-135.

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 17*(3), 577-600.

Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proc Natl Acad Sci U S A, 112*(36), 11155-11160.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound* (Vol. null).

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences, 113*(6), 1666-1671.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *J Acoust Soc Am, 118*(1), 471-482.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *Journal of Experimental Psychology: General, 113*(3), 394-412.

Creel, S. C., & Newport, E. L. (2002). *Tonal profiles of artificial scales: Implications for music learning.* Paper presented at the Proceedings of the 7th International Conference on Music Perception and Cognition.

Cross, I. (2001). Music, mind and evolution. *Psychology of Music, 29*(1), 95-102.

Daikoku, T., Yatomi, Y., & Yumoto, M. (2017). Statistical learning of an auditory sequence and reorganization of acquired knowledge: A time course of word segmentation and ordering. *Neuropsychologia, 95*, 1-10.

Han, S., Sundararajan, J., Bowling, D. L., Lake, J., & Purves, D. (2011). Co-variation of tonality in the music and speech of different cultures. *PLoS ONE, 6*(5), e20160.

Jonaitis, E. M., & Saffran, J. R. (2009). Learning Harmony: The Role of Serial Statistics. *Cognitive Science, 33*(5), 951-968.

Koelsch, S., Gunter, T., Friederici, A. D., & Schroger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *J Cogn Neurosci, 12*(3), 520-541.

Krumhansl, C. L. (1987). General properties of musical pitch systems: Some psychological considerations. In J. Sundberg (Ed.), *Harmony and Tonality* (Vol. 54, pp. 33-52). Stockholm: Royal Swedish Academy of Music.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY US: Oxford University Press.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 579-594.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Jarvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition, 76*(1), 13-58.

Landau, S. M., & D'Esposito, M. (2006). Sequence learning in pianists and nonpianists: an fMRI study of motor expertise. *Cognitive, Affective, & Behavioral Neuroscience, 6*(3), 246-259.

Leung, Y., & Dean, R. T. (2018). Learning unfamiliar pitch intervals: A novel paradigm for demonstrating the learning of statistical associations between musical pitches. *PLOS ONE, 13*(8), e0203026.

Loui, P. (2012). Learning and liking of melody and harmony: further studies in artificial grammar learning. *Top Cogn Sci, 4*(4), 554-567.

Loui, P., & Wessel, D. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Perception & Psychophysics, 69*(7), 1084-1092.

Loui, P., & Wessel, D. L. (2008). Learning and Liking an Artificial Musical System: Effects of Set Size and Repeated Exposure. *Musicae Scientiae, 12*(2), 207-230.

Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans Rapidly Learn Grammatical Structure in a New Musical Scale. *Music Perception, 27*(5), 377-388.

Loui, P., Wu, E. H., Wessel, D. L., & Knight, R. T. (2009). A Generalized Mechanism for Perception of Pitch Patterns. *Journal of Neuroscience, 29*(2), 454-459.

Marozeau, J., de Cheveigne, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *J Acoust Soc Am, 114*(5), 2946-2957.

Mathews, M. V., Pierce, J. R., Reeves, A., & Roberts, L. A. (1988). Theoretical and experimental explorations of the Bohlen-Pierce scale. *J Acoustical Soc Am, 84*, 1214-1222.

McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The Psychology of Music* (3rd ed., pp. 35-67). New York: Academic Press.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual Differences Reveal the Basis of Consonance. *Current Biology, 20*(11), 1035-1041.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature, advance online publication*.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., . . . McAdams, S. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage, 17*(4), 1742-1754.

Mithen, S. J. (2007). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body.* . Cambridge, MA: Harvard UP.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*(1), B35-b45.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2009). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *Neuroimage, 50*(1), 302-313.

Piazza, E. A., Iordan, M. C., & Lew-Williams, C. (2017). Mothers Consistently Alter Their Unique Vocal Fingerprints When Communicating with Infants. *Current Biology, 27*(20), 3162-3167.

Russo, F. A. (2009). Towards a functional hearing test for musicians: The probe tone method. In M. Chasin (Ed.), *Hearing Loss in Musicians* (pp. 145-152). San Diego, CA: Plural Publishing.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

Schwartz, D. A., Howe, C. Q., & Purves, D. (2003). The statistical structure of human speech sounds predicts musical universals. *J Neurosci, 23*(18), 7160-7168.

Schwartz, D. A., & Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hear Res, 194*(1-2), 31-46.

Sethares, W. (2004). *Tuning Timbre Spectrum Scale*: Springer-Verlag.

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *Neuroimage, 41*(1), 113-122.

Shepard, R. N. (1964). Circularity in Judgments of Relative Pitch. *The Journal of the Acoustical Society of America, 36*(12), 2346-2353.

Smith, C. P. (1951). A Phoneme Detector. *The Journal of the Acoustical Society of America, 23*(5), 632-632.

Tillmann, B., & McAdams, S. (2004). Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted With Acoustical (Dis)Similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1131-1142.

Wessel, D. L. (1979). Timbre Space as a Musical Control Structure. *Computer Music Journal, 3*(2), 45-52.

Zicarelli, D. (1998). *An extensible real-time signal processing environment for Max*. Paper presented at the Proceedings of the International Computer Music Conference, University of Michigan, Ann Arbor, USA.