New Music System Reveals Spectral Contribution to Statistical Learning

Psyche Loui

Northeastern University

Abstract

Knowledge of language and music depends upon the ability to perceive relationships between sounds in order to form a stable mental representation of grammatical structure. Although abundant evidence exists for the learning of statistical structure (specifically event probability structure) from units of sounds, little research has explored what constitutes a distinct unit of sound. Here we show that spectral content is an important cue within sound units for acquiring a stable mental representation of sound structure. Tone sequences in a novel scale were constructed and presented to participants on three different timbres with 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. Learning outcome was assessed by correlating pre/post-exposure ratings and the statistical structure of tones within the exposure period. Participants did best at learning with congruent timbres, and worst with incongruent timbres. 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

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, which evolved through complex cultural evolution over time, and are already overlearned as a result of exposure within that culture, and to some extent even by members outside the culture throughout the lifespan (Matsunaga, Hartono, Yokosawa, & Abe, 2020). 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 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 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 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 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, 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.

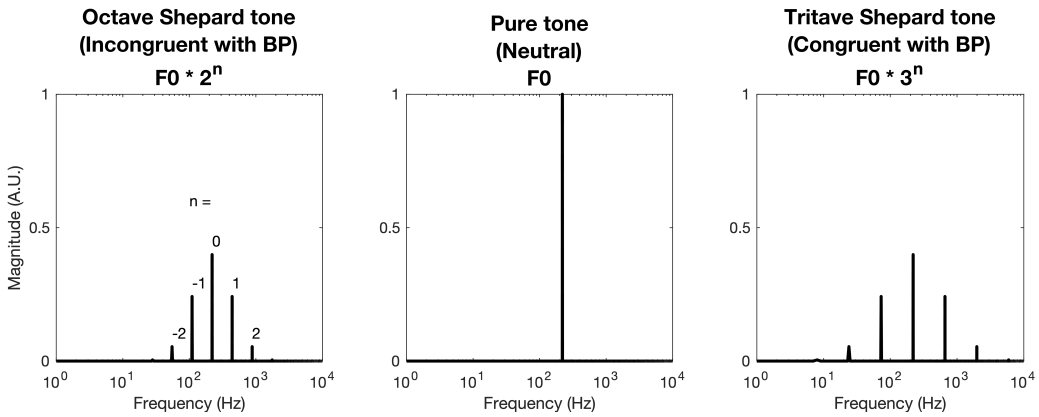


Figure 1. Frequency-domain representations of Tritave-based Shepard tones, Octave-based Shepard tones, and pure tones.

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 ($F(1,91) = 2.6833$, $p = .05$).

For all conditions, both pre- and post-exposure ratings were significantly correlated with the event distribution of exposure tones, 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: $t(23) = 3.92$, $p < 0.001$. No-exposure control condition did not show a significant increase in ratings: 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, the context also affected learning even 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), $F(3,91) = 9.89$, $p < 0.001$.

Importantly, 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), $t(45) = 2.1120$, $p = 0.04$). When we partialled out the effects of the context used to obtain the ratings, the resulting partial correlations showed that 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 ($t(45) = 2.17$, $p = .0356$).

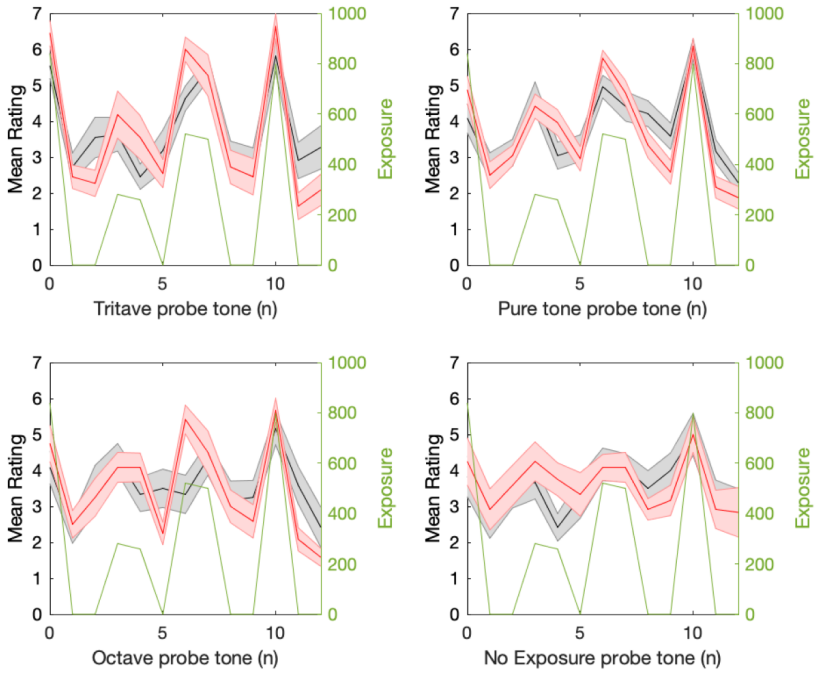


Figure 2. Probe tone ratings in the four experimental conditions. X-axis represents the probe tone in steps along the BP scale. Pre-exposure ratings are in black and post-exposure ratings are in red. The exposure profile is shown in green. The red ratings are more highly correlated with the green exposure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

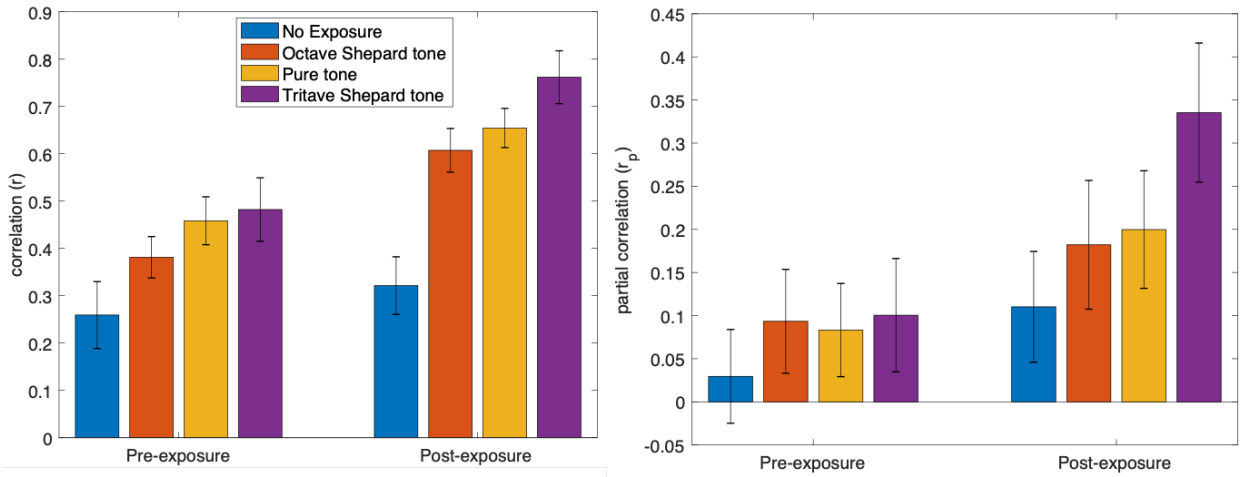


Figure 3. Correlations and partial correlations between probe tone ratings and exposure frequencies.

Discussion

Musical scales around the world are built around whole-integer mathematical ratios in frequency, which are perceived as consonant, but we do not know why consonant intervals are relatively important across many musical 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 that participants are exposed to 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 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 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, specifically sensitivity to tonality, which emphasizes consonant scale degrees. 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 event frequency 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), using the BP scale, also support the idea that tonality is rapidly acquired via similar mechanisms as statistical learning, thus converging with the present results.

Pre-exposure probe tone ratings showed significant correlation with timbre, 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 exposure) for the Tritave Shepard tone condition, which was more congruent with the tritave-based tuning system, than for the Octave Shepard tone condition. The more accurate ratings for congruent timbres compared to incongruent timbres suggests that the arrangement of frequency components in pitches affected participants' sensitivity to frequencies of tones in their input.

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 was not affected by 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. Thus, 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.

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 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 (the probe tone) fit the preceding melody, on a scale of 1 (least fitting) to 7 (best fitting).

2. Exposure phase: Participants heard 400 melodies in the Bohlen-Pierce scale, in one of four possible conditions: congruent (Tritave Shepard tone), incongruent (Octave Shepard tone), or neutral (Pure tone), or the no-exposure control condition. This exposure phase lasted approximately 30 minutes.

Congruent Condition. Four hundred melodies in **Tritave Shepard complex tones** were presented with no repeats for half an hour during exposure phase. 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 in frequency; thus the timbre of tones in this condition were congruent with a tritave-based musical system.

Incongruent Condition. Four hundred melodies in **Octave Shepard complex tones** were presented with no repeats for 30 minutes during the exposure phase. Octave-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 2:1 ratios in frequency; thus these tones were incongruous with the musical system which is based on the 3:1 ratio of the tritave.

Neutral Condition. Four hundred melodies in **Pure tones** were presented with no repeats for 30 minutes during the exposure phase. 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.

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.

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.

Matsunaga, R., Hartono, P., Yokosawa, K., & Abe, J.-i. (2020). The Development of Sensitivity to Tonality Structure of Music. *Music Perception: An Interdisciplinary Journal, 37*(3), 225.

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.

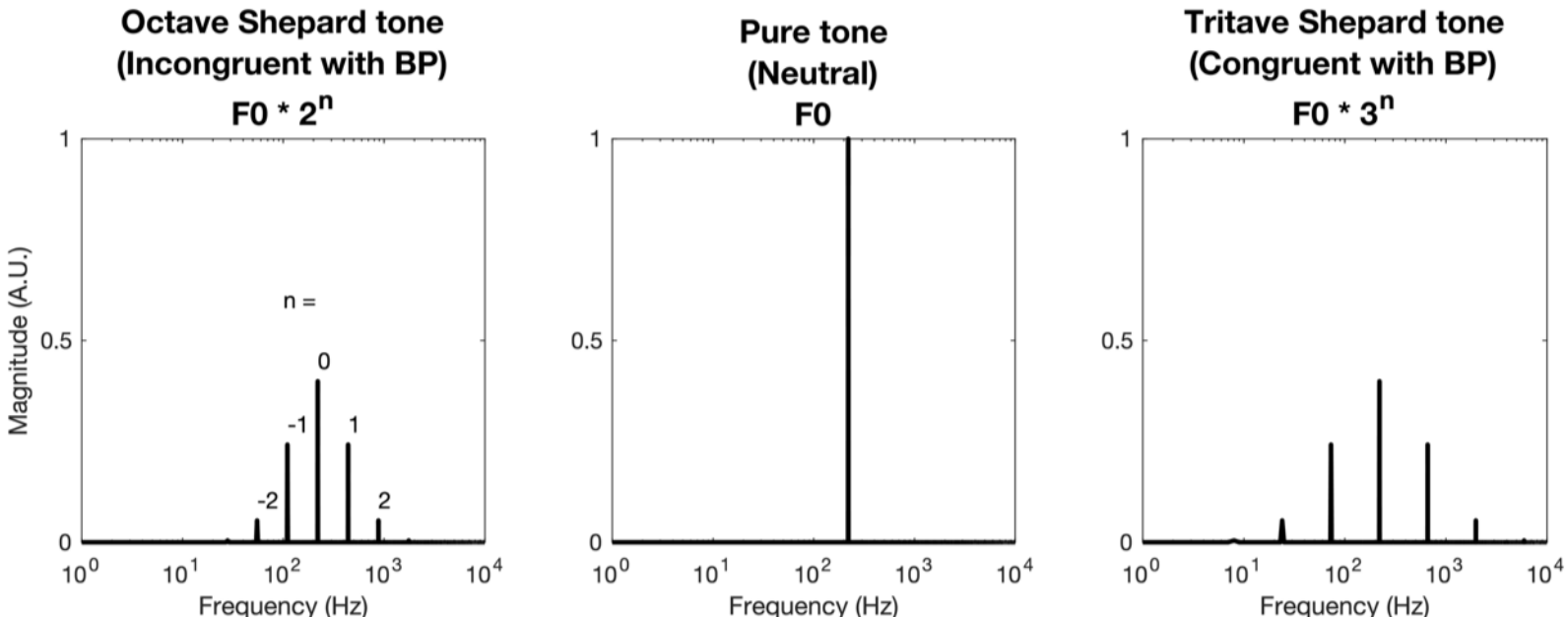
Sethares, W. (2004). *Tuning Timbre Spectrum Scale*: Springer-Verlag.

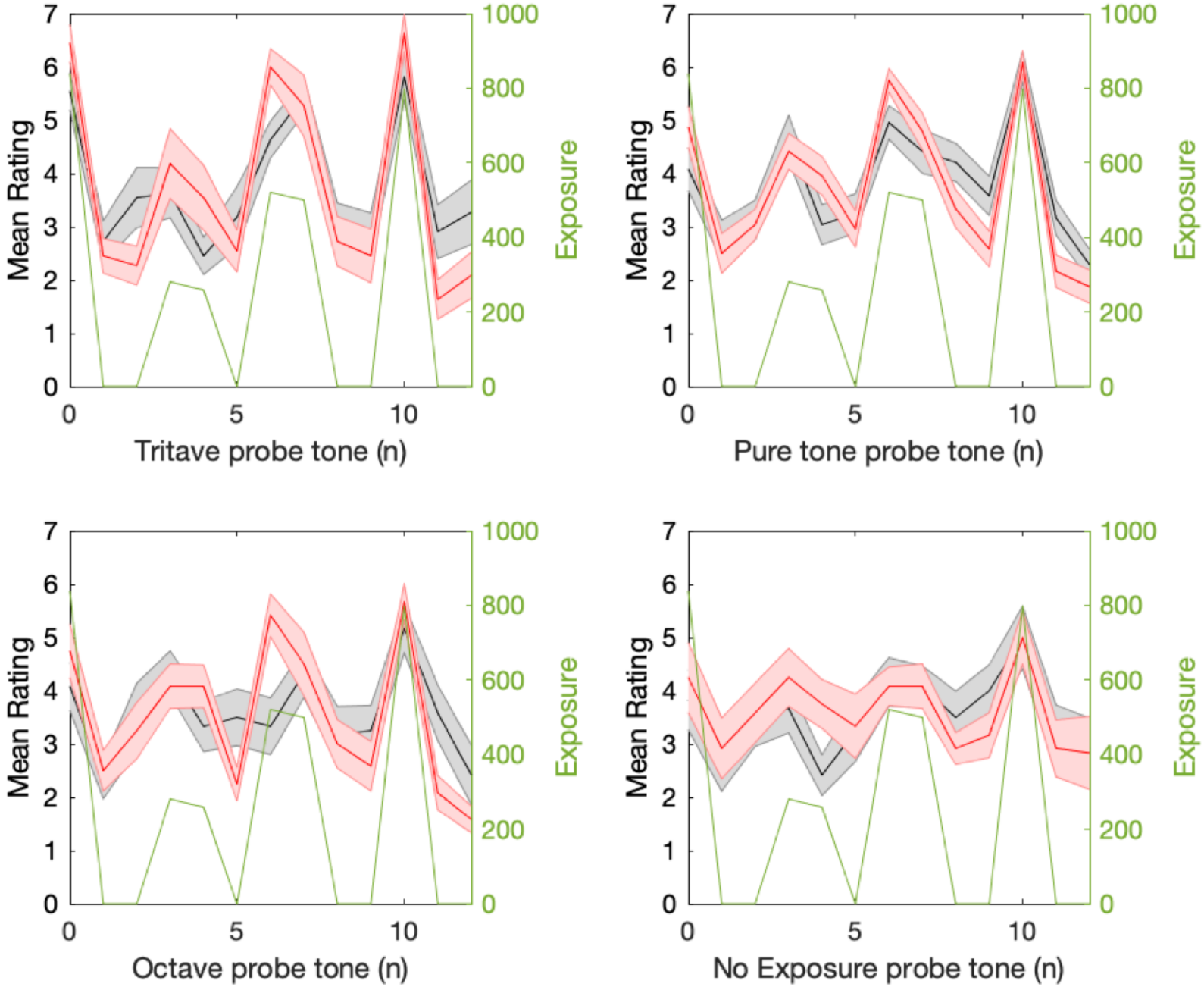
Shepard, R. N. (1964). Circularity in Judgments of Relative Pitch. *The Journal of the Acoustical Society of America, 36*(12), 2346-2353.

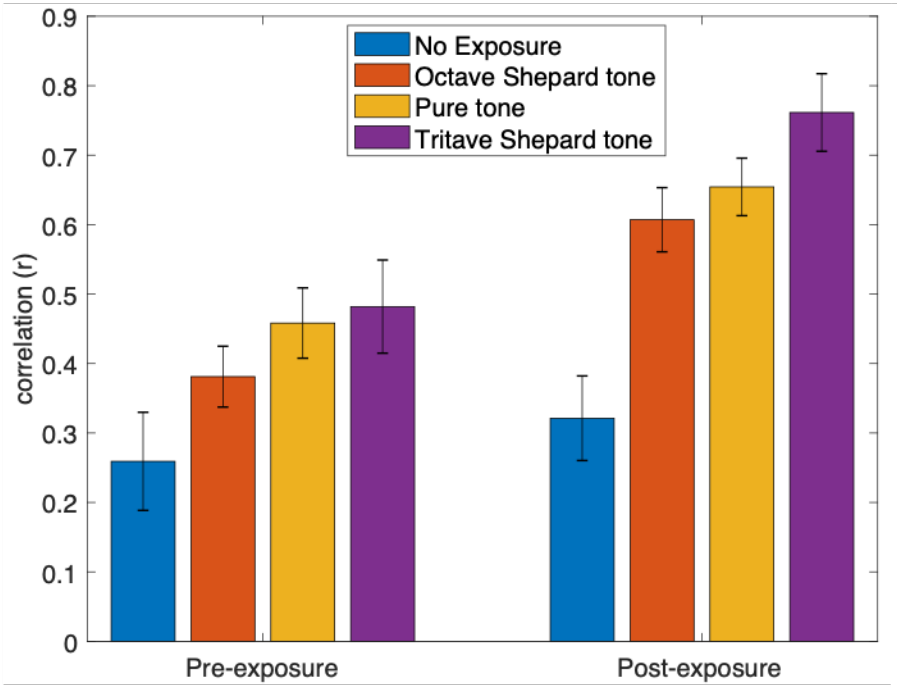
Smith, C. P. (1951). A Phoneme Detector. *The Journal of the Acoustical Society of America, 23*(5), 632-632.

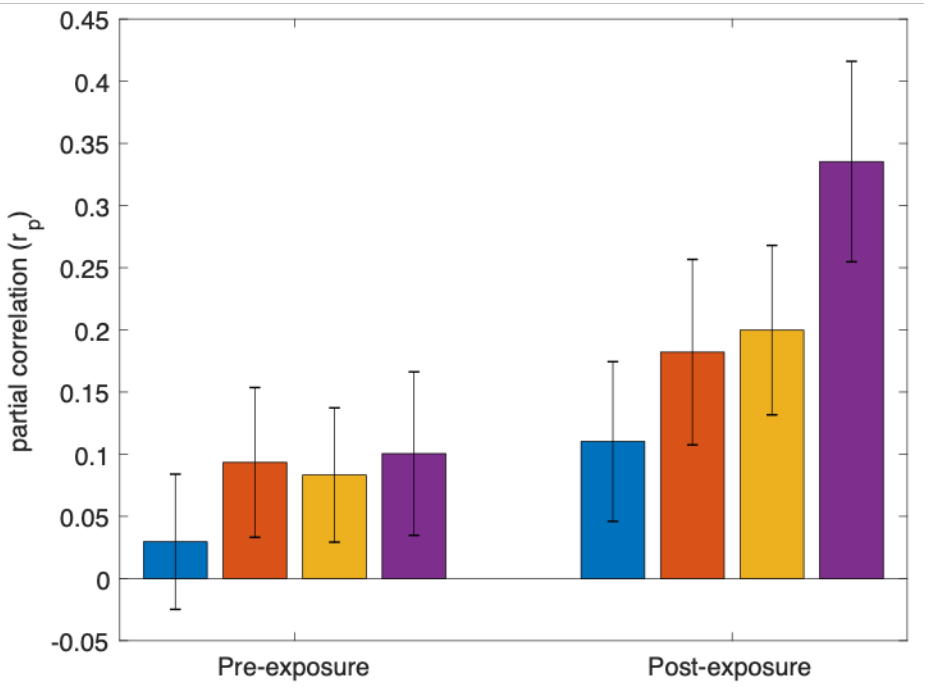
Tillmann, B., & McAdams, S. (2004). Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted With Acoustical (Dis)Similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1131-1142.

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


Octave Shepard tone
(Incongruent with BP)
F0 * 2^n
Pure tone
(Neutral)
F0
Tritave Shepard tone
(Congruent with BP)
F0 * 3^n
Magnitude (A.U.)
Frequency (Hz)
n =
-2
-1
0
1
2
1
0.5
0
10^0
10^1
10^2
10^3
10^4


Mean Rating
Exposure
Tritave probe tone (n)
Pure tone probe tone (n)
Octave probe tone (n)
No Exposure probe tone (n)
0
1
2
3
4
5
6
7
0
200
400
600
800
1000
0
5
10


No Exposure
Octave Shepard tone
Pure tone
Tritave Shepard tone
correlation (r)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Pre-exposure
Post-exposure


partial correlation (r_p)
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
Pre-exposure
Post-exposure