

Learning Harmony: The Role of Serial Statistics

Erin McMullen Jonaitis, Jenny R. Saffran

Department of Psychology, University of Wisconsin, Madison

Received 29 January 2007; received in revised form 9 September 2008; accepted 4 February 2009

Abstract

How do listeners learn about the statistical regularities underlying musical harmony? In traditional Western music, certain chords predict the occurrence of other chords: Given a particular chord, not all chords are equally likely to follow. In Experiments 1 and 2, we investigated whether adults make use of statistical information when learning new musical structures. Listeners were exposed to a novel musical system containing phrases generated using an artificial grammar. This new system contained statistical structure quite different from Western tonal music. Our results suggest that learners take advantage of the statistical patterning of chords to acquire new musical structures, similar to learning processes previously observed for language learning.

Keywords: Learning; Music; Syntax; Language acquisition

1. Introduction

How does music—at its heart a strongly nonreferential system—carry meaning for listeners? One influential framework for understanding music cognition posits that musical meaning inheres in the creation and violation of expectations in a listener’s mind about what will come next (Meyer, 1956). While many studies have provided empirical support for the existence of musical expectancies (e.g., Bharucha & Stoeckig, 1987; Schmuckler, 1989), relatively little research has directly addressed the question of the origins of musical expectancies during ontogeny. Some principles of musical organization appear to be universal (e.g., Schellenberg, Adachi, Purdy, & McKinnon, 2002). However, others vary cross-culturally (e.g., Castellano, Bharucha, & Krumhansl, 1984), so musical expectations cannot be fully explained by innate biases. By what mechanisms do humans acquire new musical expectations?

Correspondence should be sent to Erin McMullen Jonaitis, Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, WI 53706. E-mail: emjonaitis@gmail.com

The general question of how learners converge on similar interpretations of complex input has a long history in cognitive science. There is a rich literature on learning of patterns without conscious awareness, beginning with Reber's (1967, 1969) classic artificial grammar experiments, in which subjects learned to classify letter strings with regard to a pattern established by previous letter strings but were unable to explain their judgments. Conceptually similar results have been obtained with motor tasks in which a spatial pattern is covertly embedded (Hunt & Aslin, 2001; Lewicki, Hill, & Bizot, 1988). The implicit nature of these tasks makes them a good parallel for the learning of musical "rules," which are typically hard for listeners to verbalize without explicit instruction in music theory, despite general familiarity with what constitutes well-formed music within a style (Smith & Melara, 1990).

It may also be profitable to draw on insights from the field of language acquisition. For human learners, language and music represent two of the most complex systems that are typically acquired during the course of development. The two domains have some broad features in common: in their primary form, both are richly patterned, hierarchical auditory structures that serve a variety of communicative functions (Bod, 2002). They also exhibit similar patterns of emergence in ontogeny (McMullen & Saffran, 2004), where infants start with certain culturally independent predispositions (Aslin, Jusczyk, & Pisoni, 1998a; Schellenberg et al., 2002; Trainor & Heinmiller, 1999; Trehub, Cohen, Thorpe, & Morrongiello, 1986; Zentner & Kagan, 1998) which are modified upon exposure to the set of relevant elements used by their culture (Hannon & Trehub, 2005; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Thiessen & Saffran, 2003; Trehub et al., 1986; Werker & Lalonde, 1988). These developmental similarities between language and music suggest the possibility of overlapping learning mechanisms.

While it is clear that some aspects of musical and linguistic processing are subserved by different systems in the adult brain, the degree to which this specification emerges via learning is an open question (e.g., Peretz, 2005). One potential source of overlap is that listeners of all ages can acquire simple statistical regularities in each domain, such as the probabilities with which linguistic or musical primitives co-occur adjacently (e.g., Aslin, Saffran, & Newport, 1998b; Saffran, 2003; Saffran, Aslin, & Newport, 1996; Saffran & Griepentrog, 2001; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Reeck, Niehbur, & Wilson, 2005; Tillmann & McAdams, 2004) and nonadjacently (Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004). This literature suggests that probability detection may be a learning mechanism that operates over both musical and linguistic materials.

The acquisition of syntactic structure and expectations is a complex learning problem common to both music and language. In both domains, an infinite variety of legal expressions can be produced by combining smaller elements in a systematic way. In languages, these expressions correspond to morphemes combined to form sentences; in musical systems, they correspond to phrases (Lerdahl & Jackendoff, 1983). Because these combinatorial regularities vary cross-culturally, they must be learned via exposure to the native system. It is thus possible that overlapping learning mechanisms may contribute to the acquisition of the complex regularities found in music and language. This hypothesis is reinforced

by brain imaging studies suggesting similar processing mechanisms for detecting structural anomalies in music and language (Maess, Koelsch, Gunter, & Friederici, 2001; Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

Some aspects of linguistic grammatical structure are informed by statistics present in the input. Adult learners can use distributional statistics—patterns of words—to discover grammatical categories (like noun and verb) in the input (e.g., Mintz, 2002), and to discover how those categories combine into phrases (e.g., Morgan, Meier, & Newport, 1987, 1989; Morgan & Newport, 1981; Saffran, 2001, 2002). More generally, sequential regularities provide a basis for generalization beyond the input for both adults and infants (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Peña, Bonatti, Nespor, & Mehler, 2002). Even 12-month-old infants can track the distributions of words and word categories in syntax-learning tasks of varying complexity (Gomez & Gerken, 1999; Saffran et al., 2008; Saffran & Wilson, 2003).

In music, grammatical well-formedness has its closest analog in musical expectancies, which exist both globally (the distribution of pitches and chords, organized within keys) and locally (transitions from one pitch or chord to the next). Many experiments have demonstrated that listeners' expectancy judgments in a given style relate strongly to the distributional and sequential statistics of that style (Bharucha & Stoeckig, 1987; Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003; Krumhansl, 1990; Krumhansl & Shepard, 1979; Schmuckler, 1989; Smith & Melara, 1990), a relationship that suggests prior learning of the statistics of the system. Further, computational models constructed to attend to such statistical information have contributed to modern music theory (Mavromatis, 2005; Raphael & Stoddard, 2004), and many perform similarly to expert human listeners on a variety of tasks (Bod, 2002; Krumhansl, Louhivuori, Toiviainen, Järvinen, & Eerola, 1999; Krumhansl et al., 2000; Tillmann, Bharucha, & Bigand, 2000).

On the other hand, fewer studies have directly probed adults' ability to acquire the statistics of an unfamiliar musical idiom. Several studies have demonstrated listeners' attentiveness to the distributional hierarchy of pitches within a context phrase in an unfamiliar style, even when doing so requires the suppression of irrelevant prior knowledge (Castellano et al., 1984; Krumhansl et al., 1999, 2000; Oram & Cuddy, 1995). However, it remains to be seen whether local regularities can also be learned. One candidate regularity resides in the pattern of chord transitions used within a given style. In Western tonal music, for example, not all transitions between chords are equally likely; the occurrence of a given chord should constrain the listener's expectations about what may come next, and listeners' goodness ratings of chord sequences generally conform to these patterns (McMullen & Saffran, 2005; Smith & Melara, 1990). Because the particular expectations associated with Western music do not characterize all musical systems, they are presumably learned through experience with chord sequences.

The experiments that follow were designed to examine the role of statistical information in learning and making judgments about a new style of music. Adult participants were exposed to musical corpora generated by one of two counterbalanced artificial grammars, and they were tested to determine whether statistical information about chord

transitions influenced their judgments of the well-formedness of a test corpus of musical strings. Although it is certainly not the case that *all* musical regularities proceed from harmonic transitions, these transitions represent an important form of structure. Our goal was to test the hypothesis that learners can acquire new musical regularities by tracking the distribution of chords in sequence. If so, then this learning mechanism, analogous to those previously observed in language learning tasks, may play a role in the acquisition of native musical structure.

2. Experiment 1

In this experiment, we investigated participants' ability to learn a novel musical system. To reduce our listeners' reliance on Western harmonic patterns without introducing extreme dissonance, we used the Phrygian mode as the pitch framework for our stimuli. Featured in church music that predated the tonal era, the interval patterns in Phrygian are obtained by playing a scale on the piano starting at E and containing only white keys. Because this mode shares interval patterns with Western tonal music, its diatonic triads are as consonant as those found in tonal modes. However, the relative frequency of the triads differs from major- and minor-mode music. Using chords built on this scale, we constructed two counter-balanced artificial musical grammars, A and B, each containing a subset of possible chord transitions. Grammars A and B were the reverse of each other, with very little overlap in permissible transitions between the two systems. This allowed us to control chord frequency while manipulating serial order.

During the experiment, subjects were exposed to 100 items from one system. At test, they listened to 60 new items and were asked to rate the similarity of each item to the materials heard in the exposure. We used four types of test items. A-Correct and B-Correct test items consisted of novel phrases following either Grammar A or Grammar B. A-Error and B-Error test items primarily followed one or the other grammar, but contained 1–3 chord transitions that were illegal in the relevant grammar. In all cases, A items were the reverse of B items. We then compared subjects' similarity ratings of A-Correct, A-Error, B-Correct, and B-Error test items. If listeners were able to learn the chord transition statistics, we expected that they should rate within-system items as most similar to the exposure corpus. Further, within-system correct items should be rated more highly than the within-system items containing illegal transitions.

2.1. Method

2.1.1. Participants

Forty undergraduates participated in Experiment 1 for course credit. The sample contained 8 self-defined musicians and 32 self-defined nonmusicians. Participants reported a mean of 4.4 years of experience performing music ($SD = 4.2$) and a mean of 0.75 semesters studying music theory ($SD = 1.5$). Thirty-two subjects reported having taken zero semesters of music theory.

2.1.2. Apparatus

Stimuli were composed on a PC (Windows 2000) using SONAR 4 (Cakewalk, Inc., Boston, MA). The experiment was conducted using PsyScope (Cohen, Mac Whinney, Flatt, & Provost, 1993) on a Macintosh computer (OS 9.2). Participants responded using a touch-screen monitor.

2.1.3. Stimuli

Stimuli were 4- to 10-chord progressions, constructed to conform to one of two artificial musical grammars, System A and System B (see Fig. 1). These grammars, which are reversals of each other, were built on the Phrygian scale. Both grammars constrained the items to begin and end on the tonic chord (designated by the Roman numeral I¹) whose main or root note is also the primary note of the scale; however, multiple loops through the grammar were permitted. Every progression in the System B corpus was a reversal of a progression in the System A corpus.

To create our exposure corpora, we first used a computer program to generate an exhaustive list of chord progressions that followed each grammar and contained 10 or fewer chords. From that list, we selected 50 progressions according to the following criteria. First, in order to help listeners establish a tonal hierarchy, or a subjective sense of the tonic note of the key, we used the following procedure to select corpus items from all possible progressions generated by our grammars. Taking inspiration from Krumhansl (1990), we manipulated the frequency of occurrence of each chord so that the tonic (I) was the most frequent, followed by the dominant (V) and subdominant (IV), and then the remainder of the chords (II, III, VI, VII). At the same time, we equalized the frequency of each of the possible chord transitions from a given node (Table 1A, B). The resulting set formed the harmonic basis for our A and B corpora.

The chord progressions for our two corpora were voiced as follows. Although each chord was a tonal triad containing only three distinct pitch classes, progressions were voiced in four parts (four notes per chord), with the root note of each chord appearing in two voices. This allowed us to present the harmony clearly by always presenting the root in the bass

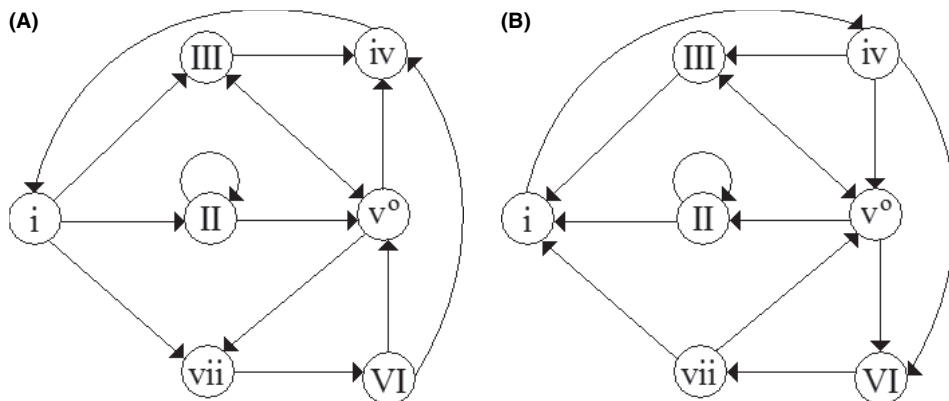


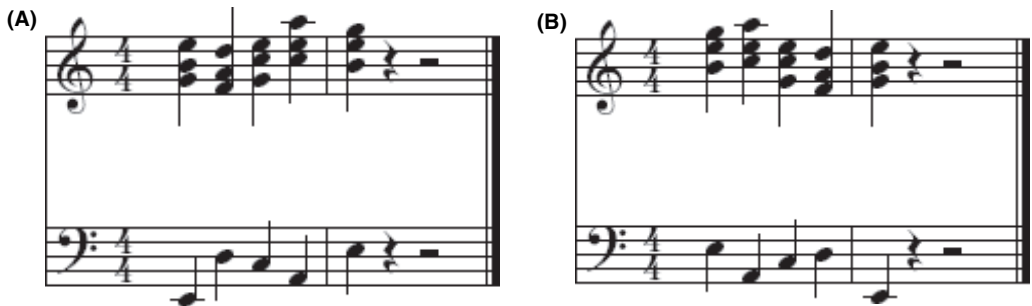
Fig. 1. Two counterbalanced artificial chord-progression grammars. (A) System A. (B) System B.

Table 1
Transition probabilities for chords in the Phrygian system

	e	F	G	a	b	C	d	End
A. Transition probabilities within System A								
e	0	0.17	0.17	0	0	0	0.17	0.50
F	0	0.50	0	0	0.50	0	0	0
G	0	0	0	0.42	0.58	0	0	0
a	1.00	0	0	0	0	0	0	0
b	0	0	0.41	0.29	0	0	0.29	0
C	0	0	0	0.50	0.50	0	0	0
d	0	0	0	0	0	1.00	0	0
B. Transition probabilities within System B								
e	0	0	0	0.50	0	0	0	0.50
F	0.50	0.50	0	0	0	0	0	0
G	0.42	0	0	0	0.58	0	0	0
a	0	0	0.33	0	0.33	0.33	0	0
b	0	0.29	0.41	0	0	0.29	0	0
C	0	0	0	0	0	0	1.00	0
d	0.50	0	0	0	0.50	0	0	0

line, while still permitting any of the chord's notes to appear in the melody. We used this flexibility to create two distinct versions, or voicings, of each chorale, with similar bass lines (modulo octave transposition) but varying upper voices, including the melody. Alternate voicings of the same progression were constrained to begin and end on a different chord tone and at a different pitch height. Basic principles of voiceleading were followed, including reversal of direction and avoidance of large jumps (Narmour, 1991). To assess our items' adherence to these principles, we calculated the expected melodic complexity ratings for the upper voice of each item using an empirically weighted sum of these factors (Eerola & North, 2000; Eerola & Toiviainen, 2004). This score represents the "weirdness" of each item's melody. A paired *t*-test revealed no difference in this score between A and B items: $t(99) = 0.68, p > .05$.² In the end, each corpus contained 100 distinct exemplars, distributed uniformly across keys. Two example items are illustrated in music notation in Fig. 2. MIDI versions of all items are accessible on the Web at <http://www.cogsci.rpi.edu/CSJarchive/Supplemental/index.html>

The test corpus consisted of 60 chord progressions that were not used in the exposure corpora. These stimuli ranged from 5 to 10 chords in length and were distributed uniformly across keys. Thirty of the progressions were "correct" items that were generated using one of the grammars (15 from System A, 15 from System B), and 30 were "error" items (15 from System A, 15 from System B) containing one, two, or three illegal transitions (five of each type per system). As in the exposure corpus, each test item within the System A group (both correct and error) was the reverse of a comparable item in the System B group. Examples are illustrated in Fig. 3. Again, voiceleading principles relating to direction and smooth contour were followed throughout. We calculated the melodic complexity of the upper voices in each subset of items (both systems, correct and error). A mixed ANOVA, with Error



ensure that participants were attending to the exposure items. Participants were not told that they would be tested following exposure.

The test phase immediately followed the exposure phase. Participants were told that they would be presented with new items, some of which belonged to the same musical system as the items they had heard during the exposure phase, and some of which did not. Their task was to judge how similar each item was to the exposure corpus, using a seven-point Likert scale ranging from “very dissimilar” to “very similar.” Six practice items preceded the 30 test items. There was no time limit given for the ratings, so the inter-stimulus interval (ISI) varied depending on the amount of time subjects took to respond to each item.

2.2. Results and discussion

2.2.1. Analysis of variance

A preliminary mixed ANOVA [Group (A-Exposed, B-Exposed) \times System (A items, B items) \times Correctness (Correct items, Error items)] showed no differences between the groups exposed to Systems A and B: The only significant effect involving the Group variable was a System \times Group effect such that higher ratings were given to test items reflecting the system heard during exposure [$F(1,38) = 42.26, p < .001$]. We thus collapsed the data from A-Exposed and B-Exposed subjects, recoding the System variable for each subject to reflect whether an item was in his or her own exposure system, and performed a 2×2 ANOVA including within-subjects factors of Correctness (Correct, Error) and System (In-System, Out-of-System). Individual listeners’ data were averaged across test items in each category to give one score per cell per listener (see Fig. 4). There was a significant effect of System [$F(1,39) = 42.5, p < .001; M_{\text{EXP}} = 4.5$ (SD = 0.58); $M_{\text{UNEXP}} = 4.0$ (SD = 0.69)], but no other significant effects [all F s < 1]. A reanalysis excluding the self-identified musicians rendered the same pattern of results.

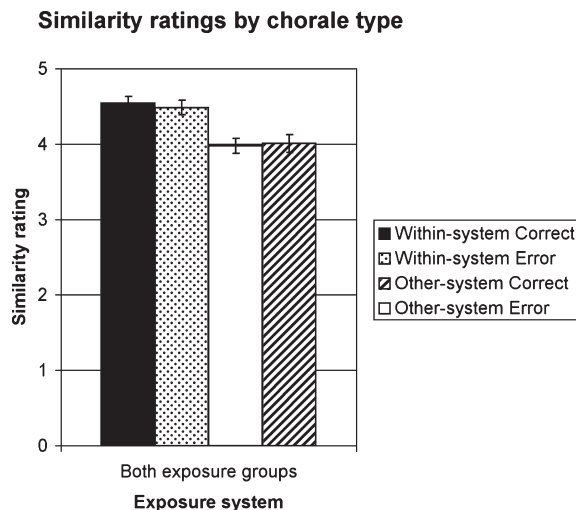


Fig. 4. Experiment 1: Similarity ratings (± 1 SE) by item type ($n = 40$).

These results confirmed our prediction that listeners can differentiate items that conform to their exposure system from items that do not. Because System A and B items differed in the serial order of their chords, but not in overall frequency content, we infer that listeners were able to learn something about legal transitions between chords in the two systems. However, listeners were unable to differentiate correct and error items within their exposure system. This finding prevents us from investigating the depth of their serial knowledge more fully; for instance, we cannot determine whether our participants were sensitive to chord-to-chord transitions in general, or merely to information contained in the cadences, or phrase endings, both of which could be used to separate A from B items. On the other hand, this result does demonstrate that the error items are not inherently “worse” than correct items for our listeners.

There are a number of possible reasons for our participants’ apparent failure to learn the details of the exposure systems. One possibility has to do with the range of test items. If the perceived difference between correct and error items is small, and the perceived difference between within-system and other-system items is considerably greater, participants may be inclined to focus on the larger distinction at the expense of the smaller. Another possibility is that the exposure time (just 100 exposure sequences) was not sufficient for detailed learning. Although the first-order transition statistics are considerably simpler than those of Western tonal music, the tonal hierarchy itself is somewhat novel, and this may increase the difficulty of the learning problem.

We addressed these issues in Experiments 2 and 3. In Experiment 2, the test included only within-system items (Correct vs. Error) in order to minimize the possibility of range-compression effects. Experiment 3 then examined the effect of doubling the exposure time.

3. Experiment 2

In Experiment 2, we examined whether participants’ performance on the similarity task could be improved by narrowing the range of test items with which they were presented. As in Experiment 1, our listeners first heard a corpus containing 100 within-system items and were then asked to rate the similarity of each item in a test corpus to the exposure corpus they had just heard. However, the test for these participants consisted only of Correct and Error items taken from that same system. If Experiment 1 listeners’ failure to discriminate Correct from Error resulted primarily from distraction by the greater dissimilarity of other-system items, removing these distractors should improve discrimination.

3.1. Method

3.1.1. Participants

Forty undergraduates participated in Experiment 2 for course credit. The sample contained 13 self-defined musicians and 27 self-defined nonmusicians. Participants reported a mean of 4.54 years of experience performing music ($SD = 4.23$) and a mean of 0.85

semesters studying music theory (SD = 2.91).³ Thirty subjects reported having taken zero semesters of music theory.

3.1.2. Apparatus

Same as Experiment 1.

3.1.3. Stimuli

Same as Experiment 1.

3.1.4. Procedure

As in Experiment 1, participants were first exposed to a corpus of 100 phrases. Because this was a pilot experiment, we began by exposing subjects only to one system, System A. Unlike Experiment 1, all test items in this experiment were drawn from the exposure system; that is, subjects were only asked to rate A-Correct and A-Error test items. Three practice items preceded the test items.

3.2. Results and discussion

As in Experiment 1, there was no significant effect of correctness [$t(39) = 0.68$, $p = .50$; $M_{COR} = 4.15$ (SD = 0.59); $M_{ERR} = 4.10$ (SD = 0.65)]. Again, a reanalysis excluding all self-identified musicians rendered the same pattern of results. Because this comparison was nonsignificant, we ended the pilot study without collecting data on a B-Exposed group.

These results indicate that the range of test items presented in Experiment 1 cannot by itself explain listeners' failure to discriminate correct from incorrect items. After one session of exposure, adults are still unable to indicate knowledge of the statistical property of even a more restricted range of items. Accordingly, in Experiment 3, we evaluated the hypothesis that increased exposure would enhance item discriminability. In this experiment, we presented listeners with the full exposure corpus on two consecutive days, testing them only on the second day. Again, to guard against range effects, we only asked listeners to rate test items originating from their exposure system.

4. Experiment 3

Experiment 3 investigated whether increased exposure would enable participants to discriminate Correct from Error items. This time, listeners came into the lab on two separate occasions on consecutive days. During the first session, they were given only the exposure task. In the second session, they were given both the exposure task and the test. As in Experiment 2, only within-system items were included. We predicted that, given more time to absorb the system's regularities, listeners in this task would give higher similarity ratings to Correct than to Error items.

4.1. Method

4.1.1. Participants

Sixty-one undergraduates participated in Experiment 3 for course credit. One participant was excluded from analysis because he did not complete the second session. The sample contained 14 self-defined musicians and 47 self-defined nonmusicians. Participants reported a mean of 5.0 years of experience performing music ($SD = 3.8$) and a mean of 1.4 semesters studying music theory ($SD = 3.5$).⁴ Forty-one subjects reported having taken zero semesters of music theory.

4.1.2. Apparatus

Same as Experiment 1.

4.1.3. Stimuli

Same as Experiment 1.

4.1.4. Procedure

Participants were exposed to one of two corpora of 100 phrases: System A or System B. The exposure phase lasted two sessions, and in each of these sessions, listeners heard the entire 100-phrase corpus once. Participants were tested at the end of the second exposure section. Like Experiment 2, they received only the test items drawn from their own system: A-Exposed subjects received only A-Correct and A-Error test items, and B-Exposed subjects received only B-Correct and B-Error test items. Three practice items preceded the test items.

4.2. Results and discussion

4.2.1. Analysis of variance

Preliminary analysis of variance showed no differences between the groups exposed to Systems A and B (all group effect F s < 1). We thus collapsed the data from A-Exposed and B-Exposed subjects into within-system correct and error items, as in Experiment 1 (see Fig. 5). Unlike Experiments 1 and 2, there was a significant effect of correctness ($t[59] = 4.13$, $p < .001$; $M_{COR} = 4.5$ [$SD = 0.71$]; $M_{ERR} = 4.2$ [$SD = 0.77$]). Again, a reanalysis excluding all self-identified musicians rendered the same pattern of results.

Taken together, the results of these three experiments suggest that given sufficient exposure, adults are capable of using statistical information to distinguish between good and bad exemplars of a particular style of music. We can be fairly confident that no intrinsic differences between correct and error items within the two systems can confound our interpretation, because the participants in Experiments 1 and 2 were unable to differentiate between Correct and Error items. However, given additional exposure, the participants in Experiment 3 responded differentially to items with and without illegal transitions. Furthermore, while the between-system discrimination evinced in Experiment 1 could be explained using cadential information alone, such an explanation for the results of Experiment 3 is unlikely,

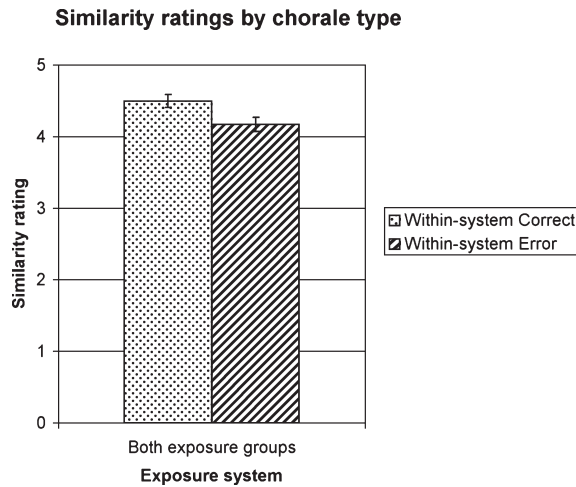


Fig. 5. Experiment 2: Similarity ratings (± 1 SE) by item type ($n = 60$).

as most of the illegal items in each system had legal final transitions, making cadential formation a relatively insensitive cue.

4.2.2. Item analyses

Our intent in constructing these stimuli was to investigate listeners' ability to learn the chord transitions in a novel system of music. However, it may be objected that rather than learning these chord transition probabilities, our learners might have listened for melodic regularities instead. Using multiple regression, we examined the extent to which these factors can predict averaged subject judgments for each stimulus. To represent the harmonic legality of each test item, we computed its average transitional probability (TP) by assessing the probability of each of its chord transitions with respect to the relevant corpus and taking the mean value. To evaluate the melodic "weirdness" of the test items, as we did for the exposure items in Experiment 1, we again used Eerola and North (2000) weighted sum of melodic expectancy factors to calculate the complexity of the top voice of each test item (MelComp).

We submitted average similarity scores given to A items by A-Exposed listeners to a two-variable regression, using TP and MelComp as predictors. Preliminary analysis of Cook's d and DFBETA values suggested the removal of three outlier items (Cohen, Cohen, West, & Aiken, 2002). Our final model reached significance ($F[2,24] = 5.85$, $p = .009$; $R^2_{\text{adj}} = .271$). TP was a significant predictor, but MelComp was not (standardized β s: $\beta_{\text{TP}} = 0.58$, $p = .003$; $\beta_{\text{MelComp}} = 0.063$, $p = .72$). Next, we likewise regressed average similarity scores given to B items by B-Exposed listeners on TP and MelComp. We removed three outliers with excessive Cook's d and DFBETA values. Our final model reached significance ($F[2,24] = 3.53$, $p = .045$; $R^2_{\text{adj}} = .162$). As in the A model, TP was a significant predictor, but MelComp was not (standardized β s: $\beta_{\text{TP}} = 0.43$, $p = .03$; $\beta_{\text{MelComp}} = 0.27$,

$p = .15$). These analyses support the hypothesis that listeners were able to make use of the chord transitional probabilities in these artificial musical systems.

One might also wonder whether, within Error items alone, the recency of an unlikely transition affected listeners' judgments. For each error item, we determined the distance between the last illegal transition and the end of the item. We used these values to predict average similarity ratings of error items only, performing separate regressions for A and B items. After the removal of one outlier from each model based on Cook's d and DFBETA statistics, both models were significant (System A: $F[1,12] = 14.74$, $p = .002$, $R^2_{\text{adj}} = .514$; System B: $F[1,12] = 4.89$, $p = .047$, $R^2_{\text{adj}} = .230$). This provides additional evidence that improbable chord transitions, particularly later in a sequence, can affect listeners' responses to music. It is possible that this finding reflects a recency effect on memory for errors. It is also possible that the effect is due to key-finding. Each phrase was presented in a different key, which meant that listeners needed to determine the key of each phrase in order to discern chord functions (I, V, etc). Since key-finding is best modeled as a process involving a window of at least a few beats (Schmuckler & Tomovski, 2005), it seems plausible that key percepts within these relatively short progressions might be more stable after at least a few chords have passed. Given that the relevant statistics concerned within-key chord functions, we would expect that errors that violate the TPs of the chord sequences would be most salient later in phrases, rather than earlier. A more detailed investigation of these competing hypotheses, using larger test corpora, may be a fruitful direction for further research.

5. General discussion

The results of these experiments suggest that learners of a new musical system can rapidly learn to discriminate well-formed from ill-formed items, and that their judgments are related to serial statistical properties of the newly learned system. In Experiment 1, subjects were able to differentiate items from two new systems after a short exposure, but they were unable to detect errors. We reduced the range of test items in Experiment 2, but this change alone was not enough to enable discrimination of correct and error items. However, the combination of reduced test item range and additional exposure provided in Experiment 3 allowed them to make this distinction, discriminating novel correct items from those containing incorrect transitions. Notably, this successful discrimination cannot be due to inherent preferences for particular test items, as participants in Experiments 1 and 2 failed to make these same distinctions between correct and incorrect items.

What was important about the extra exposure that allowed our 2-day listeners to succeed where others failed? Given the difference in tonal sophistication that has been reported between experienced and inexperienced listeners (Castellano et al., 1984), one possibility is that the extra exposure helped to solidify our 2-day listeners' understanding of the new tonal hierarchy. It is important to note that while extraction of some stable key structure from our stimuli is probably necessary for consistent performance on our task, it is not imperative that listeners center on any *particular* key structure. However, assimilation of the new data into the tonal Western system would likely make the task harder, as its statistics conflict with

those we wish subjects to learn. A deeper understanding of our Phrygian tonality, such as might be generated by repeated exposures to the corpus, could help subjects to separate what they know about the new system from their knowledge of Western music. A complementary possibility, suggested by recent evidence about the role of sleep in memory consolidation (Payne, Ellenbogen, Walker, & Stickgold, 2008), is that the relative timing of the two sessions may have played a role. Provocative results with infants suggest that sleep may promote linguistic generalization (Gomez, Bootzin, & Nadel, 2006), and other types of implicit musical learning have been shown to depend on sleep (Gaab, Paetzold, Becker, Walker, & Schlaug, 2004). Further experiments are needed to disentangle these possibilities.

In this study, we set out to investigate whether cues thought to be useful in language learning may also be helpful in acquiring musical structure. Specifically, we examined listeners' sensitivity to the predictability of pairs of harmonic elements, a process akin to discovering predictive relationships between word categories—an important building block of linguistic grammar. The predictive relationships we observed between the transition probabilities within a sequence, the recency of improbable transitions, and similarity ratings suggest that listeners are attuned to the sequential statistics of chord functions, implying that they are able to extract and use tonality information for understanding the structure of new sequences. Importantly, we do not claim that the statistics of interest in this experiment are the only relevant factors underlying subjects' musical knowledge. Clearly, "correctness" within a particular idiom depends on many additional factors. In Western music, these may include melodic continuity, rhythmic regularity, higher-order harmonic transitional probabilities, harmonic rhythm, the clarity and independence of inner-voice movement, and the metrical location of salient chords, among others (e.g., Krumhansl et al., 2000; Lerdahl & Jackendoff, 1983). However, first-order transitional probabilities are one cue that may be useful when learning harmonic structure, even after a short exposure.

Having drawn a comparison between the mechanisms used to learn language and music, we must now face the question of what we mean by the word "mechanism." This workhorse term in cognitive science can refer either to an algorithm used for problem solving or to the neural hardware over which such an algorithm is implemented. The results of these experiments do not compel one to accept either interpretation. It is conceivable, and in fact quite plausible, that similar algorithms for tracking the mutual predictability of elements are implemented throughout diverse regions of the brain, and perhaps are used for distinct purposes in each (Conway & Christiansen, 2005). On this view, different but parallel learning mechanisms may subservise musical and linguistic learning. Certainly, the preponderance of evidence on primary auditory processing of musical and linguistic sounds by nonmusicians suggests that, at least in the early stages, some separation of function is the norm (Tervaniemi et al., 1999; Tervaniemi et al., 2000). However, there is also evidence to suggest some overlap in the usage of cortical tissue to process the well-formedness of sequences, be they linguistic or musical (Maess et al., 2001; Patel et al., 1998). Further research using imaging techniques to examine the neural substrates underlying the acquisition of musical and linguistic structural knowledge may shed further light on the extent to which the faculties responsible for these two competencies overlap.

Notes

1. Throughout the paper, we use capital numerals to refer to major, minor, and diminished chords.
2. Because of the constrained relationship between A and B items (viz., each B item is the reverse of one A item), this and all similar tests were treated as paired comparisons.
3. This value was inflated by one outlier who reported having studied for 18 semesters.
4. This value was inflated by two outliers who reported having studied for 14 semesters or more.

Acknowledgments

We thank Erika Mikulic and Garrett Lee for assistance with data collection; Carol Krumhansl for helpful suggestions; John Curtin for statistical advice; Chris Alfeld for technical assistance; and Caroline Palmer, Aniruddh Patel, Erik Thiessen, and two anonymous reviewers for valuable comments on a previous draft. This research was supported by graduate fellowships to Erin McMullen Jonaitis from the University of Wisconsin—Madison, NSF, and the Beinecke Bros. Foundation, and by grants to Jenny R. Saffran from NICHD (HD37466) and NSF (BCS-9983630).

References

- Aslin, R., Jusczyk, P., & Pisoni, D. B. (1998a). Speech and auditory processing during infancy: Constraints on and precursors to language. In R. Siegler (Ed.), *Handbook of child psychology* (5th edition, Vol. 2, pp. 147–198). New York: Wiley.
- Aslin, R., Saffran, J., & Newport, E. (1998b). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321–324.
- Bharucha, J., & Stoeckig, K. (1987). Priming of chords: Spreading activation or overlapping frequency spectra? *Perception and Psychophysics*, 41(6), 519–524.
- Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1), 159–171.
- Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research*, 17, 289–308.
- Castellano, M., Bharucha, J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, 113(3), 394–412.
- Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd edition). Mahwah, NJ: Erlbaum.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257–271.
- Conway, C., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 24–39.

- Creel, S., Newport, E., & Aslin, R. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1119–1130.
- Eerola, T., & North, A. C. (2000). *Expectancy-based model of melodic complexity*. Paper presented at the The Sixth International Conference on Music Perception and Cognition, Keele, Staffordshire, UK.
- Eerola, T., & Toiviainen, P. (2004). *MIDI toolbox: MATLAB tools for music research*. Jyväskylä, Finland: Jyväskylän yliopisto.
- Gaab, N., Paetzold, M., Becker, M., Walker, M. P., & Schlaug, G. (2004). The influence of sleep on auditory learning: A behavioral study. *Neuroreport*, 15(4), 731–734.
- Gomez, R., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science*, 17(8), 670–674.
- Gomez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70, 109–135.
- Hannon, E., & Trehub, S. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, 16(1), 48–55.
- Hunt, R., & Aslin, R. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology*, 130(4), 658–680.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch* (Vol. 17). New York: Oxford University Press.
- Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual folk hymns: Convergence of statistical, behavioral, and computational approaches. *Music Perception*, 17(2), 151–195.
- Krumhansl, C. L., & Shepard, R. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579–594.
- Krumhansl, C. L., Toivainen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, 76(1), 13–58.
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, 20(1), 24–37.
- Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, 4(5), 540–545.
- Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77–80.
- Mavromatis, P. (2005). A hidden Markov model of melody production in Greek church chant. *Computing in Musicology*, 14, 93–112.
- McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, 21(3), 289–311.
- McMullen, E., & Saffran, J. (2005). *Statistical learning of a complex musical system*. Poster presented at the Society for Music Perception and Cognition, La Jolla, CA.
- Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition*, 30(5), 678–686.
- Morgan, J., Meier, R., & Newport, E. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19(4), 498–550.
- Morgan, J., Meier, R., & Newport, E. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language*, 28(3), 360–374.
- Morgan, J., & Newport, E. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning & Verbal Behavior*, 20(1), 67–85.

- Narmour, E. (1991). The top-down and bottom-up systems of musical implication: Building on Meyer's theory of emotional syntax. *Music Perception*, 9(1), 1–26.
- Newport, E., & Aslin, R. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162.
- Oram, N., & Cuddy, L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, 57, 103–118.
- Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717–733.
- Payne, J. D., Ellenbogen, J. M., Walker, M. P., & Stickgold, R. (2008). The role of sleep in memory consolidation. In H. L. Roediger III (Ed.), *Learning and memory: A comprehensive reference* (pp. 663–685). Oxford, England: Elsevier.
- Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604–607.
- Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, 100(1), 1–32.
- Raphael, C., & Stoddard, J. (2004). Functional harmonic analysis using probabilistic models. *Computer Music Journal*, 28(3), 45–52.
- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, 6(6), 855–863.
- Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, 18(1), 115–119.
- Saffran, J. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44, 493–515.
- Saffran, J. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47, 172–196.
- Saffran, J. (2003). Mechanisms of musical memory in infancy. In R. J. Zatorre (Ed.), *The cognitive neuroscience of music* (pp. 32–41). Oxford, England: Oxford University Press.
- Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Saffran, J., & Griepentrog, G. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, 37(1), 74–85.
- Saffran, J., Hauser, M. D., Seibel, R., Kapfhammer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, 107(2), 479–500.
- Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27–52.
- Saffran, J., Reeck, K., Niebur, A., & Wilson, D. (2005). Changing the tune: Absolute and relative pitch processing by adults and infants. *Developmental Science*, 8(1), 1–7.
- Saffran, J., & Wilson, D. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy*, 4(2), 273–284.
- Schellenberg, E. G., Adachi, M., Purdy, K., & McKinnon, M. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology: General*, 131(4), 511–537.
- Schmuckler, M. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, 7(2), 109–149.
- Schmuckler, M., & Tomovski, R. (2005). Perceptual tests of an algorithm for musical key-finding. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 1124–1149.
- Smith, J. D., & Melara, R. (1990). Aesthetic preference and syntactic prototypicality in music: 'Tis the gift to be simple. *Cognition*, 34(3), 279–298.
- Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *NeuroImage*, 9, 330–336.
- Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, 10(2), 74–79.

- Thiessen, E., & Saffran, J. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.
- Tillmann, B., Bharucha, J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review*, *107*(4), 885–913.
- 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.
- Trainor, L., & Heinmiller, B. (1999). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development*, *21*(1), 77–88.
- Trehub, S., Cohen, A., Thorpe, L., & Morrongiello, B. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 295–301.
- Werker, J., & Lalonde, C. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672–683.
- Zentner, M., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development*, *21*(3), 483–492.