9th International Conference on Music Perception and Cognition

Alma Mater Studiorum University of Bologna, August 22-26 2006

Acquiring new musical grammars – A statistical learning approach

Psyche Loui
Department of Psychology
Center for New Music & Audio Technologies
University of California, Berkeley
Berkeley, CA, USA
psyche@berkeley.edu

David Wessel
Department of Music
Center for New Music & Audio Technologies
University of California, Berkeley
Berkeley, CA, USA

ABSTRACT

In the present study we examine the ability of humans to acquire knowledge of new music. We designed two artificial musical grammars based on a microtonal harmonic system, and created melodies as legal exemplars of each grammar. In three experiments each participant was exposed to melodies from one grammar. Tests were conducted to assess learning, including forced-choice recognition and generalization, pre- and post-exposure probe tone ratings, and subjective preference ratings. In Experiment 1, five melodies were presented for 25 minutes. Participants correctly recognized and preferred melodies they had heard, but failed to generalize their knowledge to new exemplars of the same grammar. In Experiment 2, 15 melodies were presented for 25 minutes. Participants showed some tendency to recognize new melodies in their given grammar, and also showed an increased sensitivity to the statistics of the musical grammar following exposure. In Experiment 3, after exposing participants to 400 melodies for 30 minutes, forced-choice tests showed significant recognition as well as generalization, as well as sensitivity to the relative frequencies of tones in the new musical system. Results suggest that larger sets of exemplar melodies promote the extraction of harmonic regularities underlying the melodies, whereas smaller sets lead to better recognition and are more likely to influence subjective preference.

Keywords

music cognition, statistical learning, artificial grammar, melody, harmony, preference

INTRODUCTION

Music we encounter every day is composed of sequentially and simultaneously presented pitches. While sequential pitches give rise to melody, simultaneous pitches give rise to harmony. Together, melody and harmony are viewed as the horizontal and vertical organizational dimensions in music.

Various studies in the field of music perception and cognition have shown that humans have reliable but implicit knowledge of both melody and harmony. Melodies that end with unexpected or incorrect notes elicit electrophysiological responses (Besson & Faïta, 1995) as well as decreased goodness-of-fit ratings (Krumhansl, 1990). In the vertical dimension, when perceiving chord progressions that violate the principles of traditional Western harmony, e.g. when a chord progression resolves to a chord other than the tonic, the listener's musical expectation is violated, and such an expectation violation can be observed empirically using reaction time (Bharucha & Stoeckig, 1986) as well as electrophysiological (Koelsch et al, 2000) methodologies. These results have led researchers to ask about the source of knowledge in harmony. Some aspects of harmonic knowledge may result from the neurobiological properties of the auditory system (e.g. Tramo et al, 2001). It is also possible that certain features of harmony might be learned via exposure within one's culture. In particular, the perception of stability, as well as tension and resolution in response to chord progressions, might be partially learned via exposure to the relative frequencies and probabilities of sounds in the environment.

The learning of frequencies and probabilities of sounds has also been explored in the context of language acquisition.

Experiments in statistical learning (e.g. Saffran et al, 1996) have shown that after a brief exposure period to an artificial language, humans as young as eight-month-olds are sensitive to the transitional probabilities of one syllable following another (Aslin et al, 1998), and this has been posited to be useful in acquiring language as well as music (Saffran, 1999). Some research has addressed the learning of melodies (Creel et al, 2004) using the statistical learning method, showing that humans are able to recognize melodies following exposure to sequences of tones.

In the present study we investigate the acquisition of harmony as well as melody using a statistical learning approach. To investigate the learning of harmonic knowledge without the influence of prior exposure, we developed artificial musical grammars based on a microtonal tuning system. The artificial grammars consist of chord progressions in a non-Western scale, from which we compose melodies. Thus the new musical grammars include both vertical and horizontal dimensions of music, and while they are designed to follow some principles of existing musical systems, they are completely new in the sense that the participants of our studies have not heard any music composed using these materials.

A NEW MUSICAL SYSTEM

In the traditional Western musical scale, frequencies of notes are based around an octave, which is defined as a 2:1 ratio in frequency. Within the octave, the Western scale is divided into 12 logarithmically even increments. Thus, the frequencies of tones spanning an octave in the Western scale are defined as the following:

Frequency (Hz) = $k*2^{n/12}$

where n is the number of steps along the chromatic scale, and k is a constant and typically equals 440Hz.

For the following experiments we used the Bohlen-Pierce scale (Walker, 2001), a microtonal tuning system based on 13 logarithmically even divisions of a *tritave*, which is a 3:1 ratio in frequency. The tones in one tritave of the Bohlen-Pierce scale are defined as:

Frequency (Hz) = $k*3^{n/13}$

where n is the number of steps along the tritave scale, and k is a constant which equals 220Hz.

From the above formula we defined the following pitches in one tritave of the Bohlen-Pierce scale (see Table 1):

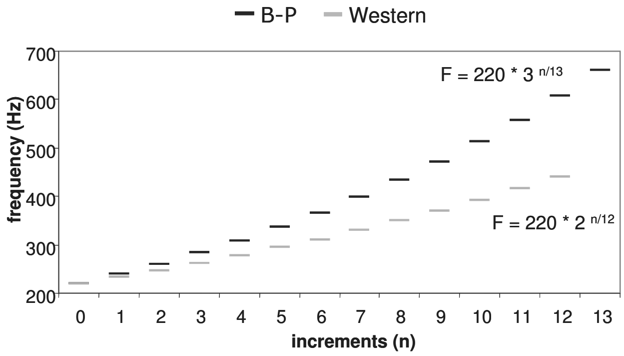

Figure 1. Frequencies along the Bohlen-Pierce scale and the Western scale.

Based on this scale, it was possible to define chords using pitches with frequencies that were approximately related to each other in low-number integer ratios, which are relatively consonant psychoacoustically. One "major" chord in this new system is defined as a set of three pitches with frequencies that approximate a 3:5:7 ratio (see Krumhansl, 1987, for a derivation of chords in the Bohlen-Pierce scale). We composed two sets of chord progressions, where each chord progression consisted of four chords, and each chord consisted of three pitches. Table 1 lists the pitches in each chord, with the frequency of each pitch given by substituting the pitch number into n in the tritave formula: Frequency (Hz) = $220 * 3^{n/13}$.

Table 1. Pitches in each chord progression.

	Pitch number			
Grammar I	10	7	10	10
	6	4	7	6
	0	0	3	0
Grammar II	10	10	7	10
	6	7	4	6
	0	3	0	0

These chord progressions formed the bases of the two sets of grammatical rules which we used to form the stimuli of the present study. By applying these rules in the fashion of a finite-state grammar, we composed sets of melodies from the above chord progressions. Figure 2 illustrates the chord progressions as two finite-state grammars, whereas Figure 3 provides one example of using Grammar I to generate a melody.

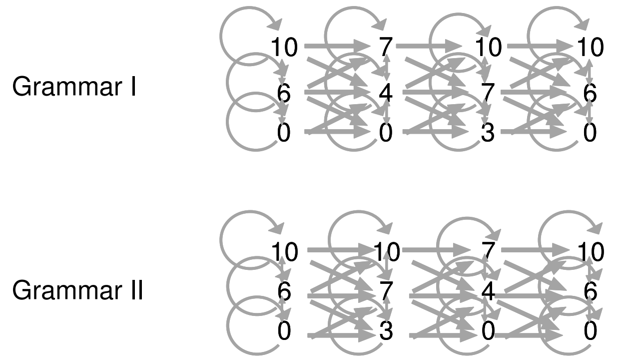

Figure 2. A finite-state grammar diagram illustrating the possible ways to compose melody from harmony.

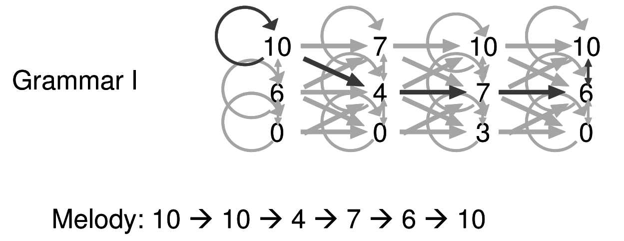

Figure 3. An illustration of applying a finite-state grammar as a set of rules to compose one melody based on finite-state Grammar I. Dark arrows illustrate the paths taken, whereas light arrows illustrate the other possible paths that are legal in the grammar. The resultant melody is shown at the bottom of the figure.

Each melody was composed according to one of the two grammars. For melodies used in Experiments 1 and 2, the following additional constraints were imposed:

- Each melody ranged from four to eight notes.
- Large intervals in one direction were followed by small intervals in the other direction (see Narmour, 1990).

These constraints were observed for Experiments 1 and 2, but not for the 400 melodies used in Experiment 3.

Taken together, this artificial musical system adheres to many of Lerdahl and Jackendoff's rules for a generative theory of tonal music (Lerdahl & Jackendoff, 1983). For example, the musical grammars are parsable, groupable, hierarchical, and statistically predictable. This makes the system viable as a new compositional tool, yet capable of generating melodies that are completely unfamiliar to all participants of our studies.

Using these two finite-state grammars as compositional rules, a small subset of melodies can be legal exemplars of both Grammar I and Grammar II. These melodies were not used in Experiments 1 and 2, but were included in Experiment 3. The musical grammars described here can support important investigations not only in music cognition, but also in statistical learning more generally. Materials constituting our finite-state grammar are such that the items themselves (the tones) are non-rehearsable in the sense that one cannot easily designate a verbal label to each individual tone. Unlike existing artificial grammar studies, such as those using artificial speech syllables (e.g. Gomez & Schvaneveldt, 1994), these artificial musical grammars force the cognitive system to represent items in a nonlinguistic manner, thus minimizing the possibility of rote memorizing the items of an artificial grammar through covert verbalization of its exemplars.

Having defined two sets of artificial harmonic grammars based on a new tuning system, we now have a tool to answer many research questions. We present three experiments in which melodies composed in the harmonic grammars are presented repeatedly to participants, and pre- and post-exposure tests are conducted to assess the extent to which participants acquired their grammar. In particular, we ask the following questions regarding learning:

1. Can participants recognize melodies (grammatical items) they had heard?
2. Can they take their knowledge of harmony, expressed here as a set of grammatical rules, and generalize this knowledge towards new melodies?
3. Can they form expectations for the underlying statistics of the new musical system?
4. Can they learn to form preferences for any aspects of the musical grammars?

To address questions 1 and 2, we employ two-alternative forced choice tests of recognition and generalization, where participants hear familiar and unfamiliar melodies in both grammars, and are asked to choose the more familiar melody. Question 3 is addressed using probe-tone tests before and after the exposure phase, where participants hear a melody followed by a probe tone, and rate the degree to which the probe tone fits the preceding melody. Finally, to answer question 4 we conduct a subjective preference rating task where participants rate their preference of each melody after hearing it.

EXPERIMENT 1

Method

Subjects

24 undergraduate students from the University of California at Berkeley participated in this study for course credit. All participants had normal hearing and more than five years of musical training. Each subject was randomly assigned to one of the two grammars.

Stimuli

All auditory stimuli were generated and presented using Max/MSP (Zicarelli, 1998) and presented via headphones at a level of 70dB. 20 individual melodies (10 in each

ISBN 88-7395-155-4

grammar) were constructed and presented in pure tones ranging from 220Hz to 660Hz, spanning one tritave in the current musical system. Each tone lasted 500ms, with envelope rise and fall times of 5ms each. The musical grammars and melodies generated from them are described in the previous section.

Procedure

Experiments were run in a sound-attenuated chamber. Each experiment was run in five phases:

- Pre-exposure probe tone ratings: A melody was presented followed by a probe tone. The participant's task was to rate how well the probe tone fit the preceding melody, on a scale of 1 to 7 (where 1 was least fitting and 7 was best fitting).
- Exposure: The second phase consisted of five melodies generated from the assigned grammar being played in randomized order for 25 minutes. Equal numbers of participants were exposed to the two grammars. Each tone lasted 500ms, and a 500ms silence separated any two melodies. Each melody was repeated 100 times over the course of exposure. While listening to the melodies, participants were given the option of drawing on provided paper as a distracter task.
- Forced-choice recognition and generalization: In the third phase, participants were given a two-alternative forced-choice task. The first part of the forced-choice trials tested for participants' recognition of melodies they had heard during the exposure phase, whereas the second assessed whether they could generalize their recognition to novel melodies in the same grammar. One block of five recognition trials preceded another block of five generalization trials. In both types of trials, two melodies were played one after another, with an inter-onset interval of 5 seconds. Participants' task was to indicate which melody (the 1st or 2nd) sounded more familiar. The alternative melody (the incorrect answer) was always a melody drawn from the alternative grammar. Thus, the right answer for one group of participants was the wrong answer for the other group.
- Post-exposure probe tone ratings: In order to measure the effect of exposure, (i.e. whether or not participants had learned anything about melodic structure) we gave them a second probe tone rating task after exposure. This task was identical to the first probe tone task, allowing direct comparison between the two sets of data.
- Preference ratings: To assess the degree to which learning influences musical preference, the fifth and final block consisted of preference ratings: participants rated each melody in both grammars (20 melodies in all) after it was played once. Ratings were on a scale of 1 to 7, with 1 being the least preferable and 7 being the most preferable.

Results

Figure 4 shows the data from the recognition and generalization tasks, plotted as percent correct. A participant's response was scored as correct when they selected either the melody that they had heard during exposure (recognition items) or the novel melody generated with the same grammar as their exposure melodies (generalization items). When the participant selected the melody generated by the other, non-trained grammar, their answer was scored as incorrect.

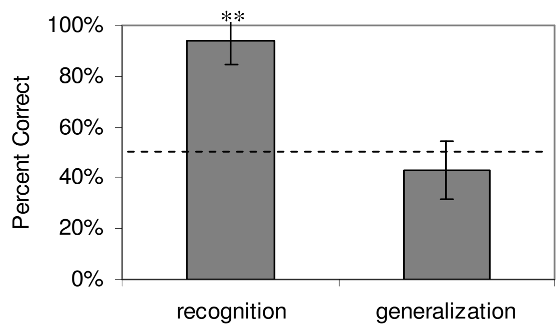

Figure 4. Two-alternative forced choice results from Experiment 1.[1]

As is evident from the figure, participants were significantly above chance in identifying the melodies they had heard, but were not above chance in being able to identify new instances of the same grammar as more familiar.

Pre- and post-exposure probe tone ratings both revealed some sensitivity to the statistics of the melodic structure. Participants' judgments of individual probe tones were correlated with the melodic structure present in their input, and there was a trend towards an increased correlation after the exposure.

Results from the preference ratings revealed some effects of exposure on subjective preference. Participants exposed to one of the grammars rated the melodies they had heard as being significantly more preferable than melodies they had not heard. A similar trend was observed in the other group, but was not significant. Ratings for novel melodies from the familiar grammar were not reliably different from ratings for melodies in the unfamiliar grammar (see Fig. 5), suggesting that a preference change was associated with familiarity but did not extend to other instances of the grammar.

[1] For all figures, ** = $p<0.01$; * = $p<0.05$

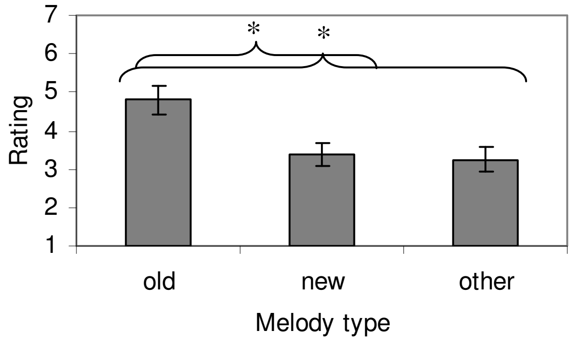

Figure 5. Preference ratings.

Conclusion

Following a 25-min exposure to 5 melodies, participants could recognize melodies they had heard, but were unable to generalize their knowledge to new melodies in the same grammar. One group of participants also rated the melodies they had heard as being more preferable, suggesting that familiarity may influence preference.

EXPERIMENT 2

Method

Subjects

24 undergraduate students participated in this study. Recruitment criteria were the same as Experiment 1. None of the participants in Experiment 2 had been in Experiment 1.

Stimuli

In addition to the 10 melodies for each grammar from Experiment 1, 10 new melodies were generated for each of the two grammars. All other stimulus parameters were the same as Experiment 1.

Procedure

Experiment 2 was identical to Experiment 1 in procedure, except for the following modifications:

- Pre-exposure probe tone ratings: a different melody (i.e. one not presented during the exposure) was used to obtain probe tone profiles in both pre- and post-exposure ratings.
- Exposure: we increased the set of presented melodies from five to 15. Each melody was presented 33 times over 25 minutes of exposure.
- Forced-choice recognition and generalization: 10 more trials were added to the recognition block, such that this phase of the experiment contained 20 trials, 15 in the recognition block and 5 in the generalization block.
- Post-exposure probe tone ratings: this was identical in materials and procedure to the Pre-exposure probe tone ratings phase of Experiment 2.
- Preference ratings: participants rated each of the 40 melodies (20 in each grammar) for their preference, using the same scale and procedure as Experiment 1.

Results

Figures 6 and 7 show results from the recognition and generalization tasks following exposure to 15 melodies. Data are shown separately for the two groups of participants based on the presentation grammar due to a significant main effect of grammar. Forced-choice tests showed that both groups of participants recognized the presented melodies. In addition, participants exposed to Grammar I significantly generalized their familiarity to the new instances of the same grammar (see Fig. 6), whereas the group exposed to Grammar II did not generalize their familiarity to new melodies (see Fig. 7).

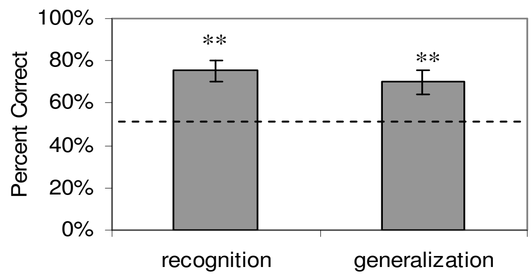

Figure 6. Grammar I forced choice results.

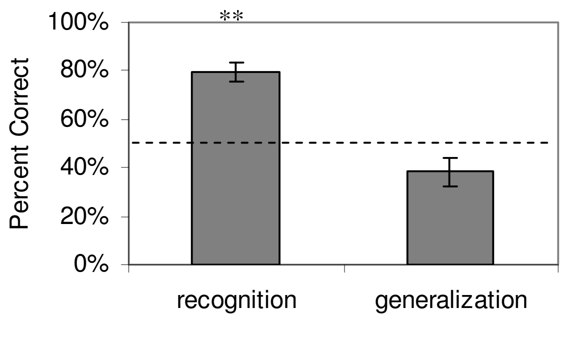

Figure 7. Grammar II forced choice results.

Probe tone tests revealed that subjects were able to pick up on the statistics of the corpus of melodies. Both the pre-exposure ratings (Fig. 8) and the post-exposure ratings (Fig. 9) were correlated significantly with the melody set, but the post-exposure ratings had significantly higher correlation. Figure 10 compares the correlations of the probe tone profiles from Figure 8 (exposure versus pre-exposure ratings) and Figure 9 (exposure versus post-exposure ratings), showing the relative statistical sensitivity reflected by pre- and post-exposure ratings. An increase in correlation is observed, suggesting that given 25 minutes of exposure to 15 exemplars, subjects became more sensitive to the underlying statistics of the new musical system.

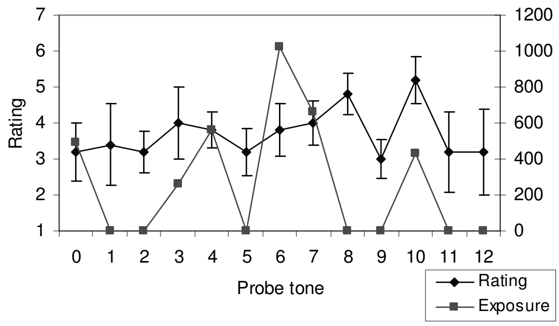

Figure 8. Pre-exposure probe tone ratings versus frequencies of occurrence of pitches during exposure.

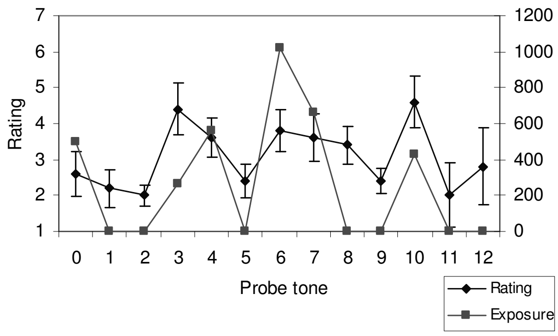

Figure 9. Post-exposure probe tone ratings versus frequencies of occurrence of pitches during exposure.

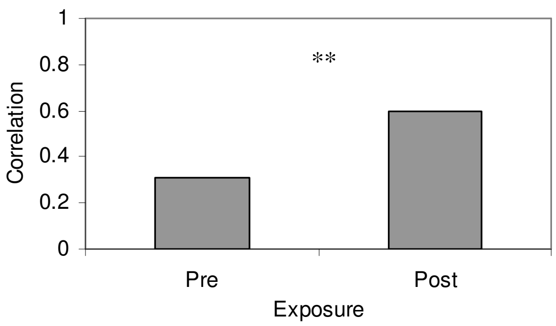

Figure 10. Correlations between profiles of probe tone ratings and frequencies of exposure, before and after exposure.

Preference ratings now did not show any differentiation between old, new, and other-grammar melodies (see Fig. 11), suggesting that familiarity was necessary for a change in preference.

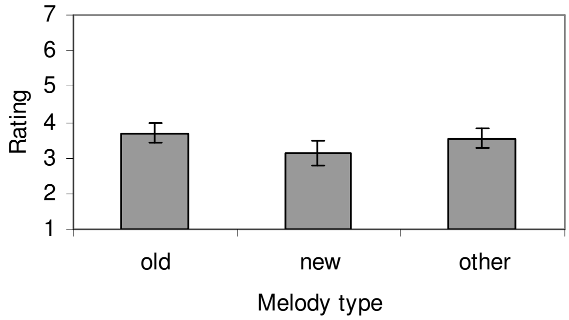

Figure 11. Preference ratings from Experiment 2.

Conclusion

After a 25-minute period of exposure to 15 melodies, participants were above chance at recognizing previously-encountered melodies. Partial generalization was also observed, as one group of participants significantly identified new melodies generated from the same grammar as being more familiar. In addition, a comparison of pre-exposure and post-exposure ratings showed that post-exposure ratings were significantly more highly correlated with the profile of relative frequencies of exposure of each tone in the new musical system. We found no increase in preference following exposure to 15 melodies, suggesting that more repetitive exposure to fewer melodies is more likely to influence preference.

Comparing Experiment 2 to Experiment 1, we note that increasing the number of exposure melodies from five to 15 seems to have aided participants in learning the underlying harmonic structure, such that generalization became somewhat plausible. To further test the hypothesis that increasing the number of melodies at exposure aids generalization, we increased the number of melodies drastically for the next experiment.

EXPERIMENT 3

Methods

Subjects

24 undergraduate students participated in this study. Recruitment criteria were the same as Experiment 1. None of the participants in Experiment 3 had been in Experiment 1 or 2.

Stimuli

500 new melodies were generated for each of the two grammars. Each melody was eight notes in length. There were no constraints on the interval size of these melodies, nor were there any constraints on the overlap between the groups of melodies that could be generated from the two grammars. Otherwise, all stimuli properties were similar to Experiment 1.

ISBN 88-7395-155-4

Procedure

Experiment 3 was identical to Experiment 2 in procedure, except for the following modifications:

- Pre- and post-exposure ratings: the melody used for pre- and post-exposure ratings was a different melody from the one used in Experiment 2, but it was identical between the pre- and post-exposure ratings for this experiment.
- Exposure: 400 melodies, each containing eight tones, were presented in random order with no repeats. The entire exposure phase lasted for 30 minutes.
- Forced-choice recognition and generalization: this phase of the experiment contained 20 trials, 15 in the recognition block and five in the generalization block. In the recognition block, 15 previously-encountered melodies, drawn from the 400 melodies in the exposure phase, were foiled against 15 melodies in the other grammar. In the generalization block, five previously-unencountered melodies in the familiar grammar were foiled against five melodies in the other grammar.
- Preference ratings: participants rated 40 melodies (20 in each grammar) for their preference, using the same scale and procedure as Experiment 1.

Results

Results were averaged across groups because no differences were significant between the two groups. Figure 12 shows results from the recognition and generalization tasks following exposure to 400 melodies. Forced-choice tests showed that both groups of participants recognized and generalized the presented melodies. Figures 13 and 14 show the pre-exposure and post-exposure ratings compared to the exposure frequencies. A significantly higher correlation is observed between post-exposure ratings and exposure frequencies compared to pre-exposure ratings. Figure 15 shows the preference ratings, where both groups showed no differences between old melodies, new melodies, and melodies in the other grammar.

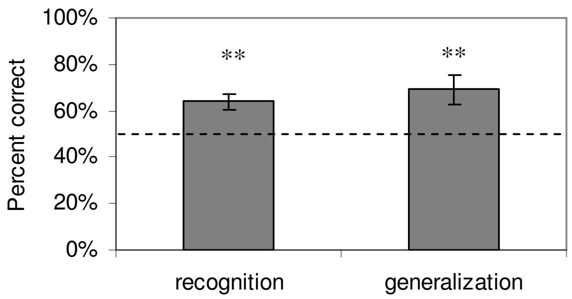

Figure 12. Two-alternative forced choice results for recognition and generalization from Experiment 3.

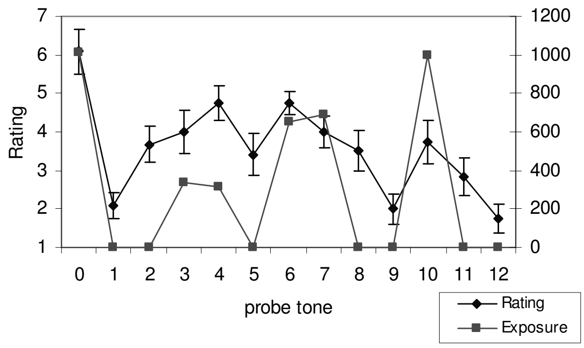

Figure 13. Pre-exposure ratings versus exposure in Experiment 3.

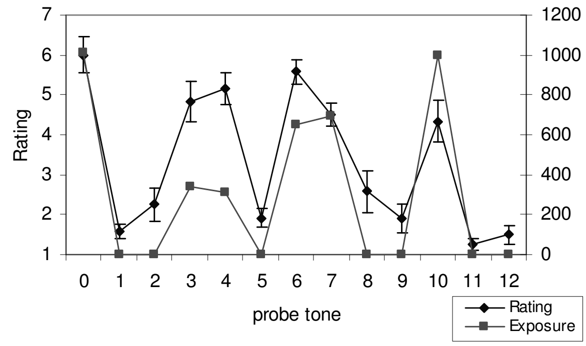

Figure 14. Post-exposure ratings versus exposure in Experiment 3.

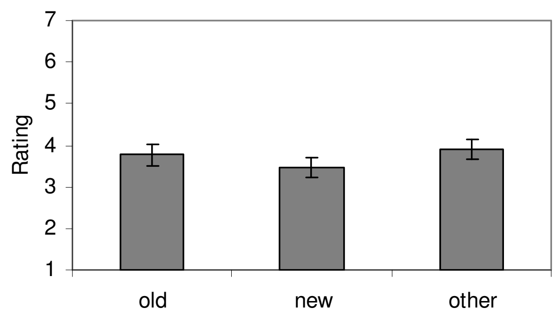

Figure 15. Preference ratings for melodies in Experiment 3.

Conclusion

Given exposure to a large number (400) of melodies, participants were able to recognize old melodies as well as generalize their knowledge to new melodies composed from the same harmonic grammars. A comparison of pre- and post-exposure ratings again show an increased correlation between post-exposure ratings and the relative frequencies of the tones presented during exposure.

GENERAL DISCUSSION

Experiment 1 showed that after participants were exposed repeatedly to five melodies for 25 minutes, they recognized and preferred melodies they had heard, but could not generalize their knowledge to new melodies composed in the same grammar.

In Experiment 2 we exposed participants to 15 melodies for the same overall length of time. Thus, participants heard each individual melody fewer times than participants in Experiment 1. This increase in exemplars and reduction in repetition affected participants' performance in the forced-choice tests: participants now showed not only reliable recognition, but also some evidence of generalization. Learning of the musical system was also demonstrated in the probe tone task, where the increase in correlation for post-exposure ratings demonstrated that participants formed expectations for tones based on the relative frequencies with which they were heard during exposure.

In Experiment 3 we exposed participants to a much larger corpus of 400 melodies with no repeats for 30 minutes. With this large increase in exemplars and no repetition, participants now showed reliable generalization as well as recognition. The increase in correlation for post-exposure ratings again demonstrated that participants were sensitive to the underlying statistics of the grammar.

In summary, these experiments suggest that given limited exposure to a novel musical system, humans can exhibit rapid statistical learning to develop expectations that conform to the musical grammar. Increasing the number of exemplars of the musical grammar may a) help to enhance sensitivity to the underlying statistics of a grammar; and b) aid learners in generalizing the acquired statistics to new instances of the same grammar. However, increasing the set of exemplars also seems to be detrimental towards preference ratings for familiar melodies. These results may suggest that the subjective experience of musical preference is not directly related to the increase of statistical sensitivity or to the understanding of grammatical structure; instead, preference may be a result of recognition and familiarity. We plan to conduct follow-up studies to further examine the link between familiarity and preference.

We believe that the artificial musical system presented here offers a new method to investigate music perception and cognition, and can also be applied more generally to behavioral and physiological research. Having demonstrated the human ability to acquire harmonic grammars, it is now possible to tease apart the individual components of learning. In ongoing research we have conducted similar studies on participants with no formal musical training; preliminary results suggest that nonmusicians behave similarly as musicians in the learning of the new musical grammars; thus these effects probably do not result from formal musical education. Ongoing work also uses Event-Related Potentials to study the perception of this new musical system; results from these studies may help characterize the neural mechanisms that enable the learning of new contexts.

ACKNOWLEDGEMENTS

The current study was funded by a grant to D.W. from the UC Academic Senate. We thank C. Hudson Kam for help on experiment design and significant contributions to a previous manuscript, to E. Wu and P. Chen for data collection and helpful suggestions on an earlier draft, to C. Krumhansl for the design of the artificial grammar, and to E. Hafter for valuable advice and use of laboratory space.

REFERENCES

Aslin, R., Saffran, J. R., & Newport, E. (1998). Computation of conditional probability statistics by 8-month old infants. *Psychological Science, 9*(4), 321-324.

Besson, M. & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: comparison of musicians with nonmusicians. *JEP: Human Perception & Performance, 21,* 1278–1296.

Bharucha, J.J., Stoeckig K. (1986). Reaction time and musical expectancy: priming of chords. *JEP: Human Perception & Performance, 12, 4,* 403-410.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant Melodies: Statistical Learning of Nonadjacent Dependencies in Tone Sequences. *JEP: Learning, Memory, & Cognition, 30*(5), 1119-1130.

Gomez, R. L., & Schvaneveldt, R. W. (1994). What Is Learned from Artificial Grammars? Transfer Tests of Simple Association. *JEP: Learning, Memory, & Cognition, 20*(2), 396-410.

Lerdahl, F., & Jackendoff, R. (1983). *Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: 'Non-musicians' are musical. *Journal of Cognitive Neuroscience, 12,* 520–541.

Krumhansl, C. L.; Kessler, E. J. (1983). Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys. *Psychological Review, 89(2)*, 334-368.

Krumhansl, C.L. (1987). General properties of musical pitch systems: Some psychological considerations. In J.

Sundberg (Ed.), *Harmony and Tonality*, Stockholm : Royal Swedish Academy of Music, publ. no. 54.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. NY: Oxford University Press.

Meyer, L. (1956). *Emotion and Meaning in Music*. University of Chicago Press.

Narmour, E. (1990) *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.

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

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics, 56,* 125-132

Walker, E. (2001). The Bohlen-Pierce Scale: Continuing Research. http://www.ziaspace.com/elaine/BP/

Zicarelli, D. (1998). An extensible real-time signal processing environment for Max. *Proceedings of the International Computer Music Conference, 1998.*