An Information-Theoretic Account of Musical Expectation and Memory

Kat Agres (kathleen.agres@eecs.qmul.ac.uk)
Centre for Digital Music and Cognitive Science Research Group
School of Electronic Engineering & Computer Science
Queen Mary, University of London
London E1 4NS, United Kingdom

Samer Abdallah (samer.abdallah@eecs.qmul.ac.uk)
Centre for Digital Music
School of Electronic Engineering & Computer Science
Queen Mary, University of London
London E1 4NS, United Kingdom

Marcus Pearce (marcus.pearce@eecs.qmul.ac.uk)
Cognitive Science Research Group, Centre for Digital Music and Centre for Research in Psychology
School of Electronic Engineering & Computer Science
Queen Mary, University of London
London E1 4NS, United Kingdom

Abstract

When listening to music, we form implicit expectations about the forthcoming temporal sequence. Listeners acquire knowledge of music through processes such as statistical learning, but how do different *types* of statistical information affect listeners' learning and memory? To investigate this, we conducted a behavioral study in which participants repeatedly heard tone sequences varying within a range of information-theoretic measures. Expectedness ratings of tones were collected during three listening sessions, and a recognition memory test was given after each session. This enabled us to examine how statistical information affects expectation and memory for tone sequences over a period of increasing exposure. We found significant correlations between listeners' expectedness ratings and measures of information theory (IT), and although listeners demonstrated poor overall memory performance, the IT properties significantly impacted on musical memory. Generally, simple sequences yielded increasingly better memory performance. High-information sequences, for which making accurate predictions is difficult, resulted in consistently poor recognition memory.

Keywords: Music cognition; information theory; computational approach; predictive models.

Introduction

Music is a fruitful domain for exploring the mechanisms responsible for learning structured temporal sequences, a type of learning that subserves a wide range of human behaviors. Research by Krumhansl (1990), Pearce & Wiggins (2006), Huron (2006), and others shows that listeners implicitly acquire knowledge about the statistical structure of music. But is this implicit learning influenced by the information contained in the musical signal and, if so, how? Using computational methods, the pitch structure of music can be manipulated systematically to help reveal the ways in which various information-theoretic properties of melody interact and influence human learning and memory.

This paper examines the process of learning novel music over time, with a focus on mental anticipatory processing and musical structure. By using carefully constructed tone sequences, we are able to test how the statistical structure of music, as measured using information theory, affects the expectedness of tones, as well as memory for specific exemplars, over a period of increasing exposure.

Information Theory and Music

Information theory has contributed to fields as diverse as engineering and linguistics by describing and quantifying the information contained in a signal. This is especially useful for clarifying how the brain processes temporal signals; and indeed, information-theoretic measures such as *entropy*, a measure of uncertainty, have successfully described and predicted how the human brain anticipates forthcoming sensory input, such as music and language (e.g., Manning & Schutze, 1999; Abdallah & Plumbley, 2009). Within the domain of music, there has been a long-standing interest in anticipation and prediction, and statistical and probabilistic approaches to learning have been influential for decades (consider Krumhansl & Kessler, 1982; and Saffran, Johnson, Aslin, & Newport, 1999). Computational models such as IDyOM (Pearce, 2005) derive information-theoretic properties of music that accurately reflect and predict listeners' expectations during music listening (Pearce & Wiggins, 2006; Pearce et al., 2010).

While statistical and computational approaches have modeled human performance on a variety of music perception tasks, these approaches have not yet been extended to modeling the learning *trajectory* of listeners: we do not yet know how information-theoretic measures capture musical learning over increasing exposure to musical exemplars, and how much exposure is necessary to learn the statistical regularities of novel music. The following research addresses these questions.

Behavioral Experiment

In the present study, computational techniques were used to create a set of tone sequences varying systematically across several information-theoretic measures. Varying the sequences' statistical structure allows us to assess which factors have the greatest impact on listeners' musical expectations and memory for tone. We focused on testing the relative influence of three information-theoretic factors based on the information theoretic concepts of *entropy rate*, *multi-information rate (*a kind of redundancy*)*, and *predictive information rate* (see Abdallah & Plumbley, 2009). These measures are defined for a random process with a known probability distribution, and hence thought of as 'objective'. However, listeners cannot know these probability distributions; they can only estimate them from observations, and so we defined variants of each measure appropriate for an observer processing events sequentially as they happen, updating its estimated probability model as it goes along: they are *dynamic* information measures based on an *adaptive* probabilistic model. Since they depend only on the actual observed sequences (rather than a theoretical statistical ensemble) and any prior expectations built into the listener model (which we may think of as summarizing the listener's previous musical experience), we can usefully think of these as 'subjective' information measures.

In our experiments, the listener model was an adaptive first-order Markov chain, as described by Abdallah and Plumbley (2009), which assumes that notes are sampled from a Markov chain with an unknown transition matrix, and tries to estimate the transition matrix from the observations. The model is supplied with an initial expectation (a Bayesian prior) that the transition matrix is similar to a first-order transition matrix derived from a large corpus of Western tonal music in a major key.

The three information measures examined in this paper are *Information Content (IC)*, *Coding Gain*, and *Predictive Information.* IC is a measure of the subjective unexpectedness of an observation. Coding Gain measures how much temporal structure or pattern there is in the sequence. And Predictive Information quantifies how much the current observation improves the listener's predictions about future observations (assuming knowledge of the previous observations). High predictive information is also associated with temporal structure or pattern, but of the sort that has more variation, requiring the observer to continually pay attention in order to follow the pattern.

These three measures are defined in the Markov model as follows: at any integer time t, let x_t be the note occurring at that time, and Θ_t be the estimated transition matrix using information available before t. Then, the IC at time t is the negative log probability of x_t given the context and the estimated model: $-\log p(x_t|x_{t-1}, \Theta_t)$, where the relevant transition probability is extracted from the matrix Θ_t. Coding Gain at time t quantifies how much the model's ability to predict the current observation depends on having observed the preceding observations, and is a difference of log probabilities: $\log p(x_t|x_{t-1}, \Theta_t) - \log p(x_t|\Theta_t)$, where the latter term is derived from the stationary distribution of the transition matrix. Predictive Information is quantified as the distance between two probability distributions over the next symbol x_{t+1}, representing the observer's probabilistic beliefs about x_{t+1} before and after the observation of x_t. The average of each of these three measures was computed for every tone sequence in the present study, henceforth referred to as *sequence statistics*.

To investigate the processes underlying musical learning, listeners were exposed to tone sequences and tested on recognition memory over several listening sessions. In each listening session, participants heard tone sequences and rated the expectedness of a tone (termed the "probe tone") within each sequence. Probe tones varied in terms of information content (representing unexpectedness) across sequences. A recognition memory test followed each listening session. This format enabled us to compute information-theoretic measures for every tone sequence, and compare the effect of these measures on probe tone ratings.

We also examined how IT measures impacted on recognition performance in the test sessions. We hypothesized that sequences featuring generally high-entropy would be difficult to remember, and probe tones would be rated with lower expectedness. Because each tone sequence was presented in every listening session, we also aimed to clarify the learning trajectory for the different classes of tone sequence; that is, how music represented in short-term memory gradually becomes more richly encoded in long-term memory, and how musical information and complexity, as measured using IT, influence this process over time.

Method

Participants

Twenty-three students (12 female and 11 male; mean age = 21.0 yrs) at Cornell University participated in this study for extra credit in a psychology course. The participants had an average of 1.61 years (*SD* = 1.88 yrs) playing music in the previous five years, and an average of 5.82 years (*SD* = 4.54 yrs) of lifetime experience playing an instrument.

Materials and Procedure

After receiving written and verbal instructions, participants listened to tone sequences in three sessions, each lasting approximately 15 minutes and followed by a brief test session. In the listening sessions, participants heard each of the 24 tone sequences (presented in a different order in each session) and were asked to rate the expectedness of a particular tone (the probe tone) within each sequence. This tone was identified visually on the computer screen via a clock counting down on the subsequent tones of the sequence. When the clock returned to midnight, participants rated the expectedness of the concurrently sounding tone on a scale from 1 to 5, where ‘1’ represented highly unexpected and ‘5’ represented highly expected.

Each listening session was followed by a test session. Sixteen test stimuli were presented in each of the three test sessions, where 8 sequences were *Old* (had been presented previously) and 8 were *New*. After each test sequence, participants responded “Yes” or “No” to whether they had heard the sequence before. Upon responding, the listener made a confidence rating on a scale from 1 to 5 where ‘1’ represented not confident and ‘5’ represented very confident.

The 24 sequences of the listening sessions each comprised 24 isochronous tones, played in a piano timbre. Each tone was 500 ms in duration, yielding sequences that were 12-seconds-long each. The sequences were generated with an alphabet of 7 pitches (representing one octave of the diatonic scale). A first-order Markov transition matrix was derived (Pearce, 2005) from the scale degrees of Canadian folk songs/ballads, Chorale melodies, and German folk songs in a major key (the same corpus described in Table 2 of Pearce and Wiggins, 2006). To construct the tone sequences, many transition matrices were generated randomly using a process biased towards the tonal transition matrix. From each matrix, one sequence of 24 notes was sampled. A subset of these was then selected manually to ensure a good spread in the 3-dimensional subjective information space formed by the information theoretic measures described above.

A distinct 500 ms white noise clip was played after every tone sequence in the listening and test sessions as a perceptual “reset” to ensure that expectedness ratings and memory judgments were based only on the current trial. The study was administered on a MacBook Pro laptop, and stimuli were presented and responses collected using Psychophysics Toolbox (Version 3) within the programming environment of MATLAB 2010a (MathWorks, Inc). Participants listened to stimuli over headphones set to a comfortable listening volume.

Results and Discussion

Whole-Sequence IT measures and Expectedness During Listening Sessions

To examine how the information-theoretic properties of each sequence influenced the expectedness of probe tones, correlations were analyzed between the IT factors and *Average Expectedness Ratings*. In terms of whole-sequence statistics, both *Sequence IC* and *Sequence Predictive Information* were significant predictors of *Average Expectedness Ratings*. As shown in the top graph of Figure 1, *Sequence IC* was correlated with *Average Expectedness Ratings* such that more predictable sequences (lower Sequence IC values) yielded higher expectedness ratings of probe tones, $R^2 = .29$, $F = 28.87$, $p < .01$. The second graph of Figure 1 displays the correlation between *Sequence Predictive Information* and *Average Expectedness Ratings*, $R^2 = .34$, $F = 36.23$, $p < .01$. The third graph shows *Sequence Coding Gain* and *Average Expectedness Ratings*, also significant in this analysis, $R^2 = .37$, $F = 41.54$, $p < .01$.

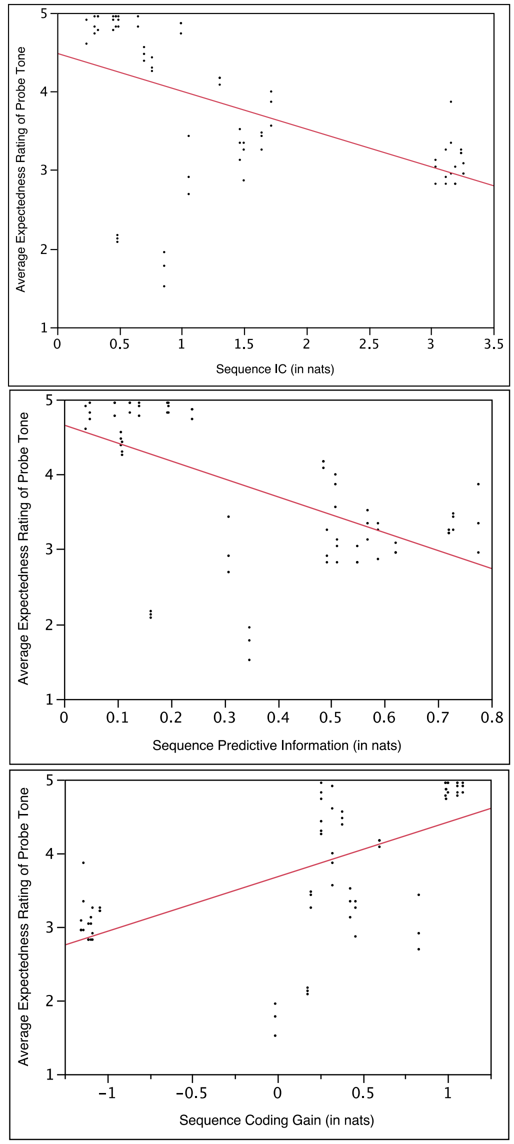

Figure 1: The main effects of *Sequence IC*, *Sequence Predictive Information*, and *Sequence Coding Gain* (in nats, where 1 nat = 1.44 bits) on average expectedness ratings during the listening sessions.

Sequences with high average IC values contain unexpectedness; the tones comprising these sequences have high average Information Content. Therefore, it is logical that sequences containing many unexpected, unpredictable tones would yield lower expectedness ratings as shown below.

Regarding the effects of *Sequence Predictive Information*, information is inextricably associated with unexpectedness: an event cannot be informative if the observer knew it was going to happen, because it will not change the observer's beliefs about the future. (Mathematically, Predictive

Information is upper-bounded by the Information Content.) Hence, sequences with higher average Predictive Information will necessarily have moderately high average information content and thus we would expect the probe tones to see relatively lower expectedness ratings.

Coding Gain is a measure of how much information was gained about the current observation from the preceding context. Therefore, the greater the average Coding Gain of the sequence, the greater the predictability of the sequence and so we would predict higher expectedness ratings in such cases.

Expectedness and Probe Tone IC To examine which factors in the listening sessions had the greatest impact on expectation, a multiple regression analysis was performed with *Probe Tone IC*, *Sequence IC*, *Sequence Coding Gain*, *Sequence Predictive Information*, and *Listening Session* as independent measures, and *Expectedness Ratings* as the dependent measure (note that all expectedness ratings were used, not the average rating for each stimulus). Listeners were included as a random effect in the analysis. There was a significant main effect of *Probe Tone IC*, F = 181.74, p < .001, with high-IC tones rated as less expected. As for the whole-sequence IT measures, there were also main effects of *Sequence IC*, F = 3.92, p < .05, and *Sequence Predictive Information*, F = 9.67, p < .01. In addition to these main effects, there were also significant interactions between *Probe Tone IC* and all three of the IT measures of sequences statistics: *Sequence IC* X *Probe Tone IC*, F = 22.34, p < .001, and *Sequence Coding Gain* X *Probe Tone IC*, F = 35.72, p < .001, and *Sequence Predictive Information* X *Probe Tone IC*, F = 91.65, p < .001. *Listening Session* did not contribute significantly to the results indicating that pitch expectation remained constant overall during the study.

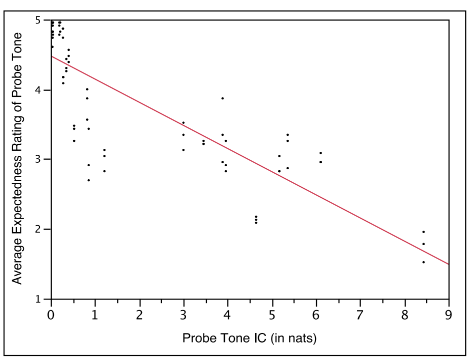

Figure 2: Probe Tone IC as a predictor of average expectedness ratings of probe tones.

Probe Tone IC had the largest effect in the listening sessions, with a significant linear relationship with *Expectedness Ratings*, R^2 = .69, F = 154.20, p < .01. In Figure 2, the average expectedness rating for each melody is shown to display more clearly the main effect on a continuous rather than discrete scale. Low IC tones do receive reliably higher expectedness ratings than high-IC probe tones over the course of listening.

Recognition Memory in Test Sessions

Data from the test sessions are reported in Table 1 as *Proportion Correct Response*. Chance performance is 0.5, and the similarity of performance for *Old* and *New* items indicates little bias towards either response.

Table 1: Recognition memory test performance (proportion correct) for *Old* and *New* sequences across listening sessions.

Listening Session	Old/ Correct (Hits)	Old/ Incorrect (Misses)	New/Correct (Correct Rejections)	New/Incorrect (False Alarms)
Session 1	0.67	0.33	0.64	0.36
Session 2	0.63	0.37	0.65	0.35
Session 3	0.70	0.30	0.65	0.35

Despite little evidence for an increase in overall memory performance over the course of the experiment, we investigated whether certain types of statistical information were being learned, and examine whether performance differed depending on the properties of the individual sequences. Therefore, to examine the effects of the IT measures on recognition scores across listening sessions, a logistic regression was performed with *Sequence IC*, *Sequence Coding Gain*, *Sequence Predictive information*, *Familiarity* (*Old* or *New* stimulus), and *Listening Session* as factors, and *Correct Response* as the binary dependent variable.

All three whole-sequence statistics showed significant main effects: *Sequence IC*, χ^2 = 16.21, p < .01; *Sequence Predictive Information*, χ^2 = 12.09, p < .01; and *Sequence Coding Gain*, χ^2 = 4.27, p < .05. *Listening Session* interacted with each of the whole-sequence IT measures: *Sequence IC* X *Listening Session*, χ^2 = 6.14, p < .05, *Sequence Predictive Information* X *Listening Session*, χ^2 = 7.98, p <.05, and *Sequence Coding Gain* X *Listening Session*, χ^2 = 6.53, p < .05, were all significant interactions.

The only significant interaction including *Familiarity* was with *Sequence Predictive Information*, χ^2 = 12.15, p < .01. As shown in the top plot of Figure 3 below, *New* sequences that are high in Predictive Information yield more correct responses than those with low Predictive Information. Conversely, *Old* sequences show the opposite trend, with worse recognition memory performance on high Predictive Information sequences. Note that *Proportion Correct Response* is used in Figure 3 rather than the categorical variable *Correct Response* for clarity of illustration.

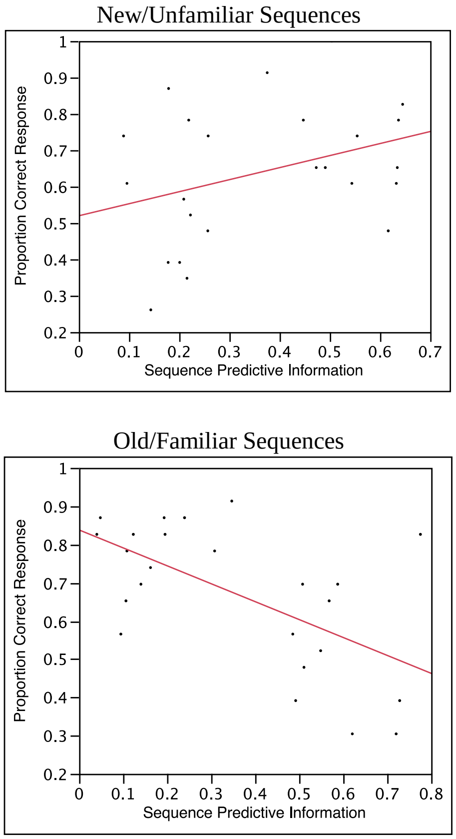

Figure 3: The differential effect of *Sequence Predictive Information* on *Proportion Correct Response* during recognition memory tests for *New* and *Old* sequences.

Confidence Ratings Confidence ratings of recognition memory judgments were collected after every test sequence; responses were made on a 1-5 scale where on a where '1' represented *not confident* and '5' represented *very confident*. A logistic regression was performed with the same factors as those used above: *Sequence IC*, *Sequence Coding Gain*, *Sequence Predictive Information*, *Familiarity* (*Old* or *New* stimulus), and *Listening Session*. This analysis yielded significant effects of *Sequence IC*, $\chi^2 = 16.44$, $p < .01$, and *Sequence Coding Gain*, $\chi^2 = 15.33$, $p < .01$, and interactions of these two factors with *Listening Session*: *Sequence IC* X *Listening Session*, $\chi^2 = 21.94$, $p < .01$, and *Sequence Coding Gain* X *Listening Session*, $\chi^2 = 23.10$, $p < .01$.

As expected, listeners made more confident memory judgments when sequences had lower IC and higher Coding Gain. For *Sequence IC*, there was a decrease in confidence (fewer 4 and 5 responses) over the course of the experiment, which was especially noticeable for low-IC sequences (because high-IC sequences rarely received 5 responses throughout the study). Similarly, there was also a decrease in highly confident ratings (4 and 5 responses) for *Sequence Coding Gain* over the course of the experiment, which was more apparent in the high-Coding Gain sequences (low-Coding Gain sequences elicited few 5 responses).

Conclusion

The analyses above highlight the significant roles that measures of entropy and predictability have on musical learning and memory. The three information-theoretic measures examined here, *Sequence IC*, *Sequence Predictive Information*, and *Sequence Coding Gain*, all impacted on learning over time (as evinced by their significant interactions with *Listening Session*). In the first memory test, *Sequence IC* had little effect on the correctness of participants' responses. In the subsequent listening sessions, a trend was displayed between increasing *Sequence IC* and number of incorrect responses ($p < .01$). Similarly, *Sequence Coding Gain* did not have a significant effect on response in the first listening session, but was *positively* correlated ($p < .01$) with correct response in the second and third listening sessions. Sequences with high average Coding Gain were more likely to yield correct responses in the memory tests. In addition, *Sequence Predictive Information* did not impact on memory performance initially, but by the third listening session, this measure was negatively correlated with Correct Response such that greater Predictive Information led to fewer correct responses ($p < .05$). Again, *Predictive Information* is upper-bounded by Information Content (unexpectedness); therefore, high Predictive Information sequences sound relatively unpredictable. To summarize, these results suggest that the global statistical properties of the tone sequences had little bearing on recognition memory judgments initially, but over repeated listenings, sequences higher in information and entropy (those that sounded less predictable) produced both lower expectedness ratings and poorer recognition memory.

As displayed by the interaction between *Familiarity* and *Sequence Predictive Information*, *New* sequences that are high in Predictive Information tend to yield more correct responses (Correct Rejections) compared with *Old* sequences that are high in Predictive Information, which yield fewer correct responses (Misses). We suggest that sequences with high Predictive information are surprising but also distinctive, making them easier to correctly reject on *New* trials but harder to remember on *Old* trials. Listeners display poor recognition memory performance for individual sequences, and appear to respond based on the statistical properties of the sequence. Follow-up studies need to be conducted to explore these complex information dynamics, but it is clear that the information-theoretic measures investigated in this study interact dynamically with both expectedness and learning over a period of increasing exposure to novel tone sequences.

General Discussion

Information-theoretic approaches have elucidated various aspects of music perception, such as melodic expectation (e.g., Pearce et al., 2010). In the IT study described above, three subjective information-theoretic factors, *Sequence IC*, *Sequence Predictive Information*, and *Sequence Coding Gain*, all significantly influenced expectedness ratings of probe tones during the listening sessions. This reveals that the perceived expectedness of events is influenced not only by properties of the event itself, but also by properties of the sequence within which it is embedded. These factors also impacted on nuanced memory performance during the recognition tests. It was also interesting to discover a significant interaction between *Familiarity* and *Sequence Predictive Information* for memory performance. The increasing effect of IT measures on recognition accuracy may result from listeners gradually learning the underlying Markov model: Upon gleaning the basic information structure of the melodies, Predictive Information has a greater effect on recognition memory. Additionally, sequences that have high average IC can also vary in Predictive Information; that is, tones may be perceived as unexpected, but they can be surprising in either a way that increases listeners' predictions of forthcoming tones, or in a way that is surprising but does not increase predictive accuracy. The significant interaction between *Familiarity* and *Sequence Predictive Information*, but not *Familiarity* and *Sequence IC*, demonstrates that it is not simply the high-information content of sequences, but rather the *Predictive* Information of these sequences that listeners can successfully use when making memory judgments.

Generally, sequences that were more difficult to predict (higher IC/Predictive Information) gave rise to worse memory performance. There was also an increasing impact of these factors on memory as exposure increased. The effect of *Sequence IC* became more pronounced as listeners repeatedly heard melodies (e.g., sequences with low average IC were more likely to be remembered by the third listening session). To our knowledge, this research is the first investigation of the time course of music learning using an information-theoretic approach.

Although listeners struggled with the difficulty of the recognition memory task, they responded differentially based on the statistical properties of the sequences. Listeners may be more adept at learning the statistical rules underlying musical sequences than the specific exemplars themselves, especially with non-stylistic music such as the sequences used in this study (see Saffran et al., 1999; Loui, Wessel, & Kam, 2010; Halpern & Bartlett, 2010). Listeners are capable of learning a vast number of songs and themes, therefore more ecological stimuli may lead to better learning and memory performance. Language research (a domain in which listeners have been shown to be proficient in statistical learning of phonological sequences) has historically revealed that people tend to remember the *semantics* of what is said, not a verbatim account (e.g., Bartlett, 1932). Therefore, it may prove more insightful to test listeners' learning of semantics (musical structure and underlying statistics) across exemplars rather than the individual exemplars themselves.

We see from this IT study that learning individual sequences is possible, but challenging. Because we see effects of the IT properties of the stimuli but no significant effect of *Listening Session*, it is likely that participants were learning the *rules* describing the underlying transition matrices rather than the particular exemplars themselves.

Because it is impossible to perform an exhaustive behavioral investigation into which exemplars and rules listeners learn, future work will develop computer models to simulate and predict the process of musical learning. Computational models can offer insight into this process by analyzing information-theoretic measures to predict human listeners' performance. Future work will also test memory differences between ecological melodies and experimentally controlled tone sequences with an expectation that stylistic, ecological exemplars will aid memory performance.

Acknowledgements

This research was supported by EPSRC via grant number EP/D038855/1: *Neural and Information Dynamics in the Processing of Musical Structure.*

References

Abdallah, S. & Plumbley, M. (2009). Information dynamics: patterns of expectation and surprise in the perception of music. *Connection Science, 21,* 89-117.

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Halpern, A. *&* J. Bartlett *(*2010). Memory for melodies. In M.R. Jones, A. Popper, *&* R. Fay (Eds.), *Music Perception.* New York: Springer-Verlag.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford, UK: Oxford University Press.

Krumhansl, C. & Kessler, E. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334-368.

Loui, P., Wessel, D.L., & Kam, C. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27*, 377–388.

Manning, C. & Schutze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. Doctoral Dissertation, Department of Computing, City University, London, UK.

Pearce, M. & Wiggins, G. (2006). Expectation in melody: The influence of context and learning. *Music Perception, 23,* 377–405.

Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage, 50,* 302-313.

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.