Statistical learning of nursery rhymes

Holly Huey
Department of Psychology
UC San Diego
hhuey@ucsd.edu

Srihita Rudraraju
Department of Psychology
UC San Diego
srrudrar@ucsd.edu

Janna Wennberg
Department of Psychology
UC San Diego
jwennber@ucsd.edu

Abstract

In music from Western cultures, melodies often follow highly structured patterns and regularities in transitions between tones. By manipulating listeners' expectations and predictions about upcoming tones, composers can take advantage of people's ability to detect statistical regularities in music to generate rich melodies. What transitions in tones characterize musical melodies? Here we develop an ideal observer model using a 1-back Markov model that detects statistically regular bigrams (i.e., transitions between two tones) in simple songs and generates novel music sequences. We find that our model extracts bigrams that are consistent with certain principles of music theory, such as the pitch proximity principle and key profile. Taken together, our study provides insight on how musical Markov models can be used to generate melodies inspired by existing songs.

Keywords: Keywords: music cognition, statistical learning, implicit learning, markov model

Our ability to discern structure from complex perceptual inputs is a fundamental aspect of cognition. From detecting regularities in auditory streams of both natural and artificial language (Saffran & Aslin, 1996; Saffran, Newport, & Aslin, 1996; Pelucchi, Hay, & Saffran, 2009) to recurring structures in visual scenes (Fiser & Aslin, 2001, 2002b, 2002a; Kirkham, Slemmer, & Johnson, 2002), research across numerous cognitive domains has revealed early developing statistical learning mechanisms that allow our perceptual system to mitigate what would otherwise be an insurmountable computational problem.

This general-purpose ability to detect statistical regularities is not only a critical means for survival in complex perceptual environments, but also a means of enjoyment in discerning structure from complex inputs, such as music. Music composers take advantage of this ability to detect statistically regular patterns in order to manipulate listeners' expectations and predictions about upcoming events (Morgan, Fogel, Nair, & Patel, 2019). In Western cultures, the majority of musical arrangements are characterized by a tonal hierarchical organization and high degree of consonance (Mencke, Omigie, Wald-Fuhrmann, & Brattico, 2019). Music theorists (Meyer, 1956; Lerdahl, 1988) and psychologists (Krumhansl, 1990) have described tonality as an vital structural property of music, because it provides the hierarchical ordering of tones in the diatonic scale such that these tones are perceived in relation to the tonic (Fig. 1). Because the tonic is often considered to be the reference point of a key, some tones can be perceived as more "stable" than others, which instead might be perceived as "moving away" or "against the acting force" of the tonic (Zuckerkandl, 1971). Studies show that listeners have increased emotional responses to tonal arrangements in which they can more easily recognize musical structure, relative to atonal arrangements (Daynes, 2011).[1] Some music theorists further posit that the ability to form expectations about upcoming tones in a melody is critically linked to pleasure in music: listeners experience satisfaction when those expectations are confirmed and surprise when those expectations are violated at strategic moments (Huron, 2006; Jackendoff, 1992; Meyer, 1956). Although arrangements are often composed by expert musicians who explicitly aim to insert regular patterns in their melodies, studies demonstrate that listeners with minimal music training can detect such patterns in harmonic progressions, directional changes in pitch, and repetition patterns (Cuddy, Cohen, & Mewhort, 1981). Further, this ability to predict which tones belong to musical scales emerges early in development (Lynch, Eilers, Oller, & Urbano, 1990).

Here we investigate how listeners may extract these regularities in music using an ideal observer model. Our work builds on recent computational advances in musical Markov chain modeling used to detect patterns in time, pitch, and duration (Verbeurgt, Dinolfo, & Fayer, 2004), identify styles of composers (Liu & Selfridge-Field, 2002), explore statistical complexity of sequences (Volchenkov & Dawin, 2012), and compose new songs (Ramanto & Maulidevi, 2017; Verbeurgt et al., 2004). In contrast to prior work, however, our

[1] Our paper will only explore tonal arrangements but for research on atonal arrangements, see (Mencke et al., 2019)

Figure 1: Roman numerals refer to the scale degrees of tones within a scale. For example, in a C major scale, C is the first scale degree and tonic (I) of the scale. G is the dominant fifth tone (V) and is the second most important note. Together, these scale degrees are central components in the I-V-I progression that is commonly used to generate a sense of completion in a musical sequence and overarching melody.

approach is to consider the statistical regularities of bigrams (i.e., two-tone sequences) that may be predicted by two different principles in music theory:

1. *Pitch proximity principle.* This principle accounts for listeners' preference for ascending or descending transitions in tones to be made in small intervals (Von Hippel, 2000; Von Hippel & Huron, 2000). For example, the pitch proximity principle predicts that melodies will contain bigrams, such as ascending scale degrees like I-ii, ii-iii, iii-IV, etc. and descending scale degrees like I-vii, vii-vi, vi-V, etc. Thus, subsequent tones will be selected from a tight normal distribution centered around the previous tone.

2. *Key profile.* Simultaneously, the key profile of a melody determines the compatibility of tonal transitions, based on the key of the scale. Because the tonic determines which sequences of tones are considered consonant or dissonant, certain tones evoke a greater sense of stability than others (Brown, Butler, & Jones, 1994; Krumhansl, 1990) and, therefore, are more probable in sequence than others (Smith & Schmuckler, 2000). For example, bigrams of scale degrees I-iii and I-V are more stable because they define the major or minor triad of a diatonic scale. On the other hand, other leaps in intervals, like IV-vii (known as the "Devil's tritone"), are less stable. Unlike the principle of pitch proximity, the key profile of a melody is oriented around the tonic and subsequently, the dominant (the fifth scale degree).

Combined, these two principles of music theory generate testable predictions for which tones may follow another, given a particular musical sequence. Concretely, the pitch proximity principle predicts that any two-tone sequence is probable so long as the interval between the subsequent and preceding tones is small. By contrast, a key profile predicts that two-sequences may contain large interval transitions, so long as the tones resolve to the tonic or dominant of the key. Thus, sequences containing pairwise combinations of I, III, V are more probable based on the key profile of the scale. Temperley's (2008) Probabilistic Model of Melody Perception utilizes these principles to identify the key of folk songs, assign probabilities to different tones in a sequence, and detect incorrect tones in melodies (Temperley, 2008). However, as Morgan et al. (2019) note, the model does not infer a moment-to-moment tone transitions in determining which sequences of tones are more consonant than others and makes minimal use of statistical learning (Morgan et al., 2019).

We constructed a 1-back Markov model to learn statistically regular bigrams in melodies and compose novel musical sequences. While musical arrangements are comprised of rich interplays between melodies, rhythm, harmonic chord progressions, and dynamics, our model will only focus on transitions in tones in melodies. Our approach this advances prior understanding of statistical learning in music in three ways: first, we develop an ideal Bayesian listener using a 1-back Markov model to detect statistically regular bigrams in classic nursery rhymes; second, we train our model to generate novel musical sequences; and third, we test whether these generated songs contain regularities predicted by the pitch proximity principle and key profile. More specifically, if our model successfully captures properties of musical described by these principles of music theory, we predict that our generated songs will contain bigrams of small interval transitions and when large interval transitions, that they will be contain more consonant interval transitions (e.g., I-V), relative to randomly generated sequences of tones. As a test case, we will also examine the frequency with which the Devil's tritone, canonical "unstable" bigram, appears in our generated songs.

Dataset

We trained our 1-back Markov model on a dataset consisting of 29 classic, Western European nursery rhymes (e.g., "Hot Crossed Buns" and "Frere Jacques"). These songs were selected because nursery rhymes are simple yet highly structured, frequently consisting of an arc-shape melody in which melodies often begin and resolve with the tonic. They often only span one or two octaves and can be played with a single instrument or sung by a single person. Our code and dataset can be downloaded from `https://github.com/hollyhuey/music211`

Although the selected nursery rhymes were composed in C major, the tones were converted in scale degrees, so that our model would be generalizable to other nursery rhymes composed in different keys. For example, the opening phrase of "Frere Jacques" written in C major is C-D-E-C, but would be transposed into the numerical sequence 1-2-3-1 to denote the progression of scale degrees. Because nursery rhymes are highly repetitive in that they use the same melody with multiple different lyrics, we only included a single stanza of each melody. Each song in our dataset had a median length of 32 tones.

Model

We used a Markov chain of order (or "memory") 1, where, given a sequence of musical notes, the subsequent note depends on the most recent note. Thus, for a given note, the probability of the next note is given by the following conditional probability:

$$P_{i,j} = P(X_1 = j \mid X_0 = i) \tag{1}$$

Before training our Markov model on the nursery rhymes, we initialized the transition matrix M as an 8×8 matrix of ones (i.e., $M_{i,j} = 1$). This constitutes our weak, uninformative prior: we assumed that each bigram would occur once and that all bigrams were equally likely to occur.

For each nursery rhyme, we constructed our transition matrix M by counting the number of times each bigram occurs. We computed the raw tallies of each song at a time to avoid counting adjacencies between the ending note of one song and the starting note of the next. Finally, we normalized each row of our matrix by dividing each value by the number of times the starting note i occurred. Thus, for starting note i and subsequent note j,

$$M_{i,j} = P(X_1 = j \mid X_0 = i) \tag{2}$$

$$= \frac{P(X_1 = j \land X_0 = i)}{P(X_0 = i)} \tag{3}$$

This transition matrix is shown in Table 1.

Table 1: A table of the transition probabilities calculated from existing nursery rhymes. The rows denote the starting note, and the columns denote the subsequent note.

	C1	D	E	F	G	A	B	C2
C1	**.40**	.24	.05	.05	.15	.04	.04	.02
D	.27	**.22**	.35	.05	.07	.01	.02	.01
E	.1	.26	**.31**	.20	.09	.02	.01	.01
F	.02	.04	.32	**.30**	.24	.06	.01	.01
G	.12	.05	.13	.20	**.28**	.17	.01	.04
A	.06	.02	.01	.07	.39	**.26**	.15	.05
B	.17	.05	.05	.02	.17	.32	**.10**	.12
C2	.03	.03	.03	.03	.21	.27	.15	**.24**

After obtaining our transition matrix M, we used it to generate 100 "songs" (sequences of notes), each with 32 notes, which was the median length of the songs used to construct the transition matrix. For each song, we sampled the first note according to the relative frequencies of the notes overall. We sampled all remaining notes using our transition matrix: assuming Tone i was the most recent note in the sequence, Tone j was sampled according to the conditional probabilities of each note given Note i. An example of a "song" is shown in Fig. 2.

Results

After generating a new dataset of 100 songs, we tested whether our 1-back Markov model extracted statistical regular bigrams from the original dataset of nursery rhymes.

Pitch Proximity Principle

In order to test whether our model extracted bigrams consistent with the pitch proximity principle, we calculated the distance of each tone to the subsequent tone in each of the generated songs. Analyses revealed that the distribution of possible subsequent tones was Gaussian, centered on zero and ranging from -7 to +7, but was quite narrow (Fig. 3). This distribution reflects the fact that smaller interval transitions were more common than large interval transitions in the generated songs, suggesting that songs generated by a 1-back Markov model capture properties of the pitch proximity principle.

As a measure of how "musical" the generated songs were, we constructed a null set of songs by generating random sequences that were uniformly sampled from the eight tones of a diatonic scale (i.e., the first scale degree to its octave). We then

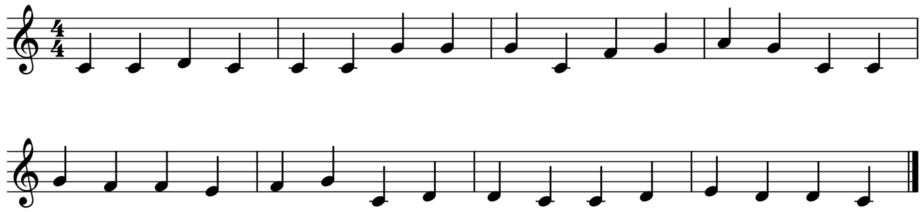

Figure 2: An example of a generated sequence

used an F test to compare the distance variation of the generated sequences to the distance variation of the null sequences. Relative to the distribution of "tones" in the null sequences, there was significantly less variance in the distribution of tones in the generated songs ($F_{3099,3099} = 0.350, p < 2.2 \times 10^{-16}$).

Key Profile

To test whether consonant bigrams were more common in generated songs, we compared the proportion of bigrams containing the tonic, dominant, and third scale degree, which comprise the major or minor triad of a scale. We found that bigrams containing interval transitions with these tones were significantly more frequent in generated songs than in null songs ($\chi^2 = 8.675, p = 0.0016$). These results suggest that generated songs contain more bigrams that fit the key profile of the melody than not. We also constructed a stacked bar chart visualizing the transition matrix of our generated rhymes (see Fig. 4).

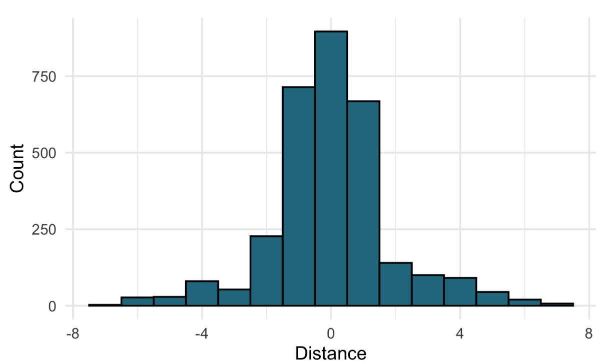

Figure 3: A histogram of the interval distances for all generated songs.

Also of interest was the Devil's tritone (IV-vii or vii-IV). Based on the key profile of the scale, we predicted that there would be disproportionately fewer occurrences of the Devil's tritone in our generated sequences. Analyses confirmed this prediction by revealing that there were significantly fewer Devil's tritone occurrences in the generated songs than in the null songs ($\chi^2 = 7.70, p = 0.003$).

Discussion

In this paper, we investigated how statistical learning models can be developed to detect musical structure and regularities. We generated a 1-back Markov model to extract statistically regular bigrams from classic Western European nursery rhymes in order to compose novel songs of 32 tones. Consistent with principles of pitch proximity and key profile accounted for in music theory, our model successfully generated songs that contained a higher proportion of bigrams consisting of small interval transitions compared to large interval transitions; of large transition bigrams, there was higher proportion of more consonant than dissonant bigrams. In particular, the Devil's tritone was disproportionately absent from generated songs.

Although our Markov model was able to extract statistically regular patterns from melodies of nursery rhymes, future research will be needed to test how well this performance matches human performance. Additionally, a key question considered by this paper was how enjoyment of music is an implicit measure of musical structure. Thus, a critical next step for this study is to test how participants with a range of musical training rate songs generated by our model as more or less consonant in comparison to randomly generated sequences of tones. This would then measure how well participants could intuit the musical structure that our model aimed to capture in the generated bigrams. Given that the perceived stability of musical sequences depends on its tonal context, it is important for listeners to be able to discern the tonic of the melody. While the bigrams of our model may be more probable than others, they are meaningless without relation to a tonic. Thus, in future work, it may be important to filter 1-back Markov sequences that contain at least one tone that is the tonic.

Furthermore, while our model captured more consonant bigrams of interval transitions than dis-

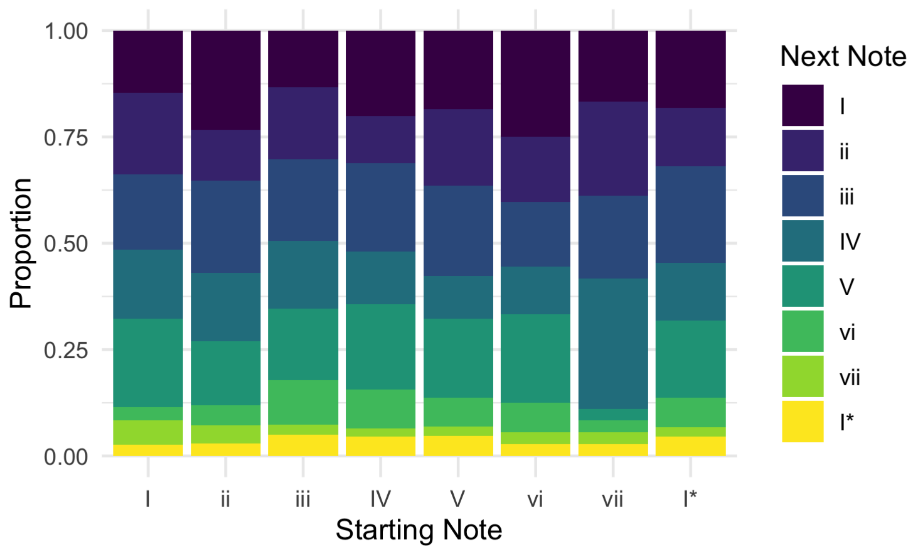

Figure 4: A stacked bar chart visualizing the transition probabilities of our generated sequences. Each bar represents the starting note, and the fill of each bar represents the next note in the sequence.

sonant bigrams, interval transitions are characterized as being more stable by comparison to less stable ones. It may be possible that songs entirely composed of stable bigrams are not perceived as structured because there is no instability or dissonance to resolve back to the tonic. Tones that are considered stable occur more frequently and for the longer durations, while unstable tones occur infrequently and for shorter durations (Smith & Schmuckler, 2000). Multiple studies show that listeners' are sensitive to this distribution of stable and unstable tones (Laden, 1994; Oram, Cuddy, & Oram, 1995). However, such sensitivity suggests that there may be a preferred proportion of stable vs. unstable interval transitions that make for more robust musical structures.

Most importantly, however, while our paper demonstrates that a 1-back Markov model may successfully capture two-tone sequences, music can hardly be considered to consist of a series of transitions between two tones. Even the simplest melodies are composed of more complex sequences such as phrases, a unit of musical meter that has a complete musical sense of its own that can be combined with others to form melodies, or periods, a larger musical unit comprised of an antecedent and consequent phrase. These musical units, which are more complex and consist of more tones than bigrams, cannot be captured by a 1-back Markov model. Thus, it will be important for future work to disentangle at which structural level an n-back Markov model would best extract statistical regularities to generate melodies that sound like music.

Musical arrangements in Western cultures follow highly structured patterns and regularities in transitions between tones. Our model presents a simple 1-back Markov model for extracting regular bigrams of two-tone sequences. While this model extracts the lowest musical unit of a melody, such work contributes to the growing literature on music composition through computational modeling. Thus, over time, advancing our knowledge of statistical regularities in music may not only lead to a deeper understanding of how people detect structure from inputs as complex as music but also the design of music composition tools.

Acknowledgments

We are grateful to Anna Mai and to Ed Vul for helpful discussions.

References

Brown, H., Butler, D., & Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Perception*, *11*(4), 371–407.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 869–883. (Place: US Publisher: American Psychological Association) doi: 10.1037/0096-1523.7.4.869

Daynes, H. (2011, October). Listeners' perceptual and emotional responses to tonal and atonal music. *Psychology of Music*, *39*(4), 468–502. Retrieved 2020-12-16, from https://doi.org/10.1177/0305735610378182 (Publisher: SAGE Publications Ltd) doi: 10.1177/0305735610378182

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*(6), 499-504. (PMID: 11760138) doi: 10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Huron, D. B. (2006). *Sweet anticipation: music and the psychology of expectation*. Cambridge, Mass: MIT Press. (OCLC: ocm61684238)

Jackendoff, R. (1992). Musical processing and musical affect. In *Cognitive bases of musical communication* (pp. 51–68). Washington, DC, US: American Psychological Association. doi: 10.1037/10104-004

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35 – B42. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010027702000045 doi: https://doi.org/10.1016/S0010-0277(02)00004-5

Krumhansl, C. L. (1990, April). Tonal Hierarchies and Rare Intervals in Music Cognition. *Music Perception*, *7*(3), 309–324. Retrieved 2020-12-16, from /mp/article/7/3/309/62922/Tonal-Hierarchies-and-Rare-Intervals-in-Music (Publisher: University of California Press) doi: 10.2307/40285467

Laden, B. (1994). Melodic anchoring and tone duration. *Music Perception*, *12*(2), 199–212.

Lerdahl, F. (1988). Tonal pitch space. *Music perception*, *5*(3), 315–349.

Liu, Y.-W., & Selfridge-Field, E. (2002). *Modeling music as markov chains: Composer identification.* Citeseer.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*(4), 272–276.

Mencke, I., Omigie, D., Wald-Fuhrmann, M., & Brattico, E. (2019). Atonal Music: Can Uncertainty Lead to Pleasure? *Frontiers in Neuroscience*, *12*. Retrieved 2020-12-16, from https://www.frontiersin.org/articles/10.3389/fnins.2018.00979/full (Publisher: Frontiers) doi: 10.3389/fnins.2018.00979

Meyer, L. B. (1956). *Emotion and meaning in music. chicago: Univ.* Chicago Press.

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23–34.

Oram, N., Cuddy, L. L., & Oram, N. (1995). Responsiveness of western adults to pitch-distributional information in melodic sequences. *Psychological Research*, *57*(2), 103–118.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child development*, *80*(3). Retrieved 2020-12-16, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883431/ doi: 10.1111/j.1467-8624.2009.01290.x

Ramanto, A. S., & Maulidevi, N. U. (2017). Markov chain based procedural music generator with user chosen mood compatibility. *International Journal of Asia Digital Art and Design Association*, *21*(1), 19–24.

Saffran, J. R., & Aslin, R. N. (1996). Statistical Learning by 8-Month-Old Infants. *Science, New Series*, *274*(5294), 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996, August). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. Retrieved 2020-12-16, from `https://linkinghub.elsevier.com/retrieve/pii/S0749596X96900327` doi: 10.1006/jmla.1996.0032

Smith, N. A., & Schmuckler, M. A. (2000). Pitch-distributional effects on the perception of tonality. In *Sixth international conference on music perception and cognition, keele* (pp. 5–10).

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, *32*(2), 418–444.

Verbeurgt, K., Dinolfo, M., & Fayer, M. (2004). Extracting patterns in music for composition via markov chains. In *International conference on industrial, engineering and other applications of applied intelligent systems* (pp. 1123–1132).

Volchenkov, D., & Dawin, J. R. (2012). Musical markov chains. In *International journal of modern physics: Conference series* (Vol. 16, pp. 116–135).

Von Hippel, P. (2000). Redefining pitch proximity: Tessitura and mobility as constraints on melodic intervals. *Music Perception*, *17*(3), 315–327.

Von Hippel, P., & Huron, D. (2000). Why do skips precede reversals? the effect of tessitura on melodic structure. *Music Perception*, *18*(1), 59–85.

Zuckerkandl, V. (1971). *The Sense of Music*. Princeton University Press. (Google-Books-ID: NGPwEeRUXs8C)