The Evolution of Music Cadence from Medieval to the Renaissance: An Iterated Learner Model

Leslie Li (rlli@umd.edu)
NACS, 4090 Union Drive,
College Park, MD 20742 USA

Abstract

Iterated learner models have been applied to various cultural changes from semantic space (Xu, Dowman, & Griffiths, 2013) to phonology (Ito & Feldman, 2021). In this project, I adapt an iterated learner model to simulate historical change in music. I found that certain assumptions such as preferring proximal movements and preferring the modern musical scales to be better fits of the specific changes in musical cadence from the 14th to the 18th century, with a few exceptions that may be improved in the future by introducing harmonic structure to the model. This is the first work to simulate cultural change of a specific structure in music, which differs from the simulations of language change in the lack of direct communication through music.

Keywords: iterative learning; music perception; cultural evolution;

Introduction

Language, among other human activities, changes along time. Historical linguistics often focuses on the pattern and elements of such changes. Also, models have been proposed to simulate language change as well as connecting language change to the inductive biases of humans.

The specific model of interest, first proposed by Griffiths and Kalish (2007), sees language as a product of cultural transmission. It sees generations of language speakers as rational learners, who infer hypotheses about the language from available data. Critically, an information bottleneck exists between generations of learners, where part of the data from the last generation are not passed onto the next generation, and the next generation of learners must infer the missing data based on their hypotheses.

A number of works have since applied the model in Griffiths and Kalish (2007) on variants of linguistic phenomena. Griffiths, Kalish, and Lewandowsky (2008) applied this framework on learning abstract functions, while groups of people cannot see the ground truth or abstract description of a function, but can only infer from the examples given by another group. Xu et al. (2013) applied this framework on color term learning, where people can only infer the reference of artificial color terms from examples given by another group. Kirby, Griffiths, and Smith (2014) applied this frame on artificial language learning. These works all find that the transmission of these concepts, from abstract function to color terms, is influenced by inductive biases. They conclude that language change can be simulated as a cultural evolution phenomenon, where certain structures in language are chosen because they are likely to be learned.

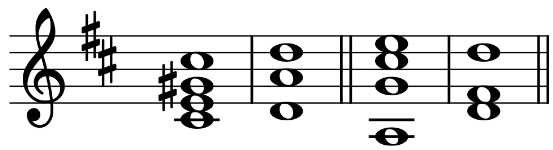

Figure 1: Two cadences of the medieval time (mm. 1–2) and during the Classical period (mm. 3–4). The Medieval cadence is adapted from Machaut's *La Messe de Nostre Dame* (composed in the 14th century), and the Classical cadence is adapted from Mozart's *Ave Verum Corpus* (composed in the 18th century).

Cadences in music

Music, on the other hand, is a domain with many shared characteristics with language. Both are unique yet universal in humans, and involve complex structure in the acoustic signal. What remains different between these two domains, however, is that music is often purely artistic and is not meant to convey semantic information. Therefore, theories about language learning and change do not necessarily generalize to the domain of music. In this project, I will examine music change in light of cultural evolution and iterated learning.

Specifically, I will focus on the musical cadence. Cadence comes from the Latin word for "to fall," and marks the sense of falling, ending, or closure of a passage (Mutch, 2015). As the description suggests, cadences often occur at the end of a musical passage. Across musical culture and history, cadences can be realized through both temporal change (such as slowing down) and harmonic change (such as a particular harmonic structure). The current project will focus solely on the harmonic aspect of cadence.

Historically in the Western (European) music tradition, the realization of cadencec underwent a series of changes. While Mutch (2015) gave a thorough review and theoretical discussion, I will give a brief summary in this paper.

For most of medieval music (up to the 14th century), the harmonic structure is quite different from what it developed into during the Classical period (ca. 18th century), which is what modern musicians are familiar with. First of all, medieval music used different musical scales than modern music — church modes. There were four main modes with some variants, from which two developed into the Major and Minor scales used today. As a result, most of medieval music would sound jarring to modern listeners due to the different scales used.

Additionally, medieval music used harmony differently from later Western music. Triads (three-note chords) were not used until late Medieval music (ca. 15th century) and was not systematically discussed by music theorists until the 18th century (Sadler & Christensen, 2001). Therefore, the notion of chord progression, or any chord at all, were not yet established in Medieval music of the 13th and early 14th century.

These factors, along with specific trends to a particular time or practice, led to great differences between cadences used by Medieval composers and cadences in later practice. A reduced example of this difference is shown in Figure 1. In the Classical cadence (mm. 3–4), one can observe the V–I harmonic motion, resolution of the leading tone (C in m. 3), downward resolution of the 7th (G in m. 3), and the independence of individual voices, which are terms one would learn in an introductory music theory class about how to write cadences. However, very few of these features are present in the Medieval cadence. This example in particular, called a "double leading tone cadence" (Rockstro, Dyson, Drabkin, Powers, & Rushton, 2001), creates two leading tones not only approaching scale degree 1 (D in m. 2), but also to approach 5 (A in m. 2). This gesture is a signature to cadences in the Medieval time, but would sound jarring to a modern listener.

The Current Study

Various accounts and theories have been proposed to explain or characterize the change of cadential structure, of which the various jargons mentioned in the last section are examples of how modern scholars characterize cadences of different time. This project, however, takes a different approach to characterize the historical change of cadential structure through cultural evolution. I will try to apply an iterated learner model on music change, and explore different hypotheses about the learners' inductive biases.

Method

Iterated Learners

Following the iterated learning models (Griffiths & Kalish, 2007; Ito & Feldman, 2021) closely, I simulated chains of iterated learners of musical cadence. A pool of a set number of cadences will be available as the training data, and at each generation, a small portion (0.05%) of data will be forgotten. [1] For each "forgotten" datapoint, the missing note is sampled based on its probability given the learned hypotheses $p(d|h)$. The learned hypotheses, in turn, are generated based on the available data.

Therefore, let the data at initialization be d_0. The first generation will obtain hypotheses h_0 based on the not forgotten part of the data (see the next section for details of this step). Then, the first generation of learners must infer the missing notes based on h_0. As they sample notes to fill in the missing notes, d_1 is generated.

Importantly, for any generation i, because 95% of d_i is identical with d_{i-1}, the data between generations (d_i and d_{i-1}) are dependent. However, since for each generation the hypotheses h_i is only sampled from d_i, h_i is conditionally independent with h_{i-1}. As a result, each set of hypotheses are only conditionally dependent on the immediately previous set of hypotheses, thereby making the chain of learners a Markov chain.

The Hypotheses of Learners

Two specific hypotheses are designed to represent learners' learning mechanism and cognitive bias.

A few other assumptions are imposed on both models. First, all notes that are not part of a modern Western scale will be "warped" or transposed to a note on the Western scale. The simulations here focused on major keys only, so any note that is chromatic to the major key will have a equal probability to move upward or downward by step. This assumption is hard-coded into every iteration of the model as a constraint, since we know in retrospect that such a change happened through music history. Also, smoothing is introduced into the model. One datapoint is created for each pair of possible notes (except for a pair of identical notes, to represent the assumption that *some* movement must happen at a cadence). This is to prevent the model from assigning zero probability to a combination it has never seen, thereby increasing the flexibility of the model.

Conditional Probability (CP) Model The first model learns only based on conditional probability. Let A be the penultimate note of a cadence that is missing and B be the final note that A cadences into. This model will calculate $p(A)$ only in terms of its probability conditioned on the final note. That is:

$$p(A) \propto p(A|B) \tag{1}$$

CP + Gestalt Model The second model learns not only based on conditional probability, but also with a continuity constraint:

$$p(A) \propto p(A|B)D(A,B) \tag{2}$$

where $D(A,B)$ stands for the distance between A and B. Note that although $D(A,B)$ is maximal when A and B are the same note, $p(A|B)$ would be zero since there is no cadence without changing a note in the corpus, even after the smoothing process. Therefore, the model is still prevented from learning cadences without changing any notes.

[1] Note that only the penultimate note will be forgotten, and the final note will never be forgotten. This is done because the final harmony changed relatively little through the centuries, and the change in the penultimate harmony is more of interest. On the other hand, if either note can be forgotten, the training outcome will become something no longer interpretable as cadence.

Data Collection

For the medieval cadences, 37 cadences are annotated from an anthology of 14-th century madrigals (Apel et al., 1950).

For each cadence, the movement of each voice is annotated, including both the final note and the penultimate note. The duration of the note is omitted and only the pitch of each note is kept. Then, the notes are transposed with respect to the tonic of the cadence (e.g., with respect to a cadence in G, E would be 6 in scale degree). The selection of notes are adjusted when the penultimate note is the same as the last note (anticipation tone) or when the penultimate note is part of an ornamentation of some other note, in which cases the appropriate notes are selected instead. The exceptions take less than 5% of the data. Since each cadence involves 3 voices, there are a total of 111 pairs of vocal movements in the original data.

For testing, chords are collected from Bach chorale harmonizations and coded in the same way. Three chords are selected from the endings of the first four chorales (with one duplicate removed).

Experiment Design

50 random initialization of each of the two models are run by specifying different random seeds. 100 generations of iterated learning are simulated for each run. As mentioned before, 5% of the data is sampled at each generation to have the penultimate note forgotten, and the data counts are smoothed by adding 1 to every possible count. These parameters above are set by intuition and have not been tested through a rigorous grid search.

After each individual model is trained, three "late" cadences are presented to it. For each of the four voices of a cadence, the model will evaluate $p(A)$ from the final dataset d_{100}. To keep the two models comparable, the continuity constraint, which is specific to the second model, is not used. $p(A)$ is only calculated based on $p(A|B)$. The log probability of $p(A)$ is summed up across all four voices of a cadence to give a score of the cadence's likelihood under the trained model. The higher this score is, it means the model is less surprised by the modern cadence, and therefore is a better representation of the evolution of cadences.

Results

First, I tested the trained models on three cadences from Bach chorales. The scores are plotted in Figure 2. Since most bars are above the reference line denoting the original data, we can conclude that the models do become more "modern" after training. The assumption of key is most likely to lead to this outcome, since it leads to higher probability being assigned to notes that are within the key. Additionally, the iterated learning process may be successful in removing some exceptional cases in the original dataset.

For the second and third chords in Figure 2, the CP + Continuity model performs better than the CP model, but only marginally. In the first chord, the CP + Continuity model performs significantly worse even compared with the untrained reference. This is likely due to a $\hat{5}$–$\hat{1}$ movement in the bass and a leading tone not resolving to the tonic in Chord 1. Although the $\hat{5}$–$\hat{1}$ movement is very common in later cadences including Bach, it will be evaluated as very unlikely under the continuity constraint, since the movement involves a leap of a perfect fourth.

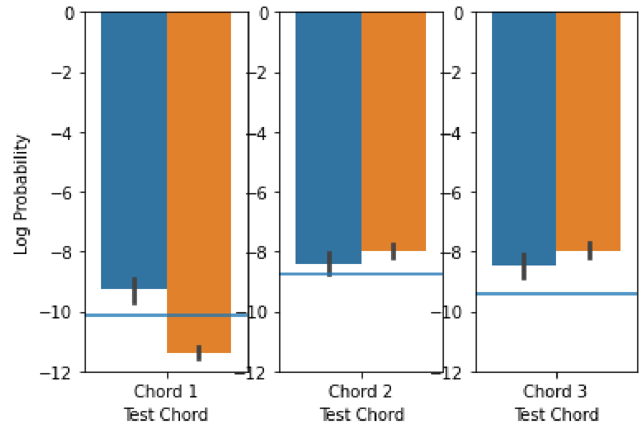

Figure 2: Training outcome of two different models. Log probability of each of the three test cadences (Chords 1–3) under the model is calculated for each model. The CP model is represented by blue bars, and the CP + Continuity model is represented by orange bars. Error bars denote the variability of the 50 initializations of each model. The blue horizontal lines denote the log probability of cadences under the original data, d_0.

To test the probability of the particular voice movements under the model, I tested three common voice movements in Figure 3. This allows observation into the models' response to very specific outcomes compared with the test in Figure 2. From Figure 3, we can see that Model 2 assigns much higher probability to $\hat{7}$–$\hat{1}$, which are only a semitone apart. The Gaussian continuity constraint will assign the highest probability to such a movement. For a $\hat{2}$–$\hat{1}$ movement, whose two notes are two semitones apart, Model 2 gives a lower probability than Model 1. For the $\hat{5}$–$\hat{1}$ movement, both models give low probabilities. This can be explained by the implied harmony of V–I motion that did not exist in the Medieval times. Furthermore, since the models only consider individual voices, there is no sense of harmony that can be possibly captured by the model. This leads to both models giving low probabilities to such a movement, while Model 2 is much worse, because the leap of a fourth involves 5 semitones and leads to low probability under the Gaussian distribution.

Discussion

This project applied iterated learner models to simulate change in music cadence. As the main findings are summarized in the last section, the interpretation and significance of this work can be seen with connections to different disciplines. In this section, I will discuss the connection between

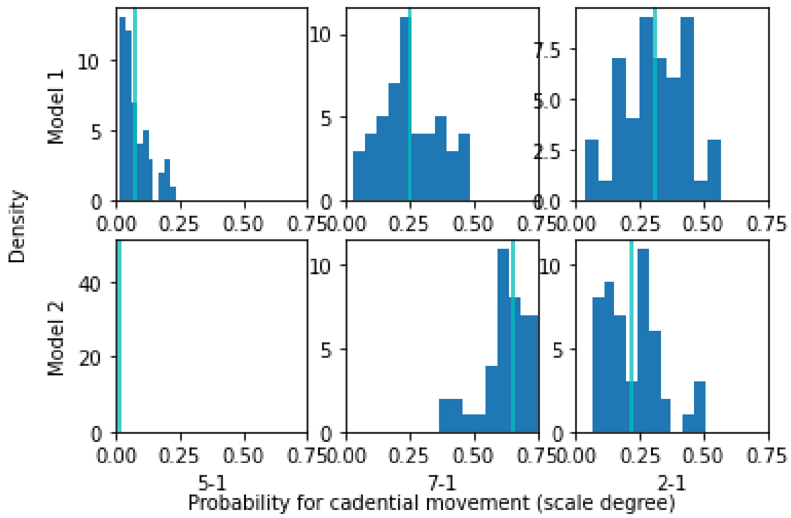

Figure 3: Probability of three vocal movements (one in each column) that end on the tonic (scale degree $\hat{1}$). The histogram denotes the density of assigned probability by each model (one model in each row). The distribution is created by testing the 50 initializations of each model. The vertical teal line denotes the average of the 50 initializations. For Model 2's probability of the $\hat{5}$–$\hat{1}$ movement, the numbers are all very close to 0 that it did not show under the normal scale. The three tested movements are chosen because they are the most common movements that ends on $\hat{1}$ taught in introductory music theory courses.

the current work and other fields.

Connection to prediction in music perception

In music cognition, the listener's prediction and surprisal to the upcoming notes has been of interest to researchers. However, formal models of this process has been limited on *n*-gram models and simple continuity models, which are compared in Morgan, Fogel, Nair, and Patel (2019). In this line of research, cadences are also particularly of interest because of the strong limitation of possible upcoming notes (and hence the strong expectation to certain notes). I would argue that the current work complements Morgan et al. (2019) by using a model that incorporates both an *n*-gram-like component (conditional probability) and a Gestalt-like component (continuity constraint). From the results, we see that addition of Continuity constraints to the CP model does not change the model's behavior in common cases, but impact the model's prediction in special cases such as the $\hat{5}$–$\hat{1}$ movement. This further suggests that implementing a harmonic component in the model may help it capture the variability in real music data that the current models fail to learn.

The current work, however, focuses on simulating music change rather than modeling the prediction and surprisal of listeners. From this perspective, the models in Morgan et al. (2019) can be seen as one generation among the chain of learners in the current study. It would be interesting to combine iterated learner models with human data such as cloze prediction and surprisal in addition to using transitional probability obtained for corpora.

Moreover, since iterated learner models can reconcile the practice of a specific culture (i.e., initialization of the chain) and inductive biases that are universal to humans, it can be used to explore what exactly the inductive biases are. For example, it had been debated whether certain consonant intervals in Western music (such as perfect fifths and octaves) are perceived as consonant/pleasant universally, or at least are easier to perceive or produce. Current works seem to suggest that the former is false (Jacoby et al., 2019) but the latter is true to some extent (McPherson et al., 2020). As Griffiths and Kalish (2007) suggested that a Markov chain of learners will eventually converge to represent the prior, and not the initialization, of the language, the same could be said for a chain of musicians. Therefore, questions about the inductive biases of music can be explored using iterated learner models.

Connection to cultural evolution

The study of how music changes along time has rarely been connected to the field of cognitive science. Compared with historical linguistics, which has attracted attention of cognitive scientists and stimulated theories of language universals and language change, music history has been mostly a focus in musicology only. However, change in music has some unique properties that make it interesting. For example, music notation are often much more transparent than language orthography. This allows us to be much more confident about how music sounded in the distant past (e.g., 13th century), and allows the modeling of music "sound change" with relative ease. Furthermore, as a modality that lies in the artistic domain and does not serve immediate communicative purpose, change in music is mostly motivated by implicit pattern-learning, and even intentional violation of expectations of the listeners (cf. the deceptive cadence). This is different from language where direct communication requires maximal amount of information to be conveyed with minimal effort. As a result, simulating music change with iterated learners suggests that a simple interaction between implicit learning and inductive biases could facilitate cultural change.

The iterated learner model could also be interpreted in an unique way in the musical context. Each generation could be seen as one individual musician, and the data available to the musician could be seen as all the compositions that the musician has access to. Due to information bottleneck, however, the musician will only be able to access some but not all the music compositions. The "forgotten" datapoints can therefore be interpreted as pieces that become out of fashion, and the musician will compose new songs based on their knowledge (i.e., hypotheses) from studying the current music still in fashion. This interpretation allows a fast lifecycle of innovation — in modern times, a new musician could become known for their works every few weeks or at most months; in the medieval ages, the life of a musical piece will likely be longer, but a new musician could still emerge ever few years. On the other hand, for iterated language learning, a new generation is often interpreted to be the next generation of people, which is about 20–40 years apart.

Connection to studies in music history

Studies of music history often do not involve computational tools or cognitive considerations. The current project seeks to bridge this gap. As early music studies often seek to obtain accurate description of past musical traditions with very limited data, simulations of longitudinal change in music have the potential to aid in this pursuit. While the current project is a very tentative step forward, it could be extended to answer questions about the historical music practice.

Additionally, the notion of harmony was not important in Medieval music. The models in this project, similar to Medieval music, consider individual voices instead of any harmonic structure formed by multiple voices. One possible expansion of the model is to consider the harmony formed by several voices. A model that incorporates harmony will be more authentic to the increased importance of harmony since the Renaissance, and may capture the $\hat{5}$–$\hat{1}$ movement better than the current models.

Conclusion

In this project, I explored the use of an iterated learner model to simulate the evolution of musical cadence. I found that it is feasible to simulate music cadence change through the model, which is the first attempt to apply iterated learning on changes of non-communicative systems such as music. Additionally, I found that the assumption of major/minor keys allows the models to capture the change of cadence in Western music to some degree, but leaves out a lot of variability unexplained still. Particularly, a potential change to the model to capture harmonic relation in addition to movement of voices along time may lead to better models.

References

Apel, W. .-., Holmes 1900-1972,, U. T. J., Linker, R. W., century. Andrieu, F. a. t., century-15th century. Anthonello de Caserta, a. t., Century., B. a. t., … century. Vaillant, J. a. t. (1950). *French secular music of the late fourteenth century*. Cambridge, Mass. SE - 1 score (xii, 39, 133 pages), 8 pages of plates : facsimiles (some color) ; 33 cm.: Mediaeval Academy of America.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, *31*(3), 441–480. doi: 10.1080/15326900701326576

Griffiths, T. L., Kalish, M. L., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3503–3514. doi: 10.1098/rstb.2008.0146

Ito, C., & Feldman, N. H. (2021). *Iterated learning models of language change: A case study of Sino-Korean accent.*

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, *29*(19), 3229–3243.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114. Retrieved from `http://dx.doi.org/10.1016/j.conb.2014.07.014` doi: 10.1016/j.conb.2014.07.014

McPherson, M. J., Dolan, S. E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E. A., … McDermott, J. H. (2020). Perceptual fusion of musical notes by native amazonians suggests universal representations of musical intervals. *Nature communications*, *11*(1), 1–14.

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and Gestalt-like principles predict melodic expectations. *Cognition*, *189*(December 2018), 23–34. Retrieved from `https://doi.org/10.1016/j.cognition.2018.12.015` doi: 10.1016/j.cognition.2018.12.015

Mutch, C. M. (2015). *Studies in the history of the cadence*. Unpublished doctoral dissertation, Columbia University.

Rockstro, W. S., Dyson, G., Drabkin, W., Powers, H. S., & Rushton, J. (2001). *Cadence.* Oxford University Press. doi: 10.1093/gmo/9781561592630.article.04523

Sadler, G., & Christensen, T. (2001). *Rameau, jean-philippe.* Oxford University Press. doi: 10.1093/gmo/9781561592630.article.22832

Xu, J., Dowman, M., & Griffiths, T. L. (2013). Cultural transmission results in convergence towards colour term universals. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1758). doi: 10.1098/rspb.2012.3073