Experimental evolution of music styles using automatic composition models

Eita Nakamura[1], Hitomi Kaneko[2], Takayuki Itoh[3], Kunihiko Kaneko[4]

[1]Kyoto University [2]Tokyo University of the Arts
[3]Ochanomizu University [4]Niels Bohr Institute

eita.nakamura@i.kyoto-u.ac.jp

Abstract

Musical culture is developed through descent-with-modification processes of complex knowledge about music creation, but how new creation styles evolve is not well understood. To study this problem, we conducted an artificial evolution experiment involving a population of machine-learning-based automatic composition models. These models evolved under evaluation by a large number of human listeners. We found that adaptive evolution of music styles can occur when the blending inheritance of high-dimensional statistics representing composition styles is incorporated into the generation update process. We also found a significant difference in musical preferences among listeners, depending on their musical experience. The results indicate the relevance of multi-parental transmission in the cultural evolution of music styles, a nontrivial fitness landscape of creation styles, and listeners' diverse preferences.

Introduction

Musical cultures develop through iterated processes of inheritance and modification of complex knowledge for creating music. Computational representation of such knowledge has long been studied in the field of automatic music composition (Nierhaus, 2009; Ebcioğlu, 1990; Pachet and Roy, 2011; Biles, 1994). Recent studies based on machine learning have shown that musical knowledge represented by high-dimensional statistics extracted from music data is effective for generating songs that imitate the features of existing music (Hadjeres et al., 2017; Sturm et al., 2016; Mittal et al., 2021; Agostinelli et al., 2023). However, how new and preferred music creation styles evolve is not well understood. Given the high dimensionality of genotype (knowledge) and phenotype (songs), and the stochastic relation between them, the conditions for adaptive evolution of music styles cannot be easily revealed by data analysis (Mauch et al., 2015; Singh and Nakamura, 2022) or by simple psychological experiments (Salganik et al., 2006).

We conduct an experiment to study the process of music style evolution in a constructive way. The framework is similar to that of the DarwinTunes experiment (MacCallum et al., 2012), which was based on evolutionary computation

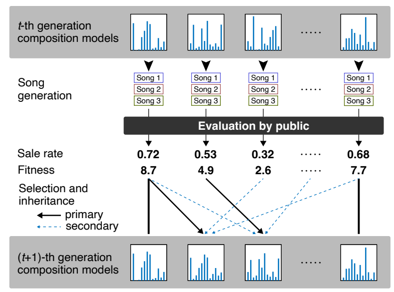

Figure 1: Overview of the experimental evolution.

involving human listeners. Instead of using sound generation algorithms as in that experiment, we employ machine-learning-based automatic composition models. These models can generate an indefinite number of songs, simulating creators who inherit a creation style by learning the statistics of songs from their cultural parents, rather than inheriting a song directly. We study the effects of blending inheritance, which mimics the fusion of different creation styles during transmission, and selection based on listeners' evaluations of the generated songs. In the following, we describe the overview of the method and the main results; the study will be presented in full detail in a forthcoming paper.

Method

In the experiment, the composition models generate melodies for given sets of chord progressions and lyrics (Fig. 1). We use as a composition model a product-of-experts model integrating a Markov model describing pitch transitions and rhythms (Nakamura et al., 2019), and a harmony action model describing the relationship between melody pitches and chords. This model has approx. 10^5 probabilistic parameters, which are treated as traits. Depending on the parameterization, a composition model can generate melodies in various existing styles, e.g. jazz and rock styles, as well as styles different from existing music.

50 composition models are present in each generation and the population is updated according to the models' fitness w_n. The fitness is measured by the human listeners' evaluations of the songs generated by the models for 20 chord pro-

gressions with lyrics. A pair of melodies generated by different composition models for the same chord progression is presented to a listener at a time, and the listener selects the one he/she prefers. By repeating these random and blind comparisons many times, we measure the sale rate r_n defined as the rate that composition model n is chosen as a preferred one. The fitness is then defined as $w_n = \exp(3\,r_n)$. About 70 undergraduate and graduate students, including students in music colleges, participated in the experiment.

After evaluation, 25 highest fitness models are included in the next generation, and the rest are removed from the population. For each of the newly introduced 25 models, a primary parent and two secondary parents are chosen from the current population with a probability $P(n) \propto w_n$. The traits are transmitted from these parents by blending inheritance: the offspring's traits θ is given as $\theta = (1-u)\theta_0 + u(\theta_1 + \theta_2)/2$, where θ_0 and θ_k $(k = 1, 2)$ are the traits of the primary and secondary parents, respectively, and $u \in (0, 1/2)$ is the oblique transmission coefficient, which is also treated as a trait transmitted from the primary parent with a random mutation. We note that this evolutionary process would end up converging into the average model obtained by averaging the traits of initial models, if there is no selection pressure.

To initialize the population, we trained 25 composition models (called progenitors) using a dataset of approx. 10^4 popular music songs, by statistical learning based on clustering and the maximum likelihood estimation. Each of the 50 models in the first generation are then constructed by randomly choosing three progenitors and averaging their traits.

Result

Generation updates were performed every few days, and by the 11th generation, most clans were removed from the population and six clans survived (the predominant clans in the tenth generation are labeled as C1 to C7). The average rating increased over generations, and a musical feature analysis showed considerable changes in the creation style of some clans (e.g. C2). An additional experiment was conducted to directly compare the 25 progenitor models, the average model, and representative models of clans C1 to C7 in the first and tenth generations. In this experiment, the models were put on a public automatic composition system (`https://creevo-music.com`) and a large number of general users evaluated them by random and blind comparisons. As a result, we observed significant increases in the sale rates in clans C2 and C7 over generations, indicating the adaptive evolution within these clans (Fig. 2A). The tenth-generation model of clan C2 was the most popular and had a significantly higher sale rate than that of the average model, indicating a nontrivial fitness landscape in the high-dimensional space of music styles.

Listeners' preferences were analyzed by their buy rates, the relative rates of choosing songs created by the predominant seven clans. A factor analysis on the buy rates revealed

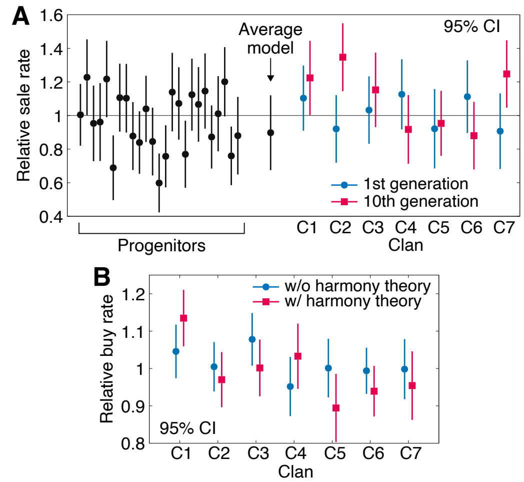

Figure 2: A: Comparison of composition models. B: Dependence of listeners' preferences on learning experience.

a difference in the preferences between listeners with and without formal learning experience of harmony theory. A result using the linear discriminant analysis showed a significant difference ($p = 1 \times 10^{-4}$) in the buy rates of these two groups (Fig. 2B).

Discussion

The evolution experiment showed that adaptive evolution occurs through blending inheritance in the high-dimensional space of music creation style, suggesting the relevance of transmission from multiple parents in music style evolution. The present evolutionary process, where the genotype is a machine learning model and the phenotype is a collection of songs, is different from that for song evolution by genetic algorithms (Moroni et al., 2000; MacCallum et al., 2012). In the former case, random mutations of genotype would not work. The experimental framework of integrating machine learning and artificial evolution can be applied to realize an intelligent system that can evolve and adapt.

We plan to continue the evolution experiment to study how the diversity of music styles can evolve and the possibility of discovering new creation styles that people prefer. Since it is difficult to create a completely new creation style only by blending inheritance, other modes of mutation should be studied. It is also important to study the fitness landscape more thoroughly and in particular analyze its dynamic part such as the frequency-dependent bias (Nakamura and Kaneko, 2019; Youngblood, 2019).

Acknowledgements

The authors thank Go Shibata, Ryo Nishikimi, Du Yicheng, Ryotaro Saito, Takehisa Oyama, Masafumi Hashizume, and the participants of the experiment for their cooperation. This work was supported by JSPS KAKENHI Nos. 21K12187, 21K02846, and 22H03661, JST FOREST Program No. JPMJPR226X, and Novo Nordisk Foundation.

References

Agostinelli, A. et al. (2023). MusicLM: Generating music from text. arXiv:2301.11325.

Biles, J. (1994). GenJam: A genetic algorithm for generating jazz solos. In *Proc. International Computer Music Conference (ICMC)*, pages 131–137.

Ebcioğlu, K. (1990). An expert system for harmonizing chorales in the style of J. S. Bach. *Journal of Logic Programming*, 8(1):145–185.

Hadjeres, G., Pachet, F., and Nielsen, F. (2017). DeepBach: a steerable model for Bach chorales generation. In *Proc. International Conference on Machine Learning (ICML)*, pages 1362–1371.

MacCallum, R. M., Mauch, M., Burt, A., and Leroi, A. M. (2012). Evolution of music by public choice. *Proceedings of the National Academy of Sciences of the USA*, 109(30):12081–12086.

Mauch, M., MacCallum, R. M., Levy, M., and Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, 2(150081):1–10.

Mittal, G. et al. (2021). Symbolic music generation with diffusion models. arXiv:2103.16091.

Moroni, A., Manzolli, J., Von Zuben, F., and Gudwin, R. (2000). Vox Populi: An interactive evolutionary system for algorithmic music composition. *Leonardo Music Journal*, 10:49–54.

Nakamura, E. and Kaneko, K. (2019). Statistical evolutionary laws in music styles. *Scientific Reports*, 9(15993).

Nakamura, E., Shibata, K., Nishikimi, R., and Yoshii, K. (2019). Unsupervised melody style conversion. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 196–200.

Nierhaus, G. (2009). *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer Science & Business Media.

Pachet, F. and Roy, P. (2011). Markov constraints: steerable generation of Markov sequences. *Constraints*, 16(2):148–172.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856.

Singh, R. and Nakamura, E. (2022). Dynamic cluster structure and predictive modelling of music creation style distributions. *Royal Society Open Science*, 9(220516):1–18.

Sturm, B. L. et al. (2016). Music transcription modelling and composition using deep learning. In *Proc. Conference on Computer Simulation of Musical Creativity*.

Youngblood, M. (2019). Conformity bias in the cultural transmission of music sampling traditions. *Royal Society Open Science*, 6(191149).