Real-Time Piano Accompaniment Using Kuramoto Model for Human-Like Synchronization

Kit Armstrong, Ji-Xuan Huang, Tzu-Ching Hung, Jing-Heng Huang, Yi-Wen Liu

National Tsing Hua University
s110061892@m110.nthu.edu.tw

Abstract. Composition and performance in the Western classical tradition represent fields of highly sophisticated artistic endeavor which have not been mastered by AI. Machine performance, though it has become an indispensable tool to composers for creating audio mock-ups, does not appear on the concert stage, where human musicians perform. An eventual goal is a machine that plays a part of a score in real time together with live musicians playing other parts, with results indistinguishable from human efforts. This work focuses on the collaborative aspect of music-making, starting with a behavior-capturing experiment that investigates how musicians adapt their playing to that of others in an ensemble. Using the empirical data thus obtained, we train a Kuramoto model for synchronization which we adapted to the context of score-based collaborative musical performance.

Keywords: Classical music, interpretation, chamber music, expressive performance, automatic accompaniment, rhythmic synchronization, Kuramoto model

1 Introduction

Within the Western classical paradigm of Composer-Performer-Listener, the composer creates a score, and the performers convert it into a performance that the listeners can experience. Both composition and performance have not been fully mastered by AI. For machine composing in general, years of research have developed well-publicized results (e.g. [1–3]), leading to commercial applications. Machine performance is simultaneously more commonplace and yet more distant: it has become an indispensable tool to composers for creating audio mock-ups via sound synthesis tools; however, machines or virtual musicians [4] do not commonly appear on the concert stage to perform alongside with human musicians.

The richness of classical music has much to do with the ways in which performers can create different experiences for the listener out of the same compositions. To do so, an ensemble of performers, individually and simultaneously interpreting their parts of the score, must synchronize with each other in real time. An AI can approach this by starting with *audio transcription*, whose purpose is to provide a machine-friendly representation of musical acoustic information [5, 6], and *score following*, the process by which the machine takes a performance and determines the point in the score to which it is most likely to correspond [7, 8]. Our motivation is to allow a machine, once in possession of these elements, to exhibit human-like behavior acting in real-time as a fellow musician of an ensemble.

In this research, we consider an environment with a single instrument, the piano, to be played jointly by one person and our model. We focused particularly on synchronization in the time-domain, with the goal being to simulate human-like behavior, not to generate a perfectly aligned accompaniment.[1]

2 Model Design

In our approach, we assume that both musicians play exactly according to a known score, and treat each musician's output as a sequence of discrete "note-on" and "note-off" events[2], which are relayed perfectly to the other musician as soon as they occur. By relating the received events to the score, our model ascertains the timing of the other musician and adjusts in real-time the timing of its own future output.

2.1 Kuramoto Model

Previous work on other instances of synchronization have provided us with inspiration regarding the specific mechanism of the adjustment. We have taken the Kuramoto model [9, 10] as a basis. Our implementation follows Heggli et al. [11], who adapted Kuramoto's approach to model human synchronization behavior in a situation where two people were faced with the task of tapping in unison.

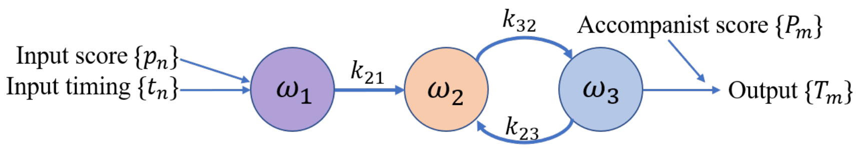

Fig. 1. Adapted Kuramoto model, playing an m-note "output" part together with a n-note "input".

The model consists of 3 oscillators $\omega_1, \omega_2, \omega_3$ that are coupled as in Fig. 1. Their positions are determined by the coupling equations:

$$\frac{d\theta_i(t)}{dt} = \sum_{j \neq i} k_{ij} \sin\left(\theta_j(t) - \theta_i(t)\right) + \Omega_i(t), \tag{1}$$

for $i, j \in \{1, 2, 3\}$, where θ_i represent the positions of the respective oscillators, $\Omega_i(t)$ their intrinsic speed, and k_{ij} the coupling coefficients (with only k_{21}, k_{23}, k_{32} being non-zero, as per Fig. 1).

[1] It is in this regard that our approach distinguishes itself from applications using score-following to automatically accompany a human player.

[2] Our focus is on the piano; for other instruments, this assumption would be less workable.

2.2 Adapting the Kuramoto Model to Musical Scores

Our oscillator model naturally deals with continuous movement, but a musical score, according to our assumptions, consists of discrete events. We reconcile the two paradigms as follows: we define each beat, in the traditional musical sense, as corresponding to a rotation of the oscillator through 2π. The score gives us a sequence $\{p_n\}$ of positions (in beats) at which note onsets in the input part are designated. Denoting t_n the time at which the n^{th} note is actually received, we set $\theta_1(t_n) = 2\pi p_n$. By linear interpolation, we construct a continuous function $\theta_1(t)$ which encapsulates the timing information of the other player.

Analogously, we obtain the output timings $\{T_m\}$ by solving $\theta_3(T_m) = 2\pi P_m$, with $\{P_m\}$ being the beat positions of the output part's notes as given by the score.

In a real-time context, the values of $\theta_i(t)$ for all t are not known beforehand. Thus it is necessary to perform the above calculations for each interval of t at the moment when the information for that interval becomes available. It seemed reasonable to introduce a parameter t_r reflecting reaction time, that is, a delay between receiving information from the input and performing the calculations based on it for adjusting the output.

We denote by $\{T^*_m\}$ the timings of the output events resulting from this process.

3 Implementation

Having set our objective as human-like musical collaboration, the first step was to investigate human behavior in a similar controlled environment. We prepared a series of MIDI recordings containing the melody of well-known music pieces, and invite subjects to accompany these recordings.[3] To train our model to simulate a subject's behavior, we input the same MIDI recording and search for the parameters k_{ij}, t_r that produce the output $\{T^*_m\}$ most similar to the performance of the subject, which we denote $\{S_m\}$:

$$\underset{k_{ij},t_r}{\arg\min} \sum_m (T^*_m - S_m)^2 \tag{2}$$

We proceed to set up the model for accompanying a human subject in real time. First we enter the score of the chosen music piece (i.e. $\{p_n\}$ and $\{P_m\}$ as per Fig. 1). With this information and the MIDI input of the subject, which provides $\{t_n\}$, the model determines $\theta_1(t)$, from which it calculates $\theta_2(t), \theta_3(t)$ by Eq. 1, and finally $\{T^*_m\}$.

3.1 Refinements

Error-handling is outside the scope of this research, which focuses on collaborative aspect of music-making under the assumption of following the score exactly. However, we found during a pilot experiment that having to start over whenever one mistakenly touched a note was unnecessarily frustrating and time-wasting. Therefore, to make our model practically usable, we made it to ignore or automatically rectify common errors.

Furthermore, we implemented a method for following the human player's dynamics based on a running average of velocities of recent input notes.

[3] We use the terms "melody" and "accompaniment" loosely here; the described procedure may be applied to any piece that can be separated into two parts to be played simultaneously.

4 Conclusion

Taking inspiration from models of the synchronization of biological phenomena, we approached the subject of musical performance from the perspective of its collaborative aspects. Our model attempts to emulate basic interactions among ensemble members, which play an important role in classical music-making. As such, it ideally would complement score-interpreting AIs, enabling machines to listen not only to identify cues but to respond and enter musical dialogues in a human-like way.

We have observed playing together with the model to be satisfying to a surprising degree. We performed a sort of Turing test, in which subjects did not know whether they were being accompanied by a human or by our model, resulting in 49/83 (59%) correct guesses, and 18/38 (47%) by listeners present. We consider this an encouraging basis for exploring the possibilities of human-AI collaboration at an increasingly high level of musicality.

References

1. G. Papadopoulos and G. Wiggins, "AI methods for algorithmic composition: a survey, a critical view and future prospects," in *Soc. Artificial Intelligence and Simulation of Behaviour Symp. Musical Creativity*, Edinburgh, UK, 1999, pp. 110–117.
2. K. Bryden, "Developing metrics for evolving a musically satisfying result from a population of computer generated music compositions," in *Proc. IEEE Symp. Computational Intelligence in Image and Signal Processing*, Honolulu, HI, USA, 2007, pp. 171–176.
3. C.-H. Liu and C.-K. Ting, "Computational intelligence in music composition: A survey," *IEEE Trans. Emerging Topics in Computational Intelligence*, vol. 1, no. 1, pp. 2–15, 2017.
4. Y.-J. Lin, H.-K. Kao, Y.-C. Tseng, M. Tsai, and L. Su, "A human-computer duet system for music performance," in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 772–780.
5. L. Su and Y.-H. Yang, "Combining spectral and temporal representations for multipitch estimation of polyphonic music," *IEEE/ACM Trans. Audio, Signal Language Proc.*, vol. 23, no. 10, pp. 1600–1612, 2015.
6. E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2019.
7. R. Dannenberg and C. Raphael, "Music score alignment and computer accompaniment," *Commun. ACM*, vol. 49, pp. 38–43, 2006.
8. R. Agrawal and S. Dixon, "A hybrid approach to audio-to-score alignment," *arXiv:2007.14333*, 2020.
9. Y. Kuramoto, *Chemical Oscillations, Waves, and Turbulence*. New York, NY: Springer-Verlag, 1984.
10. J. A. Acebron, L. L. Bonilla, C. J. Vicente, F. Ritort, and R. Spigler, "The Kuramoto model: A simple paradigm for synchronization phenomena," *Reviews of Modern Physics*, vol. 77, no. 1, pp. 137–185, 2005.
11. O. Heggli, J. Cabral, I. Konvalinka, P. Vuust, and M. Kringelbach, "A Kuramoto model of self-other integration across interpersonal synchronization strategies," *PLoS Comput. Biol.*, vol. 15, no. 10, e1007422, 2019.