

Using a Computer Accompaniment System to Study Group Music Performance

Concept for MDW PhD Program

Martin Bonev

July 25, 2017

1 Introduction

Music performance is often a collective activity, in which multiple musicians form groups and ensembles of various sizes and compositions. While the individual's interpretation of the piece of music is important as it can bring originality and uniqueness, to achieve a cohesive performance the group members must collaborate with each other, i.e., they must communicate information about both the musical structure of the piece and their intentions. Musicians thus try to be as predictable as possible to each other without sacrificing the expressiveness of their performance [L. Bishop & Goebel, 2017]. It has been determined that performers achieve such communication via a variety of mechanisms, including auditory and visual cues, psychological instruments, such as preparatory and regulatory strategies, and real-time cognitive-motor ensemble skills that facilitate mutual adaptation, attention and anticipation [Keller, 2014].

Studying the individual effect and importance in group music performance of each of these communicative mechanisms is the main goal of this PhD project. The difficulty of this task lies in the fact that in group music performance a human performer always uses these mechanisms simultaneously, thus obscuring their individual effects. To tackle this problem, an artificial performer, in the form of a computer accompaniment system (CAS), will be used. Unlike a human performer, every aspect of the behavior of such an accompaniment system will be easy to alter and customize, thus allowing for the creation of different contexts in which the communicative mechanisms could be examined individually. The CAS will also be capable of providing detailed information about the different mechanisms it uses during performance. This PhD project will be structured around two main tasks: the development of the computer accompaniment system to fill the role of the artificial performer and the undertaking of a series of experiments with the CAS as a participant to test the mechanisms that influence group music performance.

Furthermore, the PhD project will be undertaken in the context of the Coordination and Collaborative Creativity in Music Ensembles (COCREATE) project, supported by the Austrian Science Fund (FWF).

2 Preliminary System

There are two basic functions that a computer accompaniment system has to be able to carry out. First, it needs to be able to follow the solo performance and determine the exact position of the soloist in the score. Second, it has to regulate its tempo so that its own performance is temporally aligned with that of the soloist. Various research has been conducted in this area [Dannenberg & Raphael, 2006]. One of the first accompaniment systems was developed in 1984 and implements an efficient dynamic programming algorithm for finding the best match between the solo performance and the music score [Dannenberg, 1984]. Another example is a system that instead of matching the performance to the score, computes probability distributions to determine the most probable position of the played note in the score [Nakamura, Nakamura, & Sagayama, 2016]. Additionally, a system for musical accompaniment in which a computer-driven orchestra follows and learns from a soloist in a concerto-like setting has been developed [Raphael, 2010].

An initial version of a CAS, written in the Python programming language, has been recently implemented by Carlos Eduardo Cancino-Chacón and Amaury Durand. This system will be used and further developed during this PhD project. It is currently capable of adapting to fluctuations in the human performance and tracking their position in the score. Furthermore, it uses that input to determine the tempo and dynamics to align its performance with theirs. The current version of the CAS is not yet capable of real-time performance. The program is structured into three main modules – a *score follower*, a *timing model* and an *expressive performance model*.

2.1 Score Follower

The score follower allows the CAS to determine the exact position of the performance in the score, i.e., the exact note that the performer is hitting at the current moment in time. It receives monophonic MIDI input from the performer, e.g., from a MIDI synthesizer, and compares it with the score input, which is also in MIDI format. Furthermore, the score follower reads the MIDI timings and velocities to extract crucial information about the current tempo and dynamics of the performance. The score follower incorporates Hidden Markov Models (HMM) [C. M. Bishop, 2006b] to predict with high accuracy the position of the performer in the score. Currently, it is only capable of following monophonic music. This is a limitation, as in general, a human performance may include multiple tones played together, e.g, playing a chord.

2.2 Timing Model

The timing model uses the temporal information obtained by the score follower to extrapolate the future evolution of the performance and determine the tempo in which the CAS is to produce its output. This prediction is important, because an accompaniment system that tries to achieve synchrony by waiting to hear a note, and then responding, will always be late. The timing model is based on the Linear Timing model [Goebel & Guggenberger, 2015] and the Joint ADaptation and Anticipation Model (JADAM) of sensorimotor synchronization [Steen & Keller, 2013; van der Steen, Jacoby, Fairhurst, & Keller, 2015], which combines reactive error correction processes (adaptation) with predictive temporal extrapolation processes (anticipation).

2.3 Expressive Performance Model

The expressive performance model allows the CAS to interpret both the score and the human performance, through the MIDI velocities of each played note, in order to alter the dynamics to achieve expressiveness in the performance of the CAS [De Poli, 2004]. It is based on the BASIS Mixer – a prototype system for musical expression [Cancino Chacón & Grachten, 2016; Cancino-Chacón, Gadermaier, Widmer, & Grachten, 2017].

3 Project Aims

The two main aspects of the PhD project are the further development of the computer accompaniment system and its application in examining the different factors that influence interpersonal coordination in expressive group performance.

3.1 Computer Accompaniment System Development

The computer accompaniment system needs to be able to produce performances that are accurate and convincing, so that a musician would be willing to participate in an experiment with the CAS as their co-performer. The system will be written in the Python programming language as it offers great code readability and the existing basic version is already implemented in Python. Increased run-time performance will be achieved with the Cython superset of the programming language. Several improvements to the CAS are planned to be undertaken during this project.

First, the system is to be adapted for on-line matching [Dannenberg & Raphael, 2006], i.e., to work in real-time, which is crucial for co-performing with human musicians.

Second, the *score follower* is to be extended to read and track polyphonic and not just monophonic music. This would significantly increase its functionality and allow the CAS to be used with a variety of musical pieces.

Third, different machine learning algorithms are planned to be implemented in the *timing model* and the *expressive performance model*, which would allow the CAS to learn and recognize patterns, such as chord progressions, changes in dynamics and musical phrases. This would be done to give the system some leadership capacities, in order to

not simply follow the human performer, but to also attempt to guide them into a certain tempo. It would also allow for a greater degree of expressiveness through the recognition and mimicking of performances styles. The *timing model* of the current CAS program is already implemented as an artificial neural network [C. M. Bishop, 2006a].

Fourth, the CAS could be extended to produce ancillary movements with an avatar head. Such movements typically occur during ensemble performance, i.e., head sway and nodding. This would provide the human co-performer with visual cues, which are an important part of interpersonal alignment [Keller, 2014].

Finally, implementing features that allow real-time interpretation of pianists' body movements may be possible. This would be a further extension of providing an avatar head and would allow the CAS to also read ancillary movements from the human performer, thus further improving its group performance capabilities.

3.2 Tests and Experiments

After the computer accompaniment system is successfully implemented and capable of interpreting and performing music in a convincing way, it will be used in experiments to examine the different mechanisms for interpersonal coordination in expressive group performance.

The CAS will have a major applicability in such experiments in that it will have nearly unlimited technical facility, allowing the coordination of nearly arbitrarily fast notes and complex rhythms [Dannenberg & Raphael, 2006]. Furthermore, the behavior of the system can be easily altered, e.g., it can be made less or more responsive to the input from the human performer or it can be made to encourage its co-performer to play in a slower or faster tempo. These alterations can provide extensive control into the context and environment of the experiments, in a way that would be impossible with a human performer.

The experiments will consist of a number of musicians performing pieces of music with the computer accompaniment system. The CAS will then participate in storing separate pieces data from the performance for each of the factors that influence interpersonal coordination in expressive ensemble performance [Keller, 2014]. For example, online cognitive-motor ensemble skills, such as phase correction and imitation, will be monitored by the *tempo model*, through note onset information from the output of a MIDI keyboard. Similarly, auditory cues like intensity and articulation will be monitored by the *expressive performance model* through the MIDI velocities of the notes performed by the musician. Gathering separate information for the individual aspects of group music performance would give valuable insights into the importance and weight of each one of them, thus presenting a much clearer perception of the whole phenomenon.

Additionally, the difference in the movement and gestures of the musicians during performance can be investigated with a motion capture system. This could provide information about visual cues, such as instrumental movements and ancillary movements, and their influence on the whole performance.

Furthermore, both human-human and human-machine musical performances will be examined and the differences between them will be compared. This would be done

by involving another musician or by hiding from the performer the fact that they are playing with a computer. These experiments should provide information whether such differences are a result of the inability or unwillingness on the part of the musicians to conceive of computer partners as intentional and creative. They could also give a small insight into which characteristics of the computer accompaniment system are most attractive and helpful and what updates could be made, so that intelligent music systems, such as the CAS, could be turned into useful tools for musicians in the area of expressive group music performance [Widmer, 2016].

4 Applications and Conclusion

This PhD project will have practical and theoretical implications for the fields of joint action, performance science, and human-computer interaction. The development of the CAS will benefit future studies on performance coordination and provide practical applications for performing musicians and music students. By tuning and altering the parameters of the CAS, separate aspects of a music performer can be explored and their importance and value measured. Additionally, the accompaniment system would be a helpful and convenient tool for performers to practice with. Finally, the results of the experiments would allow the examination of people’s perceptions of their partners’ creative abilities, responsiveness and intentionality.

References

- Bishop, C. M. (2006a). Neural networks. In C. M. Bishop (Ed.), *Pattern Recognition and Machine Learning* (pp. 225–290). New York: Springer.
- Bishop, C. M. (2006b). Sequential data. In C. M. Bishop (Ed.), *Pattern Recognition and Machine Learning* (pp. 605–652). New York: Springer.
- Bishop, L., & Goebel, W. (2017). Music and movement: Musical instruments and performers. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition* (pp. 349–361). Routledge.
- Cancino Chacón, C., & Grachten, M. (2016). The basis mixer: A computational romantic pianist. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*. (Late Breaking/Demo extended abstract)
- Cancino-Chacón, C. E., Gadermaier, T., Widmer, G., & Grachten, M. (2017). An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music. *Machine Learning*, 106(6), 887–909.
- Dannenberg, R. B. (1984). An on-line algorithm for real-time accompaniment. In *International Computer Music Conference* (Vol. 84, pp. 193–198).
- Dannenberg, R. B., & Raphael, C. (2006). Music score alignment and computer accompaniment. *Communications of the ACM – Music information retrieval*, 49(8), 38–43. doi: doi:10.1145/1145287.1145311

- De Poli, G. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research*, 33(3), 189–202.
- Goebel, W., & Guggenberger, D. (2015). Tappingfriend—an interactive science exhibit for experiencing synchronicity with real and artificial partners. *Proceedings of the Third Vienna Talk on Music Acoustics*, 227–230.
- Keller, P. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in Music Performance: Empirical Approaches Across Styles and Cultures* (pp. 260–282). Oxford: Oxford University Press.
- Nakamura, T., Nakamura, E., & Sagayama, S. (2016). Real-time audio-to-score alignment of music performances containing errors and arbitrary repeats and skips. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(2), 329–339.
- Raphael, C. (2010). Music plus one and machine learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (pp. 21–28).
- Steen, M. C., & Keller, P. E. (2013). The ADaptation and Anticipation model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7(253), 1–15. doi: doi:10.3389/fnhum.2013.00253
- van der Steen, M. M., Jacoby, N., Fairhurst, M. T., & Keller, P. E. (2015). Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Research*, 1626, 66–87.
- Widmer, G. (2016). Getting closer to the essence of music: The Con Espressione manifesto. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(2), 19.