# EXPLORING MEANING AND INTENTION IN MUSIC CONDUCTING

*Roger B. Dannenberg*
Carnegie Mellon University
School of Computer Science

*Daniel P. Siewiorek*
Carnegie Mellon University
School of Computer Science

*Noel Zahler*
Carnegie Mellon University
School of Music

## ABSTRACT

Conducting is a high-level form of expressive musical communication. The possibility of human-computer interaction through a conducting-based interface to a computer performance system has attracted many computer music researchers. This study explores conducting through interviews with conductors and musicians and also through accelerometers attached to conductors during rehearsals with a (human) orchestra and chamber music group. We found that "real" conducting gestures are much more subtle than "textbook" conducting gestures made in the lab, but we observed a very high correlation between the smoothed RMS amplitudes of conductors' wrist acceleration and the ensembles' audio.

## 1. INTRODUCTION

Conducting remains one of the most elusive and intractable of music skills. It relies on one individual's expert knowledge of a score (an abstract graphic representation of the elements of sound as produced by traditional and electronic instruments) and an understanding of the idiomatic and historically relevant performance practice of the work's origin. It has no true lexicon of gesture (other than a loosely organized set of beat patterns), and there is no general consensus regarding what is communicated or the meaning of that communication.

Our recent work aims to discover some of the principles that govern both great and poor performances using scientific techniques to complement or substantiate (or possibly refute) wisdom gleaned from musical tradition. We believe that a highly interdisciplinary approach to the study of conducting will enable progress and potential breakthroughs in this complicated domain. Musicians pose substantial questions and offer insights that direct research in productive directions. Technologists and scientists are needed to build systems for gesture sensing and to develop new mathematical, statistical, and algorithmic tools for data processing and analysis. A compelling feature of this work is the interplay of motivations and skills among all the participants.

There are many reasons to study conducting. From a musicological perspective, conducting builds links between notation, theory, gesture, and sound. Conducting offers insights into music and music practice, with implications for music education. Conducting also has applications in the world of multimedia and games. There is much interest in conducting games and already a couple of consumer games and museum installations. Moreover, gestural communication has general applications in interactive entertainment. In the computer music field, research into conducting offers a new avenue for composers concerned with how virtual performers interpret conductors' gestures and how high-level musical intentions are translated into (virtual) instrument-specific controls and sounds. In particular, an understanding of conducting could lead to interactive music performance systems where computerized performers coordinate with humans and where a human conductor can control expressive aspects of synthesized sounds in an intuitive manner. There are fascinating implications for both new technology and new music. Finally, conducting offers a rich set of problems to stimulate the design of future wearable computers, sensors, and gesture analysis techniques. For example, we have seen an evolution from 2 to 3 accelerometers in the eWatch device (see Figure 1), and we look forward to working with the next generation of small 6-degree-of-freedom wireless sensors.

## 2. RELATED WORK

Computer-music conducting systems have been created and studied by many researchers at least as far back as the 1980's [1]-[8]. This work began to show evidence of the subtlety of conducting and tempo indication in "real" human conducting. Baird and Izmirli considered the difference in phase (or lag) between conducting and

performers [9]. Luck and Toiviainen [10] note that conductors in lab studies and studies of individual musicians obtain behaviors that differ from studies in more ecologically valid contexts. Luck and Nte [11] survey studies of conducting in natural contexts and introduce motion capture techniques to study conducting gestures and their interpretation.

## 3. STUDIES

Our goal has been to study the nature of information and communication in music conducting. We began by considering many avenues of investigation. As an interdisciplinary group, we looked at conducting from the viewpoint of the conductor (what aspects of music need to be communicated to musicians?), the musician (what information do musicians need to perform well?), music information retrieval and psychology (how tightly do performers synchronize, how can beats be detected in musical and gestural signals?), and human-computer interaction (in particular, relating conducting to work on dialog).

Non-verbal cues are the basis of communication in making music with large ensembles, and our study begins to ask questions regarding how both people and machines can decode these gestures. Conducting is much richer than merely beating time. It is clear that the measurement of physical gesture alone does not give us a comprehensive understanding of the meaning that is related between the conductor and ensemble. The rules of engagement need to be probed deeply and new tools borrowed from psychology, the computational sciences and the social sciences.

We began by capturing data from the performance of a few pieces from an orchestra rehearsal. We used a 2-axis accelerometer (the eWatch shown in Figure 1) and one video camera [12]. As might be expected, the accelerometer data was ambiguous because of the unknown orientation, and the resulting gestures could not be resolved into beats, at least not by any approaches we tried.

Following this initial foray, we captured two rehearsals, one by a small ensemble, and one by a large orchestra, using video (3 points of view), audio, and a new 3-axis accelerometer version of the eWatch. We also interviewed the conductors and selected musicians after the performances to get their insights into what information was communicated and how it was accomplished.

### 3.1. Interviews

A set of interview questions was prepared in advance for the conductor and another set was prepared for musicians. Sample questions for the conductor include:

**Figure 1. The eWatch (Daniel Meyer, conducting)**

- Can you characterize the elements of performance you communicated to the musicians?
- Were there places in the rehearsal that went better than other places?
- What made the communication better in those areas?

Sample questions for the musicians include:

- Can you characterize the elements of performance that were communicated by the conductor?
- How did he communicate those elements?
- What were the most and least effective gestures used?

After our initial interviews, we decided it would be best to interview the conductors while watching video of the rehearsals. In these interviews, we asked the conductors to simply explain what they see and try to remember what was happening. This was a very useful interviewing technique because watching the conducting apparently helped conductors to recall many details of their intentions and thought processes on a moment-by-moment basis. Because of the exploratory nature of our work, we used an informal analysis of the interviews, hoping to gain some insights that would guide further research. Some of the impressions we obtained from interviews are summarized in the following paragraph.

In many situations, the "standard" conducting gestures are not the focus of attention. This can be compared to driving a car, where steering appears to be the main task, but in fact the driver is almost unaware of the details of steering and actually may be thinking about many other

higher-level tasks. In conducting, tasks include preparing for important musical events or transitions, getting the attention of one or more musicians, and indicating specific instructions regarding timing, phrasing, or dynamics. Conductors commented that giving clear indications of every beat was not necessary and not their intention. One conductor disparaged his conducting at one point saying he was on "autopilot" and not thinking about the details of the musical line at that time. It was noted that conducting in a performance has important references to experience in rehearsals, so that the gestures seen by an audience (or by a computer interface) may not be entirely meaningful without knowledge of what transpired in previous rehearsals. Musicians commented that communication with the conductor was very intuitive and involved eye contact and facial expressions as well as arm or baton motion. Some descriptions of gestures and their meaning, such as making small gestures to get the attention of the performer, were directly contradicted by the conductor's descriptions of gestures and meaning. (It could easily be the case that gestural meaning is so context dependent that both interpretations could be true in the correct setting.)

### 3.2. Data Analysis

Our initial analysis of accelerometer data, like many previous studies, aimed to locate beats and tempo. We applied existing algorithms based on machine learning [13], but found that these algorithms do not work well for the recognition of "real world" conducting gestures. On clean data from the lab, the gesture detection ratio was 95% but with "real world" data collected in this study, the detection ratio dropped to less than 20%. This could be a result of the large amount of variation and the relatively small set of training data, but could also indicate that conductors are not always concerned with the clear communication of beats. The latter is consistent with interviews with conductors and musicians.

While our conducting gesture data seems to be very hard to interpret, there is also evidence that at least some beat gestures are very meaningful and useful to musicians. We spent some time looking for "reliability" measures. For example, one might expect amplitude peaks of acceleration to be higher at decisive synchronization points in the music. At best, we found very limited support for this conjecture in the data. Perhaps more data could lead to statistical significance. Alternatively, there may be better ways to identify the "true" points of musical interest and to characterize "decisive" beat gestures.

While working on this analysis, we had displays of audio waveforms and accelerometer waveforms in parallel tracks of an audio signal editor. To our surprise, we noticed that the two waveforms were strikingly similar. To measure this objectively, we computed the RMS (root mean squared, or the square root of average power) of the audio waveform and the RMS of the (3D) accelerometer signal using various smoothing windows. As hoped, we found a strong numerical correlation between conducting gestures and the loudness of the music. While conductors and musicians ascribed various (and sometimes conflicting!) meanings to the size of gestures, we found statistically significant correlations between the average audio amplitude and the average acceleration, both expressed as the RMS over 10-second windows. For the two ensemble/conductor combinations we studied, these correlations were 0.85 and 0.52. (See Figure 2).

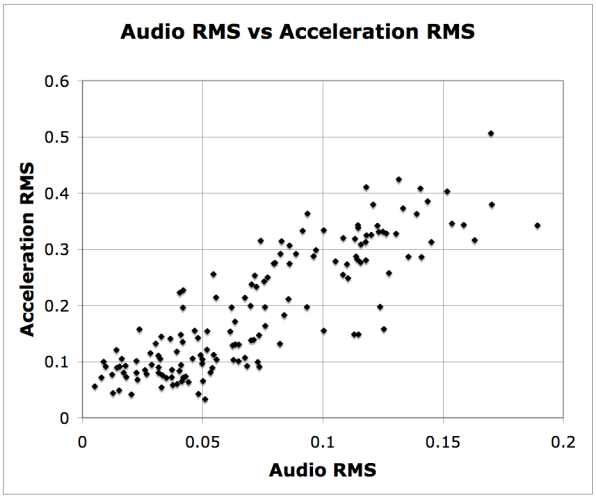


**Figure 2. Scatter plot of short-term average audio amplitude (y-axis) vs. short term average magnitude of conducting gesture acceleration (x-axis). Correlation = 0.85.**

## 4. CONCLUSIONS AND FUTURE WORK

Previous work has largely focused on obtaining tempo and beat information from conducting gestures. Our work provides evidence that conducting by professionals in "real" musical performance is not oriented toward providing a clear beat, at least not all of the time. Conductors have many other tasks. It is interesting but not too surprising to find a very high correlation between musical amplitude and conducting acceleration, which would also correlate with the velocity and size of gestures. This is the first actual measurement of this correlation to our knowledge.

We cannot establish any causality in this correlation. Moreover, dynamic values are based on relative, not absolute, values, and every group of performers will interpret a dynamic marking based on many variables (the relative strength of the performers, the qualities of the instruments used in the ensemble, the acoustical space, etc.) An interesting question for future research, then, is to what extent is the conductor is directing dynamics to the orchestra, is the conductor is responding to the orchestra's

dynamics, or are both simply responding to the underlying music composition?

In the course of this research, conductors who teach have asked about the possibility of virtual orchestra conducting systems for training. It is interesting that one was interested in detecting when students successfully cue entrances, while the other was interested in timing and tempo. The latter did not feel it was important to control tempo continuously, but only at the beginning of a piece and at important places such as meter and tempo changes. This provides additional evidence that there are interesting conducting gesture recognition problems to be solved, but the idea that a computer should (or could) detect every beat and follow the conductor closely at all times is probably not well founded. (This assumes that the goal is to model human orchestras; there have certainly been systems to conduct machines where new conducting techniques must be learned and applied. There is nothing wrong with these systems, but we should be clear that they are not accurately modeling traditional conducting.)

Conducting-oriented interfaces can be viewed as one component of a much larger framework for musical interaction. Although considerable work remains to be done on conducting, it is also interesting to consider the more general problem of coordinating human and computer musicians, using a variety of interaction techniques, including tapping, gesture sensing, and audio analysis. Because conducting seems to rely on musicians to apply basic musical knowledge, rehearsal knowledge, score following, and visual cues (such as watching the concert master's bow) as well as following the conductor's baton, a systematic study of all these techniques may lead to the most interesting and musical interactive performance systems.

## 5. ACKNOWLEDGEMENTS

This work was "conducted" with additional help from Brian French. We would like to thank the Pittsburgh Youth Symphony, Daniel Meyer, the Pittsburgh New Music Ensemble, and the Edgewood Symphony for help and participation in our studies. Dominik Schmidt ran our new data through his conducting gesture software, allowing us to compare results from "lab" data to those from "real-life" data.

## 6. REFERENCES

[1] Matsushima, T., S. Ohteru, S. Hashimoto. "An Integrated Music Information Processing Systme: PSB-er," *Proceedings of the 1989 International Computer Music Conference,* San Francisco: International Computer Music Association (1989), pp. 191-198.

[2] Morita, H., S. Hashimoto, and S. Ohteru. "A Computer Music System that Follows a Human Conductor," *IEEE Computer 24, 7* (1991), pp. 44-53.

[3] Dannenberg, R. and K. Bookstein. "Practical Aspects of a Midi Conducting Program," *Proceedings of the 1991 International Computer Music Conference,* San Francisco: International Computer Music Association (1991), pp. 537-540.

[4] Lee, M., G. Garnett, and D. Wessel. "An Adaptive Conductor Follower," *Proceedings of the International Computer Music Conference,* San Francisco: International Computer Music Association (1992), pp. 454-455.

[5] Usa, S. and Y. Mochida. "A Multi-Modal Conducting Simulator," *Proceedings of the International Computer Music Conference,* San Francisco: International Computer Music Association (1998), pp. 25-32.

[6] Ilmonen, T. and T. Takala. "Conductor Following with Artificial Neural Networks," *Proceedings of the 1999 International Computer Music Conference,* San Francisco: International Computer Music Association (1999), pp. 367-370.

[7] Nakra, T. Marrin. "Synthesizing Expressive Music through the Language of Conducting," *Journal of New Music Research 31 (1)*, (June 2002), pp. 11-26.

[8] Lee, E., T. Karrer, and J. Borchers. "Toward a Framework for Interactive Systems to Conduct Digital Audio and Video Streams," *Computer Music Journal 30(1)*, (Spring 2006), pp. 21-36.

[9] Baird, B. and O. Izmirli. "Modeling the Tempo Coupling between an Ensemble and the Conductor" *Proceedings of the International Computer Music Conference* (2001).

[10] Luck, G. and P. Toiviainen. "Ensemble Musicians' Synchronization With Conductors' Gestures: An Automated Feature Extraction Analysis," *Music Perception 24, 2* (2006), pp. 189-200.

[11] Luck, G. and S. Nte, "An Investigation of Conductors' Temporal Gestures and Conductor Musician Synchronization, and a First Experiment," *Psychology of Music 36, 1* (2008), pp. 81-99.

[12] Schmidt, D., R. Dannenberg, A. Smailagic, D. Siewiorek, and B. Bruegge. "Learning an Orchestra Conductor's Technique Using a Wearable Sensor Platform," *11th IEEE International Symposium on Wearable Computers, 2007*. New York: IEEE (2007), pp. 113-114.

[13] Schmidt, D., *Acceleration-Based Gesture Recognition for Conducting with Hidden Markov Models*. Diploma Thesis, Munich: Ludwig-Maximilians-Universitaet (2007), http://eprints.comp.lancs.ac.uk/1655/