The *UBS Virtual Maestro*: an Interactive Conducting System

Teresa Marrin Nakra
Department of Music
The College of New Jersey
P.O. Box 7718
Ewing, NJ 08628
nakra@tcnj.edu

Yuri Ivanov
Mitsubishi Electric Research Labs
201 Broadway
Cambridge, MA 02139
yivanov@merl.com

Paris Smaragdis
Advanced Technology Labs
Adobe Systems Inc.
275 Grove Street
Newton, MA 02466
paris@adobe.com

Chris Ault
Interactive Multimedia Program
The College of New Jersey
P.O. Box 7718
Ewing, NJ 08628
ault@tcnj.edu

Abstract

The *UBS Virtual Maestro* is an interactive conducting system designed by Immersion Music to simulate the experience of orchestral conducting for the general public attending a classical music concert. The system utilizes the *Wii Remote*, which users hold and move like a conducting baton to affect the tempo and dynamics of an orchestral video/audio recording. The accelerometer data from the *Wii Remote* is used to control playback speed and volume in real-time. The system is housed in a UBS-branded kiosk that has toured classical performing arts venues throughout the United States and Europe in 2007 and 2008. In this paper we share our experiences in designing this standalone system for thousands of users, and lessons that we learned from the project.

Keywords: conducting, gesture, interactive installations, *Wii Remote*

1. Introduction

In early 2007, representatives of Jack Morton, Inc., a global experiential marketing agency, contacted Immersion Music, Inc. to build an interactive conducting system for the UBS AG Corporation, headquartered in Switzerland. Widely recognized for its substantial support of classical music organizations, the client was interested in designing a computer system to travel along with a European youth orchestra, whose tour was sponsored by UBS. The system was envisioned to provide a compelling interaction and enhance the concert experience by allowing audience members to try their hand at conducting the orchestra in its signature repertoire.

Immersion Music, a not-for-profit music technology company headquartered in the Boston area, undertook this challenge and developed the *UBS Virtual Maestro*, an interactive application that was built for this purpose. The system has been touring major concert venues in the U.S. and Europe since November 2007. Housed in a portable kiosk with a 42" plasma screen and an overhead parabolic speaker (**Figure 1**), it has seen thousands of users and lived up to most expectations in the course of its existence.

Figure 1. The *Virtual Maestro* System in Los Angeles, November 2007 (photo courtesy Jack Morton Inc.)

The system was designed and developed during the Summer of 2007, and tested and integrated into the kiosk in Fall 2007. It debuted in November 2007 at the annual UBS Thanksgiving concert at the Boston Symphony Orchestra, with over 2000 in attendance. It remained at Boston's Symphony Hall for a few weeks, before moving on to other classical venues around the U.S. and Europe. Other stops on the tour included the Walt Disney Concert Hall in Los Angeles, Avery Fisher Hall in New York City, the Kimmel Center in Philadelphia, orchestra halls in Utah, Seattle, Charlotte, Cleveland, and Minnesota, and the Ravinia, Verbier, and Montreux Jazz Festivals.

The design challenge was to build a complicated software/hardware solution that would be simple to set up and maintain while on a demanding travel schedule. In particular, the baton was an important but vulnerable design element. Unlike other similar systems for the public, it would have to be rugged enough to withstand the tremendous stresses of the tour as well as the abuse of potentially several thousand users at each location.

While a custom baton would have had advantages, it would have been expensive to manufacture and repair. Ultimately, the Nintendo *Wii Remote* was selected; the rationale for this decision is described below.

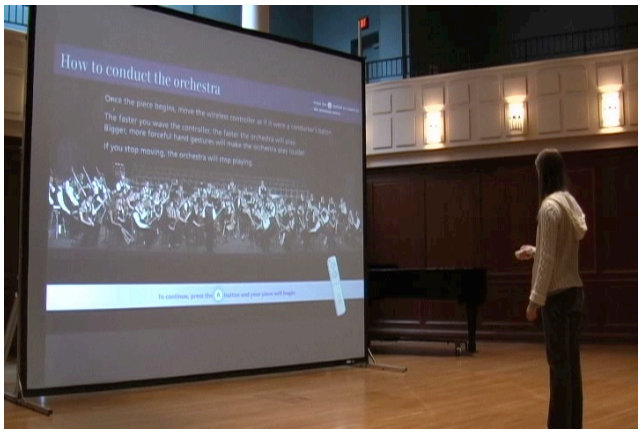

Figure 2. Student using UBS *Virtual Maestro* system at The College of New Jersey

1.1 Description of Interaction

The *UBS Virtual Maestro* system was designed for audiences of any age. Users pick up a baton from a music stand at the front of the kiosk, and navigate through three selection screens. The first screen invites them to conduct the orchestra and press "A" to begin. The second screen offers users to select from among three works, including excerpts from the 4th movement of Tchaikovsky's *Symphony no. 5*, Rossini's *William Tell Overture*, and the 4th movement of Berlioz' *Symphonie Fanstastique*. The third screen provides instructions on "How to conduct the orchestra," suggesting that users should move the interface like a conductor's baton. Users are instructed that the speed of their gestures will affect the speed of the orchestra's performance; that the force of their gestures will affect the loudness; that soft and gentle gestures will create a quiet performance; that if they stop moving the "baton," the performance will stop; and that strong gestures on the beat will cause the orchestra to follow. Users can move the "baton" freely; typically, however, they mimic the gestures used by classical conductors.

UBS provided four orchestral audio/video recordings, taken in the summer of 2007, including a tuning-up sequence. The project was documented in a video produced by the Office of Public Affairs at The College of New Jersey, which is available on Youtube under the title "Wii Conductor Hero" [10]. Figure 2 presents an image of a user interacting with the third screen.

2. Previous work

Curtis Roads once wrote: "the original remote controller for music is the conductor's baton."[9] The *Virtual Maestro* system has represented an attempt to combine the ubiquity of the remote-controller with the expressive potential of conducting. This project has benefited from the prior work of numerous predecessors, spanning decades of research in controllers and interactive music systems. In particular, we acknowledge several prior successful interactive conducting systems, and admire in particular the important contributions of Max Mathews, Joseph Paradiso, Donald Buchla, and Jan Borchers.

Max Mathews, whose work has deeply influenced so many areas of computer music, invented the "Daton" in the 1980s at Bell Labs. He then collaborated with Bob Boie to extend its functionality and developed the "Radio Baton," which has been used by composers and conductors to perform electronic versions of orchestral scores. It was the first documented system that enabled the possibility of interactive conducting, and has since been used in numerous concerts, demonstrations, and exhibits [6].

In a 1996 collaboration with Joseph Paradiso et al of the MIT Media Lab, Teresa Marrin Nakra designed the "Digital Baton" sensor interface to conduct a MIDI-based music system in Tod Machover's "Brain Opera." The baton was a hand-held gestural interface that was designed to be used like a traditional conducting baton by practiced performers. It featured a ten-ounce molded polyurethane handle that held a thin plexiglass tube and infrared LED at the tip. It tracked the position of an IR LED at its tip, and used a 3-axis accelerometer array to detect beats and large gestures. It also had five force sensors that measured finger pressure with a resolution of seven bits. The number and variety of sensors on the "Digital Baton" enabled highly expressive control over the parameters of an electronic music performance. In the period between 1996 and 1998 the "Digital Baton" was featured in hundreds of performances of the "Brain Opera" worldwide [5, 8].

Teresa Marrin Nakra's "Conductor's Jacket" project (1998-2000) was undertaken to improve upon the "Digital Baton" by incorporating new physiological and gesture-capture sensors. The "Conductors Jacket" is a wearable physiological monitoring system that has been built into the clothing of an orchestral conductor; it was designed to provide a testbed for the study of emotional expression as it is conveyed through conducting technique. Its design reflected the need for an array of different sensor inputs that would be unencumbering to the wearer while also capturing both the expressive and mechanical aspects of conducting technique. The "Conductors Jacket" was used to gather and analyze data from professional and student conductors in Boston during numerous rehearsals and performances [7].

Donald Buchla's "Lightning" baton interface consists of a pair of handheld infrared tracking devices that output MIDI data on position coordinates, beats, and gestures. The "Lightning" wands have been commercially available since 1991 and used as batons in several prior interactive conducting systems [8]. In 1998, Jan Borchers et al [1] built an interactive exhibit called the "Personal Orchestra," which has since been on continuous display at the *Haus der Musik* in Vienna. This system uses the Buchla "Lightning" batons to control the playback speed and volume of an audio/video recording of the Vienna Philharmonic. This system allows the users to seamlessly switch between tracks of pre-time-stretched audio and video segments. The "Personal Orchestra" provides some motivational and critical feedback to users in the form of prerecorded commentary and applause from the orchestra.

In 2003, Teresa Marrin Nakra and Jan Borchers collaborated to build a new interactive conducting system for the public. This project combined some of the gesture-sensing approaches from Nakra's "Digital Baton" and "Conductor's Jacket" projects with the software interaction paradigm of the "Personal Orchestra." Borchers updated his audio system for this project, developing a new real-time audio stretching algorithm. The new system, called "You're the Conductor," featured footage from the Boston Pops Orchestra and was on display at the Boston Childrens Museum in 2003, 2007, and 2008. It toured to Children's Museums around the United States through 2004-2006 [3].

Finally, in recent years, musical video games such as Nintendo's *Wii Music* have provided interesting and useful models for interactive music systems. In *Wii Music*, there is a conducting simulation that allows up to four conductors to play simultaneously, and assigns scores to each performance. The algorithm for judging the success of a performance seems to be weighted towards consistency and synchronization. Playback seems to involve discrete note events in a MIDI-like framework, different from the continuous time-stretching approaches taken by other recent conducting simulations. The *Wii Music* game was not commercially available in the U.S. prior to the *Virtual Maestro*, but its wide commercial availability and usage by the general public must be acknowledged as relevant to the continued future success of the *Virtual Maestro*.

The *Virtual Maestro* system differs from other prior conducting interfaces, simulations, and exhibits, in that it is specifically designed to enrich the audience experience in concert hall lobby settings. As a result, it has been designed not just for entertainment but also for education and enrichment. A variety of supplemental education programs that were offered to children using the *Virtual Maestro* are described in section 4, below. The *Virtual Maestro* also differs from videogames that feature conducting in that it does not attempt to make a qualitative assessment of the accuracy of the user's technique or tempo consistency/synchronization, and does not give the user feedback other than the direct experience of controlling the contours of tempo and volume.

3. System Specifics

The *Virtual Maestro* system consists of a *Wii Remote* controller, a USB Bluetooth transceiver, a PC running Windows XP, a fast consumer video card from NVidia, and a professional-quality audio card from E-Mu. Custom software is used to acquire and process the incoming *Wii Remote* data, guide the user through three selection screens, run the conducting application, and coordinate the real-time audio- and video-processing algorithms.

The movements of the *Wii Remote* affect the tempo and dynamics of an orchestral video/audio recording. Incoming accelerometer data from the *Wii Remote* are used to control playback speed and volume in real-time, while some of the buttons are mapped to users' selections at the beginning of the interaction.

3.1 Rationale for Selecting *Wii Remote*

The Nintendo Wii game console has been tremendously successful in attracting large numbers of so-called "casual gamers" during the past two years. This is perhaps due to the intuitive design of its wireless, motion-sensitive interface. Rather than remembering an arcane combination of buttons to execute a game action, such as throwing a ball, a user simply swings the Wii controller over his or her head in a typical throwing motion. The fact that a natural action leads to an expected result serves to lessen intimidation on the part of many users, and also allows users to begin interacting with the system quickly, without studying complex instructions.

A physical baton interface that is immediately understandable and intuitive is essential to the success of the *Virtual Maestro* in the crowded and hurried setting of a concert hall prior to performances or during intermissions (Figure 3). The desire to maintain that immediacy also drove the design of the on-screen interface, which favors simple graphical menus and instructive animations over lengthy written information. The performance footage features generous bands of negative space both above and below the orchestra on the stage, allowing user interface elements and the orchestra to appear on-screen simultaneously, further minimizing the barrier between learning to use the system and actually experiencing it.

In addition to the intuitive aspects of the *Wii Remote*, there were also practical considerations in adopting it. The *Virtual Maestro* was specified to be a stand-alone system that could easily be packed, transported and set up by non-experts in remote locations. This requirement motivated the decision to select an off-the-shelf controller instead of building a custom baton interface. Although this choice sacrificed some sophistication in the sensors and ergonomics, it favored more practical features of robustness, usability, and longevity. We also chose standard wireless communication protocols and commonly

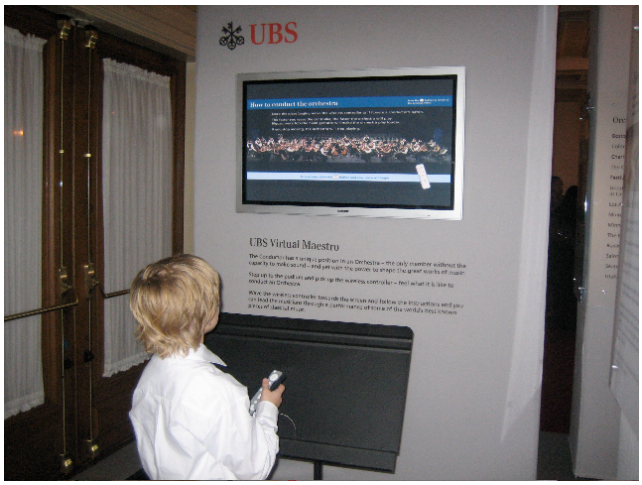

Figure 3. User conducting with the *Wii Remote*

available hardware to make the inevitable required repairs straightforward to manage remotely. Ultimately, the *Wii Remote* has been a successful choice because it is rugged, easy and cost-effective to replace, ergonomic, and not too bulky.

3.2 Acquiring and Processing *Wii Remote* Data

The *Virtual Maestro* system uses a *Wii Remote* to gather acceleration information from the user's hand. The *Wii Remote* device includes both infrared fiducial marker tracking capabilities for estimation of its position and orientation, as well as a 3-axis accelerometer package for direct measurement of accelerational forces.

We decided to use the accelerometers on the *Wii Remote* instead of the infrared optical position/orientation sensor. Given that our system would be primarily situated in a public location without attendants to manage it, we accepted the necessity of a tradeoff between accuracy and robustness. We opted for the noisier but more robust accelerometer data, because they are not subject to unknown lighting and crowding (occluded sight-line) conditions. In addition, we anticipated that big conducting gestures would cause the infrared camera on the *Wii Remote* to move out of view of the Sensor Bar. Forcing users to point at the Sensor Bar while conducting, we determined, would have been an inappropriate and unrealistic constraint for the type of gestural interaction the *Virtual Maestro* system was attempting to encourage.

The accelerometer data is filtered by a recursive low-pass filter and subsequently integrated to obtain the estimated velocity. The running mean of this signal is removed to avoid drift. Then a simple peak detection algorithm calculates the positions of the peaks in the velocity magnitude and estimates the beat rate as well as the relative magnitude of the force in the conducting gestures.

To each sample of the accelerometer data, $\mathbf{r}_t^a$, we first apply a recursive low-pass filter to remove spurious noise and obtain a de-noised sample, $\mathbf{r}_t$:

$$\mathbf{r}_t = \mathbf{r}_{t-1} + \alpha\left(\mathbf{r}_t^a - \mathbf{r}_{t-1}\right)$$

where α is a constant from a range between 0 and 1 typically set to a value close to 1 to achieve fast signal tracking. The running mean of this quantity is calculated to estimate the signal baseline, or average acceleration. We use a similar recursive relation with a parameter, β, such that $0 < \beta < \alpha < 1$:

$$\bar{\mathbf{r}}_t = \bar{\mathbf{r}}_{t-1} + \beta\left(\mathbf{r}_t - \bar{\mathbf{r}}_{t-1}\right)$$

The resulting estimate for an average acceleration removes the bias from the signal such that all deviations from it are due to beat gestures. The magnitude of the residual velocity is calculated by integrating the residual and finding its norm:

$$M_t = \left\| \sum \mathbf{r}_t - \bar{\mathbf{r}}_t \right\|$$

This is now a scalar function that peaks in synchrony with hand gestures. In order to estimate the instantaneous beat rate of the conducting gesture, we can now simply calculate the time interval between peaks of M_t:

$$BPM = \frac{1}{T_{peak_t} - T_{peak_{t-1}}}$$

The playback beat rate is again estimated by a recursive relation similar to the above with a parameter $0 < \gamma < 1$ that defines the overall system sensitivity to change in conducting tempo.

The reason we selected the velocity magnitude for tempo control is because we wanted to accommodate a wide variety of gesture styles by a large number of users. This choice made our algorithms robust enough to handle gestures made in any direction or orientation. We were able to accurately detect the beat rate regardless of how the user held the *Wii Remote*, or where he/she pointed.

Additionally, the gesture magnitude range was continuously estimated to compensate for the user's height and arm length. This allowed users of different heights to use the system, with relatively equivalent results – that is, tall people didn't always generate loud volumes and short people didn't always generate soft volumes.

Once the beat rate and gesture magnitude values are computed, they are then passed to a custom audio and video player to play the recordings at the estimated tempo and volume. Despite a time lag that the beat estimation algorithm introduces into the system, we preferred this method to the more tightly-coupled, direct control of tempo demonstrated in similar earlier systems because our method of playback control produces smoother, more audibly pleasing, results and avoids abrupt jumps and transitions in tempo and volume. In this mode, the actual

lag becomes irrelevant and doesn't play any role in the user experience.

3.3 Audio Processing Algorithms

In order to allow a user to control the tempo of the performance, we need to be able to modify, in real-time, the playback speed of the audio recording. This necessitates the use of an algorithm that is able to change the speed of the recording while also maintaining its pitch characteristics. There are various algorithms that have been developed to perform this function, each best suited for specific types of signals. In this case, where the input is a complex polyphonic recording, a safe approach is to use an algorithm based on the phase vocoder [2], which is best suited for complex multi-pitch and potentially inharmonic signals.

The approach we chose to use is a straightforward implementation. Each music recording we used was segmented in 48ms frames $\mathbf{s}_t$ with an overlap of 32ms. We applied on each $\mathbf{s}_t$ a Hann window and then performed its Discrete Fourier Transform (DFT). We then extracted the magnitude and phase of the resulting set of complex frames. More precisely, for each $\mathbf{s}_t$:

$$\mathbf{f}_t = \mathrm{DFT}(\mathbf{s}_t)$$

$$\mathbf{f}_t^m = \|\mathbf{f}_t\|$$

$$\mathbf{f}_t^p = \angle \mathbf{f}_t$$

where $\mathbf{f}_t$ is the DFT for frame $\mathbf{s}_t$, and $\mathbf{f}_t^m$ and $\mathbf{f}_t^p$ are the magnitude and phase of $\mathbf{f}_t$ respectively. These computations were performed for all the music pieces that the system used and were stored on disk for later real-time rendering when needed.

During the runtime portion of the system, we obtain a new timeline from the user that we use to sample the spectral frames $\mathbf{f}_t^m$ and $\mathbf{f}_t^p$. To do so we need to obtain a magnitude and phase spectrum for any new specified time τ. For the magnitude spectrum, if the desired time τ matches an already known time t from our collection of magnitude spectra, then we just use $\mathbf{f}_t^m$. If this is not the case then we linearly interpolate the data between the two closest bounding samples in $\mathbf{f}_t^m$. After obtaining $\mathbf{f}_\tau^m$, we then need to obtain an appropriate $\mathbf{f}_\tau^p$ in order to recover the full DFT frame at time τ. To do so we track the phase differentials from frame to frame and then accumulate them appropriately to synthesize a plausible phase track. Once we obtain $\mathbf{f}_\tau^m$ and $\mathbf{f}_\tau^p$ we can compute $\mathbf{f}_\tau$ by:

$$\mathbf{f}_\tau = \mathbf{f}_\tau^m e^{\sqrt{-1}\mathbf{f}_\tau^p}$$

We can the use the inverse DFT to obtain the appropriate time domain data for that time point. To ensure smooth transitions we use an overlap and add resynthesis process. The time scaling is done independently on both channels of the stereo recording. This helps further reduce the audibility of any phase reconstruction artifacts, which can often appear in implementations of this approach.

The amplitude of the signal is also appropriately modulated to follow the dynamics indicated by the conductor's gestures. Others have used the modified phase vocoder algorithm for similar conducting applications [3]. Our implementation is straightforward, with the only minor difference being that we operate on both channels of a stereo stream independently so as to conceal any potential resynthesis artifacts.

3.4 Video Playback Algorithms

It is widely known in the Video game industry that the public is more forgiving of mistakes in the video stream as long as the audio quality is good. We tried to follow this principle by using the audio processor's clock to drive the video playback. Audio typically plays at a much higher rate than the video (44.1kHz vs. ~60Hz refresh rate), so it is preferable to use the higher frequency audio clock for all purposes in the system.

During playback, the audio player calculates the position of the next audio frame to be played. This position is converted to the universal clip time, from which the next video frame is calculated. This master-slave relation between the audio and video streams guarantees the synchrony of the playback and continuity of the audio stream.

4. Evaluation and Future Work

The *UBS Virtual Maestro* system succeeded in achieving our goals on a number of fronts. First of all, in terms of our design goals, the anecdotal feedback we received from numerous users indicates to us that it was well suited to address the situation and local needs in concert hall lobbies. Secondly, the wide attention and coverage in press and media sources has indicated to us that there is a great deal of public enthusiasm for a system that can provide a satisfying and realistic simulation of orchestral conducting.

In terms of future work, we have identified some areas for improvement in the implementation of the algorithms. First of all, the processing delays from gesture to audio output may be minimized, although they cannot be removed entirely. Secondly, limitations in the accelerometer data from the *Wii Remote* should be addressed, including noisiness and inherent processing delays in beat tracking. Specifically, the sample rate of the Wii Remote (approximately 100Hz) has been demonstrated to be insufficient for capturing the high-frequency components in human gestures [7]. Different sensors and higher sampling rates would allow for better tracking and response to specific conducting gestures, including cues, articulations, and fermatas.

However, we believe that the primary areas to focus on for future work include applications of this system to different musical scenarios. For example, individual

Figure 4. Children participating in the Outreach program at Boston Symphony Orchestra using *Virtual Maestro*

orchestras might wish to customize such a system for their own lobbies, providing valuable marketing not only in their concert halls but in local civic spaces where they could be reaching out to the public in more engaging ways.

We are also deeply interested in the possibilities for using the *Virtual Maestro* in classical music education and outreach programs. In particular, we are interested in the ways in which the Boston Symphony, Philadelphia Orchestra and Minnesota Orchestra have used the *Virtual Maestro* in their programming. In particular, these orchestras created additional enrichment opportunities for the system in Youth Concert programs. At these events, the orchestra's conductor described the *Virtual Maestro* system during the concert, and special events were held in the lobbies to attract children to try out conducting (Figure 4). At the Boston Symphony event in January 2008, the conductor personally came out to the lobby after the concert to spend time engaging with children in impromptu conducting lessons and discussions. At the Cleveland Orchestra event in July 2008, *Virtual Maestro* users entered a drawing to meet the orchestra's Assistant Conductor in a special reception including dinner and discussion.

At Chicago's Ravinia Festival in July 2008, festival President and CEO Welz Kauffman said, "This is such a welcomed exhibit, because the role of the conductor may be one of the most misunderstood in all of performance art. Ostensibly, the most important man on stage comes out and turns his back to the audience the whole evening making magical gestures. This ingenious game gives everyone just a glimpse into what's going on up there." (Chicago Daily Herald, July 9, 2008)

We believe that these kinds of events represent very promising ways to engage younger audiences in classical music events. The *Virtual Maestro* project for us has provided an important platform for exploring numerous issues related to audience development that are becoming increasingly important for organizations that support and present the traditional performing arts.

5. Acknowledgments

The authors would like to thank the officers and board members of Immersion Music, Inc. for the managerial, financial, and logistical support they provided in the development of this project. Also, we would like to thank our collaborators at the marketing agency Jack Morton, Inc., especially Sarah Shulman, Christine Spiel, and Frank Moran, who maintained the system while on tour. Thanks especially to the UBS AG Corporation for initiating, overseeing, and funding this project. We are indebted to Myran Parker-Bass and Andrew Russel of the Boston Symphony Orchestra for their support and encouragement.

References

[1] Borchers, J., Lee, E., Samminger, W. and Mühlhäuser, M. "Personal Orchestra: A real-time audio/video system for interactive conducting." ACM Multimedia Systems Journal Special Issue on Multimedia Software Engineering, 9(5): 458-465, March 2004.

[2] Dolson, M.B. "The Phase Vocoder – a Tutorial." Computer Music Journal, Winter 1986, pp. 14-27.

[3] Lee, E., Nakra, T.M., and Borchers, J. "You're the Conductor: A Realistic Interactive Conducting System for Children." In NIME 2004 International Conference on New Interfaces for Musical Expression, pp. 68-73.

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

[5] Marrin, T. Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton. M.S. Thesis, Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology, 1996.

[6] Mathews, Max V. "The Conductor Program and Mechanical Baton," in Max V. Mathews and John R. Pierce, eds., Current Directions in Computer Music Research. Cambridge, Massachusetts: The M.I.T. Press, 1989. pp. 263-281.

[7] Nakra, T.M. "Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture." Ph.D. Thesis, Media Laboratory. Cambridge, MA, MIT, 2000.

[8] Paradiso, J. "American Innovations in Electronic Musical Instruments." http://www.newmusicbox.org, October 1999.

[9] Roads, C. "The Computer Music Tutorial." MIT Press, February 1996 (page 653).

[10] Wii Conductor Hero, YouTube Video, College of New Jersey, http://www.youtube.com/watch?v=rX-bQR4bkqY