# *MAESTRO:* USING TECHNOLOGY TO IMPROVE KINESTHETIC SKILL LEARNING OF MUSIC CONDUCTORS


## ABSTRACT

The use of technology in music conductor training is a growing area of interest. The expressive, subtle, and meaning-rich gestures that are used in conducting, serve as fruitful ground for innovative research in areas such as artificial vision, gesture following, and musical mapping. While it is known that the kinesthetic skills of conducting are acquired through hours of intensive training, practice with real time audio and visual feedback is severely limited by availability, focus, and good will of live musicians. The current project, titled *Maestro,* builds upon previous work and provides a new approach for training beginning conductors: a system allowing the conductor to practice basic to advanced baton skills accompanied by a virtual orchestra that responds to the conductor's baton gestures affecting tempo, duration, articulation, and dynamics. By incorporating gesture anticipation and tracking, machine learning for gesture analysis, utilization of physical modeling for high-quality audio, *Maestro* provides immediate feedback that is directly related to subtle variations of performed conducting gestures.


## 1. INTRODUCTION

Performing music, whether playing an instrument, singing, or conducting, requires a combination of aural, cognitive, and kinesthetic skills that require specific practice to improve [1], [2]. Such skills could include learning the fingering patterns of major and minor scales on a particular instrument or the weight on the bow of a stringed instrument. Kinesthetic skills are also the foundation of beginning music conducting skills [3]. Beginning conducting students must learn a plethora of movements that include instruction on torso, head, and arm positions and a variety of expressive movements intended to bring about a response from performers.

The acquisition of such skills is a challenging task, which is historically achieved with individual or group instruction, followed by individual practice. Indeed, several technological innovations address this effort by putting an emphasis on the development of kinesthetic skills related to performing music or providing sophisticated feedback (either in real-time or non real-time) to act as a virtual music teacher.

Such tools present different solutions for the practical issues as well as the psychological aspects of acquiring musical skills. Practicing in front of a teacher, peers, and eventually an audience may cause different responses ranging from indifference to anxiety [4], [5]. Creating individualized instructional tools and allowing more comfortable practicing environments can be invaluable to many populations that are affected by such difficulties. We contend that use of the *Maestro* system in such traditional learning environments would enhance the learning experience and encourage kinesthetic awareness and overall musical skill development.

The project seeks to advance previous conducting technology and pedagogy through two core advances: a) the delivery of rich real-time audio and visual feedback through the *Maestro* system to enable the refinement of kinesthetic skills of conducting gestures affecting variations of speed, articulation, dynamic, and speed, and b) the ability to practice conducting gestures without the need for live musicians or peers. The *Maestro* system introduces technical innovation-based research in three main areas: a) gesture anticipation and tracking; b) machine learning for gesture detection and classification; c) utilization of physical modeling for high quality, subtle musical feedback. This work is designed to foster more opportunities for meaningful learning experiences through the beginning conductor's discovery of subtleties of gestures and their effect on musical performance.

## 2. RELATED WOKS

In recent years, there have been several attempts to simulate the conductor's baton. Developments in mobile technology and the wide availability of sensors and accelerometers encouraged researchers to explore the hitherto relatively uncharted realm of conducting. The Radio Baton [6] was one of the first systems developed in this field. It offered an interactive conducting experience by controlling the tempo of a MIDI sequence as a feedback to the gesture. Other systems in later years incorporated sensors for more precise input analysis, such as measuring the pressure on the baton [7], tracking the conductor's muscle tension [8], and using a built-in camera on the baton [9]. Improvement over the years included transition from MIDI to audio-based musical feedback [10] to more sophisticated and realistic forms of sound generations [11].

Similar projects targeted simulation of the conducting experience as a way to experience controlling an orchestra, rather than for researching the subtleties of conducting gestures and their musical effect. In 2004, Borchers offered children the opportunity to conduct the Vienna Symphony Orchestra. The 'conductor' would stand in front of a video screen and control the tempo of an actual performance [12]. Two other systems with similar focus are iSymphony [13] and Pinocchio [14], developed a few years later.

Along with programs designed to familiarize and introduce the conducting experience to non-musicians, some conducting systems have been developed with educational and research goals in mind. A system designed to analyze and classify hand gestures of conductors was implemented on a basis of Hidden Markov Models (HMM) and developed for MAX/MSP [15]. Similar ideas and goals can be seen in *Conga*, a gesture analysis system using graph theory [11], analysis without real-time component [16], video analysis of conductor's gestures [17], and a baton simulation using a Wii remote [18].

Some of the projects introduce complex algorithms and systems for conductors' gesture analysis, and within the constraints and limitations they impose on the gestures, they report high accuracy. Most of these related works focused on the activity of conducting in its highest level – developing a digital system that allows an individual to conduct a short excerpt of or an entire musical composition. These systems focus primarily on the use of movement to control the speed of prerecorded pieces of music, seeming to aim for the education and entertainment of the general public rather than the learning of kinesthetic skills in order to produce effective conducting gestures to indicate a combination of tempo, duration, articulation, and dynamics.

## 3. INNOVATIONS

In all the projects described above, there is a missing component which the *Maestro* system improves upon: previous systems have been developed based on the assumption that the conductor's gestures convey mostly (or only) temporal information; when in practice, a conducting gesture must convey additional aspects of sound generation, such as articulation, volume, and duration.

In order to detect and provide feedback for various aspects of gestures, the *Maestro* system introduces technical innovation based on research in three main areas: a) gesture anticipation and tracking; b) machine learning for gesture analysis; c) utilization of physical modeling for high-quality audio feedback.

### 3.1. Gesture anticipation and tracking

Since any delay that occurs between a performed gesture and its audio feedback is undesirable within a music-conducting system, gesture anticipation, allowing precision of a few milliseconds is an essential requirement. The *Maestro* system uses a high-speed sensing device that provides a data-sampling rate close to 100 Hz in 3D space. Such high-resolution data, combined with pre-trained gestures, allows the anticipation of gestures and achieved accuracy of a few milliseconds.

### 3.2. Machine Learning for Gesture Analysis

Once a gesture is detected, *Maestro's* machine learning algorithm requires two kinds of analyses a) real-time **classification** of a performed gesture by comparing it with a set of pre-trained gestures, and b) real-time **identification** of higher-resolution characteristics of the classified gesture. Both comparisons will be performed with the ultimate goal of mapping any subtle change in a gesture to subtle parameters that will influence the audio feedback.

### 3.3. Physical Modeling

Physical modeling is a set of audio signal processing and synthesis algorithms and models that have been developed based on intensive research of the behavior of acoustic instruments. These models allow the synthesis of realistic-sounding audio with relatively low computational and technological resource cost [19], [20].

The high-resolution sensing and tracking devices, along with the proposed machine learning-based gesture classification, will allow for an intuitive utilization of physical modeling synthesis, where we will map one gesture to multiple parameters of physical modeling-based musical response. Previous conducting projects have used either MIDI [6], [10], [18] or sampled sounds [11], [12], [13] for audio feedback. Physical modeling is another major step towards a realistic conducting environment.

## 4. SYSTEM DESIGN

The system consists of four interconnected modules as illustrated in Figure 1. The modules will include a conductor's baton; a tracking and sensing system; computer software to analyze the gestures, and an interface for audio and video feedback. The baton will serve as the physical interface for the user. Spatial coordinates of performed gestures are sent through an IR transceiver to the desktop application, where they are recorded and analyzed. Once the analysis algorithm recognizes the completion of a gesture, the system generates audio and visual output that correlates to the performed gesture.

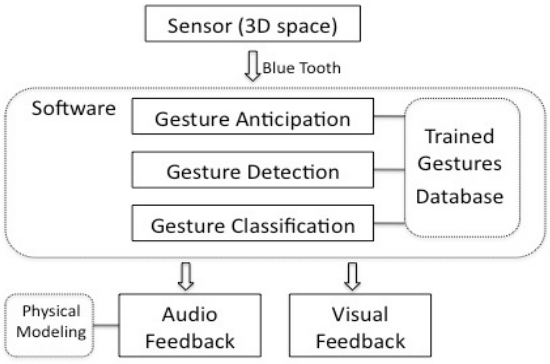


**Figure 1**. Schematic representation of the design.

### 4.1. Conductor's Baton

The baton is a real conductor's baton, fashioned with an infrared LED (Light Emitting Diode) at its tip to allow movement tracking in a 3D space. Infrared sensors were chosen since they track only the movement of infrared light sources, thus avoiding confusion with other objects in space [21]. The baton is wireless to help simulate a real conducting environment.

In addition to the infrared sensor on the baton, higher-level detection (with a lower sampling rate) of the

conductor's head and torso movements are also sensed and allow the detection of skeletal movement in the 3D space. This analysis, combined with the baton movement, allows the rendering of the visual feedback.

### 4.2. Anticipation and Tracking

Once data are fed into the system (raw coordinates of baton movement), 2D representations of the baton movements are reconstructed by the software, and are analyzed in two parallel stages. A gesture detection algorithm distinguishes between random movement, system noise, and intentional gestures. The system then searches for specific characteristics of the conducting gestures (e.g. beginning and end of a gesture). A second algorithm anticipates the end of a gesture (i.e. attack – when the baton movement stops) that allows a time-accurate, audio feedback without discernable time delay. This algorithm gathers information on a gesture before the parallel algorithms determines that the current movement is indeed a gesture.

### 4.3. Gesture Classification

Classification of gestures relies on two orthogonal algorithms, providing two layers of detection accuracy. First, gesture statistics pertaining to the current gesture characteristics (e.g. vertical gesture length, acceleration, attack characteristics) are gathered by the anticipation algorithm, and are compared with gathered statistics of the trained gestures. The second layer is a Hidden Markov Model (HMM) algorithm, commonly used for gesture classification and following, and specifically for conducting gesture classification [22], [15]. The HMM algorithm compares the gesture as a whole once the statistical analysis is complete, and there is a positive match between a performed gesture and a trained one.

The two algorithms complement each other to achieve two goals: anticipate the next gesture to provide audio feedback with no discernable time delay, and prevent false positives for cases in which random baton movements might be mistaken to be real gestures.

### 4.4. Audio and Visual Feedback

Once classification is successful, the musical content is constructed and the recognized gesture is translated to audio and visual feedback.

#### *4.4.1 Audio Feedback*

Parameters gathered from the detection algorithm, along with the classified characteristics of the gesture are mapped to produce a tailored sound, correlating in dynamics, duration, and articulation to the performed gesture. By mapping the rich space of subtle gesture analysis to the rich space of physical modeling sound generation, we are able to provide a sophisticated and intuitive response that would imitate the response of a real orchestra.

The high-resolution sensing and tracking devices, along with the proposed machine learning-based gesture classification, allow for an intuitive utilization of physical modeling synthesis, with which we map one gesture to multiple parameters of a physical modeling-based musical response.

#### *4.4.2 Visual Feedback*

The visual feedback is provided to the user in multiple ways. First, the user is able to see a replication of the path of the baton via the infrared LED at the baton's tip. This path is viewed as a 2D plot that traces the gesture as a whole so that the entire gesture can be viewed from start to finish. Additionally, the interface enables the user to view a mirror image of their torso, arms, and head while performing a gesture in real time. Both of these visualizations provide rich valuable feedback to the user in combination with the audio response.

## 5. CONCLUSIONS

The main achievement of this work is the development of a complete conducting system that allows a conductor to perform gestures and receive multi-dimensional feedback in real-time that matches the musical intent conveyed by the conductor. In particular, several goals were achieved: the anticipation algorithm allows the system to provide audio feedback with time delay of 5 ms from the end of the gesture (attack). The system was pre-trained with 12 different gestures that vary by attack style and dynamic intention, while tempo information was extracted from the gesture in real-time. During tests with the authors conducting, the detection rate (judging if a certain baton movement is a gesture) was 92%, while the classification rate (match between the conductor's intent and the perceived audio feedback) was 81%. The system detects discrete gestures and plays back audio feedback comprised of one instrument, and displays the visual feedback in real time as a mirror image of the gesture.

## 6. FUTURE WORK

The second iteration of the system will include an expansion of the number of sets of trained gestures and melodic excerpts in order to provide a richer learning environment. These will build on the current discrete gestures to successive gestures and multiple meter patterns. Additionally, future work with audio feedback will move beyond a single instrument sound to allow the user the option to hear full orchestra, band, or vocal sounds in response to their gestures. A second iteration will also include a sophisticated, yet intuitive user interface to allow the user to change sound preferences, move between practice modules, visually and audibly record their session, and change camera viewpoints.

The desired end result of this work is to provide a new, meaningful tool to music conducting pedagogy that enhances conductors' development of subtle gestures affecting a full range of musical expression. The *Maestro* system is being developed iteratively and incrementally with input from conductors of various competency levels. An accompanying curriculum is also being developed and will be deployed within the context

of a music conducting class of undergraduate music majors. The system will be disseminated and evaluated in an undergraduate introductory conducting course, evaluated by participating students and the course instructors. Following the analysis of the evaluations, further modifications to the *Maestro* system and collaborative curriculum will be made before another iteration of the study the following year.

Future potential uses of the project include widespread accessibility to conductor training programs and the appropriation of the project for use by individuals at all levels of musical skill and age. System components and techniques that will be developed as part of the project could also be used in medical research such as communicative and movement abilities of disabled persons, sign language technologies for people with visual disabilities, novel gaming interfaces, and music creation software.

## 7. ACKNOWLEDGEMENTS

The authors would like to thank Marcelo Cicconet for his help with the initial setup of the hardware interface, and his continuous help with the implementation of the HMM algorithm.

## 8. REFERENCES

[1] Costa-Giomi, 2005; Does music instruction improve fine motor abilities? *Annals of the New York Academy of Sciences*. 1060, 262-4.

[2] Dickey, 1992; A review of research on modeling in music teaching and learning. 133 (Summer), 27-40.

[3] Haithcock, M., K. Geraldi, and B. Doyle. 2011. *Conducting Textbook*. (self-published).

[4] Kenny, D. 2011. *The Psychology of Music Performance Anxiety*. Oxford University Press.

[5] Wilson, G. D. 1997. Performance anxiety. In *Hargreaves DJ, North AC (eds) The social psychology of music*. Oxford University Press, Oxford, pp 229–245.

[6] Matthews, M. V. 1991. The radio baton and the conductor program, or: Pitch—the most important and least expressive part of music. *Computer Music Journal* 15(4), 37–46.

[7] Marrin, T. 1997. Possibilities for the digital baton as a general-purpose gestural interface. *CHI*, pages 311-312. ACM.

[8] Nakra, T. M. 2000. *Inside the Conductors Jacket: analysis, interpretation and musical synthesis of expressive gesture*. PhD thesis, Massachusetts Institute of Technology.

[9] Murphy, D., T. H. Andersen, and K. Jensen. 2003. Conducting audio files via computer vision. *Proceedings of the Gesture Workshop*, Genova.

[10] Ilmonen, T. 2000. The virtual orchestra performance. *CHI*. ACM.

[11] Grull, I. 2005. *Conga: A conducting gesture analysis framework*. Masters Thesis, University of Ulm.

[12] Borchers, J., E. Lee, and W. Samminger. 2004. Personal orchestra: A real-time audio/video system for interactive conducting. *Multimedia Systems* 9, 458–465.

[13] Lee, E., T. Karrer, H. Kiel. 2006. iSymphony: An Adaptive Interactive Orchestral Conducting System for Digital Audio and Video Streams. In *Proceedings of CHI*, Montreal, Canada, 259-262.

[14] Bruegge, B., C. Teschner, P. Lachenmaier, E. Fenzl, D. Schmidt, and S. Bierbaum. 2007. Pinocchio: conducting a virtual symphony orchestra. In *Proc. ACE 2007*, ACM Press, 294-295.

[15] Kolesnik, P., and M. Wanderley. 2004. Recognition, analysis and performance with expressive conducting gestures. In *Proceedings of the 2004 International Computer Music Conference (ICMC2004)*, Miami, Fl.

[16] Je, H., J. Kim, and D. Kim. 2007. "Hand gesture recognition to understand musical conducting action," in *Proc. of. IEEE International Conference on Robot & Human Interactive Communication*.

[17] Nakra, T. M., A. Salgian, and M. Pfirrmann. 2009. "Musical analysis of conducting gestures using methods from computer vision." *International Computer Music Conference*, Montreal.

[18] Peng, L., and D. Gerhard. 2009. A Wii-based gestural interface for computer-based conducting systems. In *Proceedings of the 2009 Conference on New Interfaces For Musical Expression,* Pittsburgh, PA, USA.

[19] Scavone, G. P. 1997. *An acoustic analysis of single-reed woodwind instruments with an emphasis on design and performance issues and digital waveguide modeling techniques*, Ph.D. thesis, Music Dept., Stanford University.

[20] Smith, J. O. 2004. Virtual acoustic musical instruments: review and update *J. New Music Res. (*33**),** 283–304.

[21] Guy, E. G., F. Malvar-Ruiz, and F. Stoltzfus. 1999. Virtual conducting practice environment. *In Proceedings of the International Computer Music Conference*. ICMA. Pages 371-374.

[22] Usa, S. and Y. Mochida (1998). A conducting recognition system on the model of musicians process. *Journal of Acoustical Society of Japan 19*(4), 275–287.