Gesture Interaction for Electronic Music Performance

Reinhold Behringer

Leeds Metropolitan University, Headingley Campus, Leeds, LS6 3QS, UK
r.behringer@leedsmet.ac.uk

Abstract. This paper describes an approach for a system which analyses an orchestra conductor in real-time, with the purpose of using the extracted information of time pace and expression for an automatic play of a computer-controlled instrument (synthesizer). The system in its final stage will use non-intrusive computer vision methods to track the hands of the conductor. The main challenge is to interpret the motion of the hand/baton/mouse as beats for the timeline. The current implementation uses mouse motion to simulate the movement of the baton. It allows to "conduct" a pre-stored MIDI file of a classical orchestral music work on a PC.

Keywords: Computer music, human-computer interaction, gesture interaction.

1 Introduction

In the minds of many people, the worlds of "classical music" and "computer science" have no intersection. However, there are several areas in which these two fields approach each other. Computers can be used as a tool for automatically generating music, as an aid for the conventional human composition process, or as tools for performance, where their large capabilities in the acoustic processing, e.g. sound synthesis and design are employed. The musical avant-garde has embraced these capabilities and has used computers since the 1950s to generate novel music, for example based on strict mathematical principles for human players (Xenakis [1]) or for directly performing music [2]. Software tools, applications, and frameworks (such as CSound and Max/MSP) have been developed for making these capabilities available to composers. The standardization of communication between synthesizers and computers by the MIDI (Musical Instrument Digital Interface) standard (1983) [3] brought the use of computers in music composition and performance into the mainstream of music production.

1.1 Large Number of Parameters for Music and Sound

The use of computers provides infinite possibilities for the creation of sounds, no longer limited by the physical constraints of conventional acoustic music instruments but solely by the computing power and memory of the computer system in use. They can control electronic synthesizers or act themselves as synthesizers. One of the possibilities is to create "virtual instruments" which are simulations of sound creation mechanisms based on true physical principles. However, to access the vast number of

J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552, pp. 564–572, 2007.

possibilities and parameters for the creative use of synthesizers and virtual instruments is often not intuitive to the player. A musician can learn how to play conventional music instruments and express very subtle nuances of musicality. But using the computer as a musical instrument for expressive live play is in general quite difficult, due to the large number of parameters to be controlled simultaneously and due to the non-standardized interfaces to such computer music systems. As a consequence, "electronic" music generated by computer-controlled synthesizers often lacks the aesthetic quality of human live music play [4], because of an insufficient control of all possible sound creation parameters. This is also true in general for most electronic synthesizers which are mainly controlled by buttons, knows, and sliders. This interface paradigm has been transferred to computer-controlled synthesizers, as the Graphical User Interface (GUI) of music software systems often emulates this synthesizer operation of a one-on-one controller mapping. An exception is the Theremin, which operates on the principle that the position of the player's body relative to an antenna is directly translated into the creation of a sound, based on analog electronics principles. This allows the player to directly interact with the generated sound in an intuitive way, enabling the player to shape the sound with intuitive musicality. There have been many adaptations of this technique into software interfaces for computer simulations of the Theremin principle (e.g. [5]), simulating the effect of the interaction for the sound creation [6]. However, the acoustic possibilities of the Theremin are limited to monophonic music (unless several Theremins would be combined) and to the unique continuous pitch change of this instrument.

1.2 Sampled Instrumental Sounds

These shortcomings of synthesizers and computer interfaces have prevented the wide-spread use of computer technologies in the performance of traditional non-avant-garde classical music. A step forward has been the introduction of sampled sounds, based on recordings of acoustic instruments. This of course limits the freedom of the sound creation process to the production of "naturalistic" instrument sounds. But this removes from the player the burden of creating a musically pleasing sound, as the recorded instruments naturally have incorporated centuries of musical heritage and experience in them. Examples of such sample libraries of orchestral instruments are the Vienna Symphonic Library [7] or the Garritan Orchestra Libraries [8]. The number of parameters for the sound generation is significantly reduced in these libraries, because the various playing techniques have been recorded as separate sample sets. To change from one sample to another can be easily mapped to a single keystroke and hence can be executed very rapidly. Shaping the sound can be achieved with very few parameters, as the main sound characteristics are in the pre-recorded samples. This can be used for a real-time performance given by a human player of a particular sampled instrument.

1.3 Music Timeline

With these sampling techniques of reproducing the sound of acoustic instruments, it is possible to create "natural" and realistic sounding renditions of classical orchestral

works. This broadens the application of computers in the musical context away from music types which made specific use of the inherent sound characteristics of synthesized sounds and computerized rendition. However, there is still one big hurdle in the creation of renditions of traditional classical music: to create a "musical" aesthetically pleasing time flow of the music rendition. A music performance by a human instrumentalist has often subtle deliberate tempo variations, which are intuitively generated by the player. If the player is in a live performance and plays the instrument, then this timeline is generated naturally. However, if a music part is created "offline" as a rendition by programming a sequence of music events (notes, controller chances, etc.) using a (software) sequencer, then the programmer has to create this timeline manually, either by editing tempo variations or deviations from the musical beat.

In this paper we address the issues of creating such an offline music timeline intuitively, using gesture interaction, as used by orchestra conductors.

2 Related Work

Since the mid 1980s, much research and development has been done in the area of interaction with music instruments, in order to make the capabilities of computer music available for the music creation and performance process.

Already in the 1960s, Max Mathews had experimented with a light-pen [9] for using graphical interaction as a composition tool. He later developed the radio baton [10] for allowing improvising and conducting music, based on 3D tracking of radio frequency tracking emitted from the baton. Other early conducting systems have been developed in the 1980s [11][12].

Teresa Marrin Nakra's "Conductor Jacket" [13][16] is a complete system worn by an orchestra conductor for interacting with a computer system. The "Digital Baton" in this system was also used in MIT's "Brain Opera" [14]. It is comprised of 3 sensor systems: an IR LED at the baton's tip, an array of 5 force sensitive resistor strips at the grip handle, and 3 orthogonal 5G accelerometers [15]. The visual tracking was done through the IR LED which emits modulated (20 kHz) light, detected by a 2D position-sensitive photodiode. Also magnetic tracking can be used as a simple interface for orchestra conducting, as demonstrated by Schertenleib et al. [17]. An infrared baton that is actively emitting IR for tracking by triangulation has been used in conducting demonstrations such as the "Personal Orchestra" by Borchers et al. [18]. Common to these systems is that they are not completely seamless and require the conductor to either wear devices and sensors, use specialized batons, or move in a specific location relative to a receiving device.

A large area of research is how sensor data from those interfaces can be mapped into acoustic or musical parameters (e.g. [19]). Research has shown that more complex mapping schemes which allow simultaneously controlling several parameters with a reduced number of control input, allow the user to do more complex tasks [20]. This cross-coupling of several parameters actually emulates real acoustic instruments, in which such a cross coupling frequently occurs: the input controls often act on several sound parameters simultaneously.

3 Concept

The idea of our approach of an interface between a musician / conductor and an electronic computer-controlled synthesizer is to leverage from the repertoire of gestures that have been employed in the interaction of the conductor with a traditional orchestra. This will allow that the timeline for a synthesizer rendition can be created naturally, enabling the musician to create an aesthetically pleasing musical performance recording. There is a set of pre-defined gestures which the orchestra conductor executes in order to synchronize orchestra players and to tell the orchestra members how to play their music part [17]. The information conveyed by the conductor is basically tempo (beat) and "expression". The latter one is a complex set of parameters, mostly consisting of instructions on loudness and phrasing. Some of this is conveyed not only by hand motion but also by facial expressions, using the whole range of human-human interaction for conveying the musical intent of the conductor to the orchestra members.

3.1 Elements of Conducting an Orchestra

The goal of this research project is not to exploit the facial interaction (at this time) of the conductor with the orchestra, but rather to focus solely on the gestural interaction. In this section we briefly will revisit a few essentials of traditional conducting [21]. The main parameter conveyed through the hand motion is the beat of the music tempo. A long vertical motion from top to bottom indicates the first beat in a measure. Vertical hand motion indicates the pace of the music to be played, and the beat itself (rhythmic pulse, ictus) is by convention on the lower bounce of the hand. In other words: the beat is characterized by the sudden change in direction of the hand motion at the bottom of the motion.

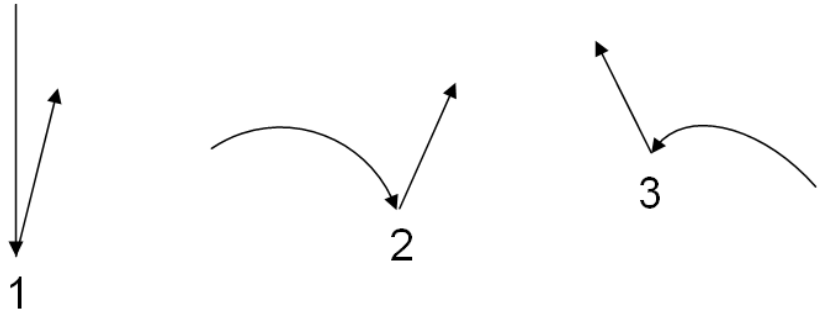

Fig. 1. Motion of the conductor's hand of a THREE beat: the numbers indicate the beat within the measure. The ictus is at the deflection of the hand motion (change of direction from down to up = bouncing).

In Fig. 1, the standard figure for conducting a THREE beat is shown [21]. The initial motion is from the top downward. The first beat is at the bounce of the motion at the bottom. The hand then raises up, only to move down again to create the 2nd

bounce (beat #2). The third beat is again a bottom bounce, but higher up. From there, the hand returns back to the upper default location, for indicating the next bar.

3.2 Mapping of Conductor Hand Motion

Based on the conductors' practice, the following seems a reasonable approach to map the hand gestures to recognition:

- The longest vertical down motion indicates the beginning of a bar.
- The beats are the time when the inversion of motion direction between downward and upward occurs (bounce at the lower end).
- The amplitude of the hand motion – both in vertical and horizontal direction – is mapped to the overall volume of the music.
- In a more sophisticated version, the "hemisphere" of the predominant hand position can indicate the parts of the orchestra to which the conducting gestures are addressed. This is done by conductors to highlight individual instrument groups and could be used in the same way for an automatic separation of those groups of synthesizer instruments.

The tempo can be computed from the time duration between two beat points. This will give only a coarsely quantized tempo map, as a new tempo would only be determined at those ictus inflection points. However, in a real performance the tempo can change between the beats, indicated by a slowing hand motion. In order to bridge these gaps between beats which can be in the order of 200 ms to 2 seconds [4], it is necessary to continuously monitor the hand motion and to derive a suitable mapping onto the resulting tempo.

4 System Architecture

The development of the envisioned system is done in modules. The system will analyze the motion of the conductor's hand and play an electronically stored music file (in MIDI format), with the tempo and expression controlled by the conductor.

The visual tracking module is observing the two hands of the conductor in a video stream, obtained from a connected camera. Based on detection of hand and face (using color cues and expected relation between face and hands), the tracking is initialized. The system outputs the values of the 2D location of the hand in the image synchronized to the video framerate. In the current implementation, this module is simulated by mouse motion over a given screen area.

In order to recreate the hand position at any arbitrary time and to increase the "resolution" to the required smallest musical time (5ms), a spline algorithm creates a smooth representation of the motion. This allows to interpolate the location and the time of the beats at a higher resolution than the video frame sampling.

An analysis is done to detect the beginning of a bar (longest vertical stretch of motion) and the beat (velocity zero-transition at lower end). These values lead to the tempo which is sent to the synthesizer module.

An optional module for acoustic tracking can be connected, to provide feedback from the sound/audio of other ensemble players [22][23]. Currently, such a module is

not being integrated, but the architecture provides input for this. This would basically be a score follower, comparing the captured audio with the expected audio.

In order to compensate for lag and latency (processing time), it is necessary to extrapolate from the current measurements to the current time.

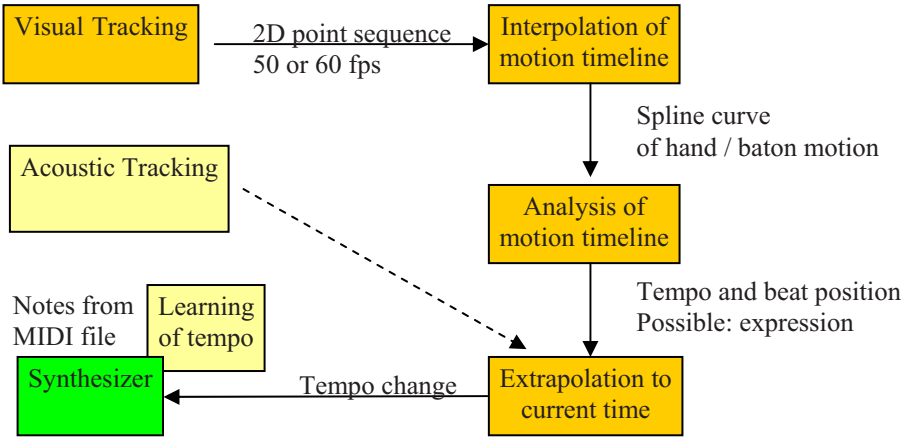

Fig. 2. Architecture for conduction a MIDI file

Since the MIDI event storing is usually done in musical time, that is in measures and bars, there needs to be a translation of the beat interval time (real time) into musical time – this happens through the tempo. Based on the beat timing, this tempo is computed and sent to the synthesizer. The synthesizer is playing a pre-stored MIDI file ("score") and uses the tempo to adapt the replay speed. From the conducting amplitude, the volume can be derived and sent to the synthesizer.

Since the tempo in a classical music piece can vary quite significantly, it is difficult to predict the tempo with the lag compensation correctly: in most cases it will be right, however in cases of a sudden tempo change, the prediction / extrapolation of the tempo will be not correct, if it is plainly extrapolated from past tempo chances. Therefore, it is necessary to embed the expected tempo changes into the MIDI file. In practice, at the first run the MIDI file does have a "standard" default tempo. As the music piece is played for the first time, the tempo captured from the conducting is stored and placed into the MIDI file as a reference tempo. At the next runs, this tempo map allows a more correct prediction of the tempo variations.

5 Implementation

The system is being implemented on Windows XP, using the Windows32 APIs and DirectX for replay of a MIDI file. The software modules have been written in C++ for fast real-time operation.

Since it is important to use a precise timing, the Windows high precision timer was used instead of the multimedia timer.

There are two types of tempo changes which are computed as a consequence of the beat detection: the "true" tempo change, based on the detected beat. These tempo changes are stored and synchronized to the MIDI file timer, so that a new replay of the MIDI file (without conducting) will result in the timing of the conducted timeline. In order to reproduce the live replay as it is being conducted, another tempo needs to be computed: as there is a lag between the conducting control output and the actual MIDI file play, the MIDI file is usually "ahead" of the conductor analysis. In order to compensate and to slow down or speed up the replay, so that it is again synchronized to the conductor, requires this additional set of live tempo changes. These need to compensate the error which had been created by the processing lag, and allow that the sounding music during the conducting is in sync with the conductor's motion.

6 Results

To obtain realistic test data for development, we have recorded two orchestra conductors during rehearsal of a variety of classical music works – the overall video data covers about 15 hours on MiniDV tape. The camera was mounted on a tripod, behind the orchestra, facing towards the conductor above the heads of the orchestra members. This viewpoint puts the camera in place of a regular orchestra member. The zoom lens of the camera was set so that the arms of the conductor fill the image in extreme moments.

This data collection has partially been transferred to a hard disk and reduced to 320x240 pixels – this resolution is deemed to be a good compromise between precision requirement and computation time economy.

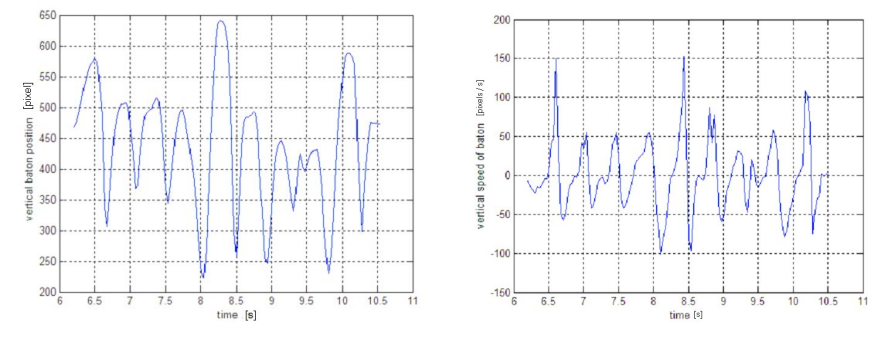

Fig. 3. Left: Vertical baton/hand motion. Right: vertical speed of the hand/baton.

From previous experiments [24], we have obtained the following typical hand / baton motion patter shown in Fig. 3 (solely the vertical motion component). It clearly can be seen how the bar detection and the beat detection can be done by analyzing these motion patterns.

We have implemented the replay of standard orchestral MIDI files with complex structure and many control parameter events. One example is Johann Strauss' waltz

"At the Blue Danube". We hope to have the system ready for HCI 2007in order to show a demonstration of the capabilities.

7 Conclusions and Outlook

There is still a lot to do until a system for capturing the conductor's gestures through computer vision can be used reliably in a situation where an electronic instrument is supposed to play in an orchestral example together with human players. At the time of the submission of this paper, our prototype was not yet in a state for us to make an assessment of the suitability for such an envisioned application. However, such a system is feasible, and the current state of technology is mature enough to envision that such a system can be built.

There are many possible applications of such a system:

- Professional musicians (soloists) will be enabled to rehearse their part of a performance at home, using an automatic computer-controlled accompaniment system, in a "music-minus-one" fashion, but with an automatically adaptable timeline of the accompanying other instruments, fitting to the interpretation of the soloist.
- Music could be written specifically for a piece for "orchestra and synthesizer" where the synthesizer part would be played by a computer. This would raise the computer into the status of an individual orchestra member, playing an electronic instrument according to the instructions of the conductor.
- This system could also be used in education and training, to teach facts about music, interpretation, and performance.
- Such a system could be developed into an application giving "hobby home conductors" the ability to give individual performances within their home.

References

1. Xenakis, I.: Musiques Formelles (1963), reprinted Paris, Stock (1981)
2. Doornbusch, P.: The Music of CSIRAC, Australia's first computer music. Common Ground (2005)
3. Heckroth, J.: Tutorial on MIDI and Music Synthesis. MIDI Manufacturers Association Inc. (2001) (accessed 12 February, 2007) http://www.midi.org/about-midi/tutorial/tutor.shtml
4. Dannenberg, R.B.: Music Understanding by Computer. 1987/88 Computer Science Research Review, Carnegie Mellon School of Computer science, pp. 19–28 (1988)
5. Bolas, M., Stone, P.: Virtual mutant theremin. In: Int. Computer Music Conference, San Jose, CA, USA, pp. 360–361 (1992)
6. Theremin World: (accessed 14 February, 2007) http://www.thereminworld.com/software.asp
7. Vienna Symphonic Library: (accessed 15 February, 2007) http://www.vsl.co
8. Garritan Orchestral Libraries: (accessed 15 February, 2007) http://www.garritan.com
9. Mathews, M.V., Rosler, L.: Graphical Language for the Scores of Computer-Generated Sound. In: Perspectives of New Music, pp. 92–118 (1968)
10. Boulanger, R.C., Mathews, M.V.: The 1997 Mathews radio-baton and improvisation modes. In: Proc. of ICMC 1997, Thessaloniki (1997)

11. Brill, L.M.: A Microcomputer based conducting system. Computer Music Journal 4(1), 8–21 (1980)
12. Haflich, F., Burnds, M.: Following a Conductor: the Engineering of an Input Device. In: Proc. of the 1983 International Computer Music Conference, San Francisco (1983)
13. Marrin Nakra, T.: Immersion Music: A Progress Report. 2003. In: Int. Conf. On New Interfaces for Musical Expression (NIME), May 22-24, Montreal (2003)
14. Paradiso, J.A.: The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance. Journal of New Music Research 28(2), 130–149 (1998)
15. Marrin Nakra, T., Paradiso, J.A.: The Digital Baton: a Versatile Performance Instrument. In: Proc. of the 1997 International Computer Music Conference, San Francisco, pp. 313–316 (1997)
16. Marrin, T.A.: Towards an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton. MSc thesis, MIT, Boston (1996)
17. Schertenlaib, S., Gutierrez, M., Vexo, V., Thalmann, D.: Conducting a virtual orchestra. IEEE Multimedia 11(3), 40–49 (2004)
18. Borchers, J.O., Samminger, W., Mühlhäuser, M.: Personal Orchestra: conducting audio/video music recordings. In: Proc. of Wedelmusic, Darmstadt (2002)
19. Rovan, J.B., Wanderley, M.M., Dubnov, S., Depalle, P.: Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. In: Wanderley, M., Battier, M. (eds.) Trends in Gestural Control of Music, Ircam, Paris (2000)
20. Hunt, A., Wanderley, M.M., Paradis, M.: The importance of parameter mapping in electronic instrument design. In: Proc. of the 2002 Conference on New Instruments for Musical Expression (NIME), Dublin, Ireland (2002)
21. Green, E.A.H.: The Modern Conductor, 6th edn. Prentice Hall, Upper Saddle River, New Jersey (1997)
22. Vercoe, B.: Teaching your computer how to play by ear. In: Proc. of 3rd Symposium on Arts and Technology (1991)
23. Dannenberg, R.B., Grubb, L.: Automated accompaniment of musical ensembles. In: Proc. of the 12th National Conference on Artificial Intelligence, AAAI, pp. 94–99 (1994)
24. Behringer, R.: Conducting Digitally Stored Music by Computer Vision Tracking. In: 1st Int. Conf. on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS), Florence (2005)