Conductor Following With Artificial Neural Networks

Tommi Ilmonen
Helsinki University of Technology
Telecommunications Software and
Multimedia Laboratory
P.O.Box 5400
FIN-02015 HUT
+358-9-451 4735
Tommi.Ilmonen@hut.fi

Tapio Takala
Helsinki University of Technology
Telecommunications Software and
Multimedia Laboratory
P.O.Box 5400
FIN-02015 HUT
+358-9-451 3222
Tapio.Takala@hut.fi

Abstract

This paper presents techniques based on artificial neural networks to extract rhythm and nuance information from conductor's movements. Neural networks are combined with heuristics to achieve a natural reaction by a virtual musician. Main emphasis has been on tracking the rhythm implied by conductor's movements. We have approached the conductor following as a modular gesture recognition and musical expression synthesis task. The process of conductor following is divided into three tasks: gesture identification, gesture meaning analysis and musical expression synthesis.

1 Introduction

Automatic conductor following is a way to create more flexible performances with computer-controlled instruments. A follower software runs in a computer and reacts to conductor's movements by playing music – like a real musician. Such technology can be useful in conductor training and live performance.

The earliest systems were made in the eighties by Max Mathews (1991). Successive projects have refined the methods to use artificial neural networks (ANNs) and Hidden Markov models for motion analysis along with heuristics to provide better following (Brecht & Garnett, 1995; Usa & Yasunori, 1998). Tobey reports a comprehensive system, but little information on it is available (1996).

Our research was ignited by desire to understand and model implicit non-verbal communication between humans. The conducting presents a challenging case of communication where no words are used. Due to our research interests we have used ANNs extensively and combined them with heuristic rules. The system acts as virtual musician, the conductor can conduct any piece represented as a MIDI file (figure 1). As a novelty the system may be combined with animated virtual musicians who play the piece. Thorough information on the system can be found in author's master's thesis (Ilmonen, 1999).

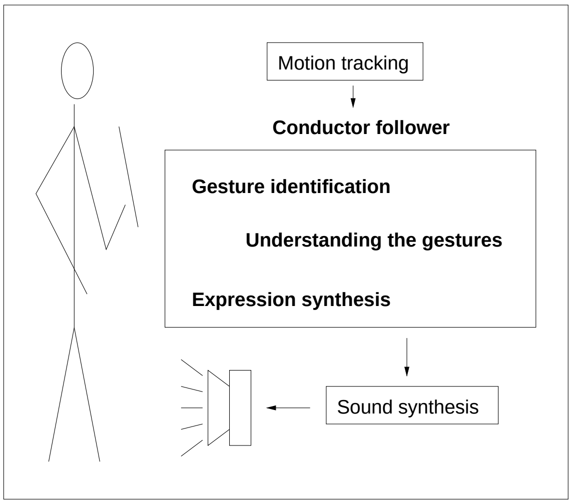

Figure 1: A conductor follower in its environment.

The follower has appeared as part of the DIVA setup in SIGGRAPH'97 conference (1997) and at Finnish Science Center Heureka. The virtual DIVA band has also appeared with a live band in concert.

2 Identify, Understand and Express

Any conductor follower must accomplish three tasks: Identify the gestures the conductor makes, understand their meaning and produce musical output in real-time (figure 1). These tasks can be combined to a single algorithm. Mathews's (1991) system is an example of such approach. Recently most projects have adopted a modular approach; for example we and Usa (1998) have taken this approach.

We have grouped the modules to three layers that correspond to the three tasks of the follower.

The modules interact strongly by sharing information. Thus the system combines modular structure with holistic reasoning.

The first layer contains modules to detect gesture primitives such as beat type, dynamics and staccato. Some gestures can easily be tracked by direct calculation. Other gestures are more difficult because they have different manifestations depending on the conductor. Since neural networks are inherently adaptable to variations in their environment and the networks can be trained anew for each conductor they are an attractive choice for gesture analysis. A multilayer perceptron (MLP) ANN is used to track continuous tempo, a self-organizing map (SOM) (Kohonen, 1995) to classify different beat types and yet another SOM to predict the type of an upcoming beat.

The second layer combines the results of the gesture detectors to form an understanding of the conductor's intentions. This includes coherent, reliable tempo estimation and other performance parameters – dynamics, nuancing, articulation etc. A priori knowledge of the piece (expected dynamics and beat patterns) is used to make sure the gesture data is interpreted correctly. Rigid rules and heuristic pattern matching methods estimate the intended score position and nuancing.

The third layer caontains heuristics that control sound synthesis via MIDI. The event stream from a source MIDI file is transformed to individual notes and enriched with musical hints such as fermata and staccato markers. Information provided by the lower layers is used to create musically valid timing and dynamics for the enriched score representation.

3 The Gestures

The conductor communicates his intentions to the orchestra with gestures. He/she can control any musical parameter – tempo, dynamics, nuances etc.. The follower detects the most important gestures – gestures that affect tempo (beat phase, beat type) and some dynamics and nuancing gestures (staccato and legato).

Basic knowledge of the gestures has been obtained from books (eg. McElheran, 1989) and by discussion with conductors. The algorithms used in the software are derived from this knowledge. For example the tempo estimation is based vertical and horizontal movement of the baton tip as these indicate the beats and beat pattern. Often the vertical motion alone is shows the tempo. In these cases the lowest point of the physical beat matches the musical beat. Measurements have given new knowledge of the motion. Figure 2 shows movement that contradicted some of our experts' views of how the baton moves – the physically lowest point of beat does not necessarily indicate musical down-beat exactly.

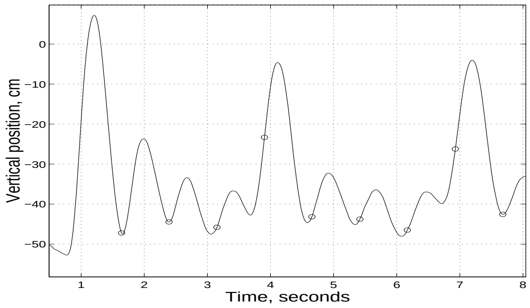

Figure 2: Measured baton tip vertical movement. Little circles represent even division of beats for two bars. The first and last point have been selected manually and the others are interpolated evenly.

When applying the analysis results one needs to be careful. It has been pointed to us that a conductor may use staccato technique while conducting non-staccato passages. As a counter-measure the effect of certain gestures can be disabled.

4 Motion Tracking

Tracked motion is the basis for motion analysis. The more accurate and thorough the measurement the more one can extract information from it. Figure 3 shows an example of how the follower sees the conductor. We have limited the motion measurement to the features a musician can see from anywhere in the orchestra. This includes body and hand movements. Eye movement and other less-visible forms of behaviour are not taken into account, unlike Usa (1998). Our system uses a magnetic motion tracker to track six degrees of freedom per sensor. Figure 4 shows such sensors attached to the conductor. Accelerometer sensors that track two degrees of freedom have been used in a demo application where low price is important.

To get motion data we asked conductors to wear the data suit and to conduct a short passage of music with a given tempo and nuancing. A metronome was used to give tempo and get reference timing.

5 Tempo

Tempo is shown with the motion of the right hand (given a right-handed conductor). The motion repeats a cyclic waveform. The conductor's motion is divided into beats. Each beat indicates advance in the score position.

Multiple modules participate in determining how the music should be played. One module estimates beat phase; eg. the timing within one beat (figure 5). The beat phase can be estimated at all times.

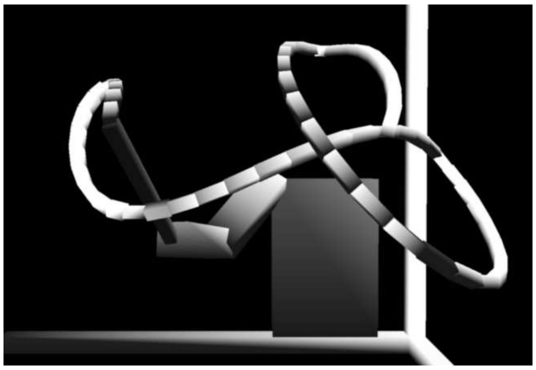

Figure 3: A 3D visualization snapshot. The curve represents baton tip trajectory.

Figure 4: The magnetic sensors for motion tracking.

Therefore one can react to future beats even before they take place.

We have used ANNs to determine the beat phase – like Brecht (1995). As ANNs we have used MLP variations and self-organizing maps. At each motion sample the phase of the motion is estimated. The ANN input vector contains current baton tip position, current velocity and delayed velocities. Both vertical and horizontal dimension are used. A single network is capable of tracking tempos between 40 and 192 bpm – exceeding the limits of practical conducting.

The system follows beat patterns by classifying individual beats and matching successive beats to expected beat patterns. To determine the beat type (first, second etc.) motion of the two latest beats is collected to a single parameter vector. SOM is used to classify the the beats.

Each piece has an expected beat pattern. Since the conductor may add extra beats (for example by subdividing a beat) or leave beats out without a warning the system needs to be very robust. One module matches estimated beat types (measured in real-time) to expected beat pattern (supplied by user in advance).

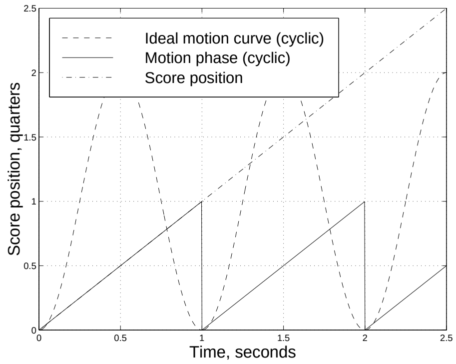

Figure 5: The mapping from vertical motion to beat phase and score position.

The possible beat timings are presented as beat classification vectors with a priori expectation. The system tries to match the new beat classification to expectation of the beats. Few latest beats (three to five) are matched against the expected beats to find the best interpretation. The best match implies the current score position with accuracy of one beat.

The beat pattern and beat phase calculations estimate the current score position and tempo. The final module filters out noise caused by overly accurate beat phase output and combines the phase data with the score position data. Since the ANNs react quickly to changes, irregularity of conducting motion causes notable changes in estimated tempo. Heuristic rules are used to obtain more even tempo. The same heuristic rules also handle special situations like the beginning of the piece, the end of the piece, and fermatas.

The tempo must react swiftly when the conductor wants it. At other times the system must not react fast to changes in shown tempo. Figure 6 illustrates the problem. Most of the time the heuristics try to keep the tempo as stable as possible. As major changes in conducted tempo are detected the heuristics respond faster to conductor's movements.

6 Beyond Tempo

Showing tempo is only part of conductor's work. Nuances, dynamics and emotions are also expressed with gestures. These gestures are commonly embedded in the movement – there is no specific hand movement for a particular emotion. One must determine the implied parameters from the form of the motion.

For example staccato and legato are easily detected with heuristic algorithms. Monitoring the difference between short term maximum and median acceleration indicates fairly well legato and staccato: high peak acceleration indicates staccato, while low overall acceleration indicates legato. De-

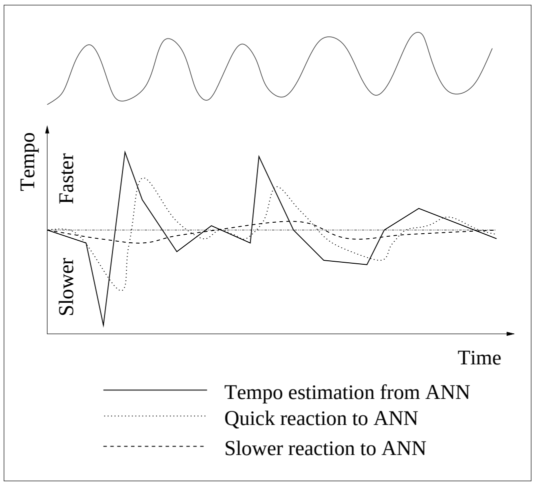

Figure 6: Three alternate ways to react to estimated tempo changes. Note that the tempo is intended to be completely stable.

tecting emotions is more difficult. There is little knowledge of how exactly emotions are manifested in movements. We believe that finding explicit rules that could detect emotion may be very difficult. We plan to use ANNs for this purpose.

Certain special gestures are detected with the aid of a priori knowledge. Such gestures are the beginning and the end of the piece. Both the beginning and the end are typically shown with explicit motion that is easily detected. In these cases algorithms to detect the gestures can be derived directly from expert knowledge.

7 Discussion

The follower has proved to work well when conducted carefully. The system has implicit limitations. The conductor must use standard conducting technique if the machine is to understand his movements. For the best performance, the ANNs should be trained anew for each conductor. With properly tuned ANNs the follower can react faster than a musician can, but its reaction is not as musical as the interpretation of a real musician.

New analysis methods are needed to detect emotional content of the motion. The musicality of the system needs tuning, especially producing emotional features. A player model could solve this problem. While a comprehensive player model is lacking algorithms have been already published by Bresin and Friberg (1998). Currently the system creates only one interpretation of the music. Multiple player models need to be bundled to mimic the interplay between musicians and to the give feeling of a real orchestra.

8 Acknowledgments

This work was funded by the Academy of Finland, Tacit knowledge in complex mind-environment systems project. This work has also been supported by Finnish Technology Development Centre Tekes, the Finnish Science Centre Heureka, Veikkaus and VTI Hamlin.

References

Brecht, B., & Garnett, G. 1995. Conductor Follower. *Pages 185–186 of: Proceedings of the International Computer Music Conference.* San Francisco: Computer Music Asociation.

Bresin, R., & Friberg, A. 1998. Emotional expression in music performance: synthesis and decoding. *Pages 85–94 of: KTH, Speech, Music and Hearing, Quarterly Report*, vol. 4. Stockholm, Sweden: Kungl Tekniska Högskolan - Royal institute of technology.

Hiipakka, J., Hänninen, R., Ilmonen, T., Napari, H., Lokki, T., Savioja, L., Huopaniemi, H., Karjalainen, M., Tolonen, T., Välimäki, V., Välimäki, S., & Takala, T. 1997. Virtual Orchestra Performance. *Page 81 of: Visual Proceedings of SIGGRAPH'97.* Los Angeles: ACM SIGGRAPH.

Ilmonen, T. 1999 (April). *Tracking Conductor of an Orchestra Using Artificial Neural Networks.* M.Sc.Tech. thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory.

Kohonen, Teuvo. 1995. *Self-Organizing Maps.* Berlin: Springer.

Mathews, M. 1991. The Radio Baton and Conductor Program, or: Pitch, the Most Important and Least Expressive Part of Music. *Computer Music Journal*, **19**(4).

McElheran, B. 1989. *Conducting Technique for Beginners and Professionals.* Oxford/New York: Oxford University Press.

Tobey, F., & Fujinaga, I. 1996. Extraction of Conducting Gestures in 3D Space. *Pages 305–307 of: Proceedings of the International Computer Music Conference.* San Francisco: Computer Music Asociation.

Usa, S., & Yasunori, M. 1998 (October). A Multimodal conducting simulator. *Pages 25–32 of: Proceedings of the International Computer Music Conference.* International Computer Music Association, Ann Arbor, Michigan USA.