Tracking conductor of an orchestra using artificial neural networks

Tommi Ilmonen
Tommi.Ilmonen@tcm.hut.fi
Laboratory of Telecommunications Software and Multimedia
Helsinki University of Technology

Abstract

Live performance using a combination of musicians and electronics has so far been problematic art. Getting the two components to create a coherent sound is a major problem. Automatic conductor following is one way to do this. In this case, the conductor can control the electronics as one controls the musicians. Conductor following requires multiple information technology tools. Pattern recognition techniques are used to recognize the gestures (movements) conductor makes. The identified gestures need to be matched with general knowledge of the conducting technique and the knowledge of the piece being conducted. Finally the conductor follower must create musical, expressive output. Little research has been done on this field. Most systems only track the tempo of the piece. Few can track dynamics or finer nuances like staccato. Our main research goal is interpreting the tempo of the conductor and reacting musically in real-time. A multi-layer perceptron network is used to estimate the phase of the conductor motion. This enables very accurate tracking with fast reaction. An oscillator is used to track slow tempo changes when accurate tracking is not necessary. Applications of this technology include conductor training, live performance and music synthesis control.

1 Introduction

Conductor following is a challenging problem of modeling human-human interaction using a computer. The features of conductor following as a research topic are: natural gesture recognition, understanding these gestures, modeling a human reacting to these gestures and information processing in temporal environment.

Only a handful of reports have been made on conductor following. The pioneering work was done by Mathews [1], whose "Radio baton" is a simple conductor following implementation. Later projects have improved the techniques. Among these are Keane [2], Carosi [3], Brecht [4] and Tobey [5]. The latest systems use ANN-based (artificial neural network) beat tracking with additional heuristics to prevent ANN errors from disturbing the following.

Our work brings to the field new ANN solutions with improved motion tracking mechanism.

1.1 The aim

The conductor follower should resemble orchestra (or musician) as much as possible.

The tasks of the virtual musician can be divided conceptually into three tasks. The gestures of the conductor need to be identified. Being natural gestures they can be difficult to analyze computationally.

The identified gestures need to be matched to the knowledge of the piece. One can determine how a given gesture should be understood in a particular context and how the simulated musician should react to it.

Finally one must produce musical output in real-time. This requires modeling of a musician playing a piece. In most projects this player model is either implicit or based on heuristics.

2 Gestures

The most visible form of communication between conductor and orchestra are the gestures conductor makes. The most audible is the music. In training situation both conductor and band members may speak.

We have narrowed the study of communication from conductor to orchestra to the gestures.

The gestures are meant to direct the musicians, tell them the tempo and mood of the piece. The most crucial task in performance is showing tempo - making sure all players are playing the same music in the same time.

2.1 Tempo

Tempo is usually shown with the right hand (given right-handed conductor) using predefined beat patterns (figs. 1 and 2). The time one physical beat takes is ideally the time between two local minima in vertical baton tip (or hand) motion. The vertical movement should show the tempo and the exact phase within beat (fig. 3). The combination of vertical and horizontal movement indicates which part of bar is currently played.

The strength and regularity of vertical motion suggests it could be used as the basis for rhythm tracking.

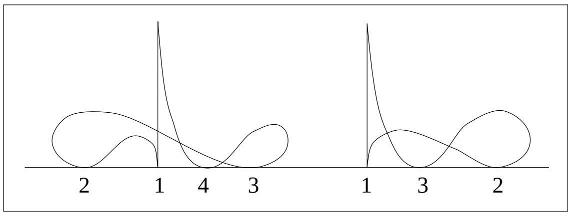

Figure 1: Beat patterns for time signatures 4/4 and 3/4. View is that of the conductor.

Rhythm can be controlled with the horizontal movement also. This is not a recommended way to conduct music, but often used [6]. Most conductor following systems cannot combine vertical and horizontal modes successfully.

2.2 Other gestures

Besides tempo, conductor can control many expression parameters. These gestures include - for example - staccato (shortened note duration), legato (notes glued together) and dynamics (louder or softer).

Some of these can be expressed with explicit gestures. Often they are embedded in other gestures. Typically the conductor shows a number of expression parameters

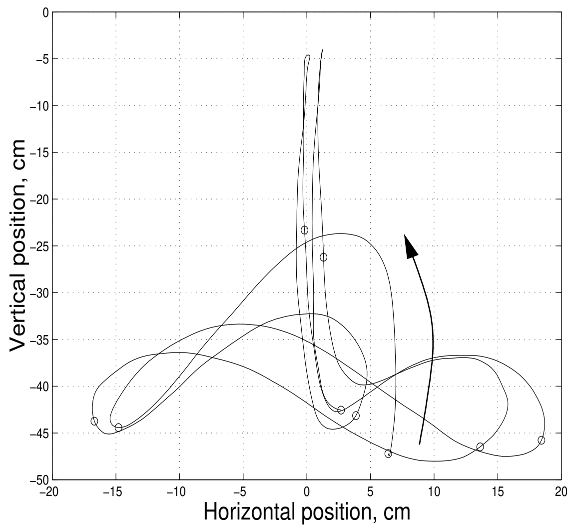

Figure 2: Real baton tip movement. Two bars of 4/4 by conductor Hannu Norjanen. Conducted very carefully. The little circles represent the point of curve where musical beats are supposed to be, see figure 3

with baton. The form of motion implies how the music should be interpreted.

Not all gestures should be understood literally. Often conductors make dramatic movements in concert situation for the audience, not for the orchestra.

In our system the dynamics of the music can be controlled. Heuristics are used to detect left hand's "softer" and "louder" gestures.

2.3 Motion tracking

The first link in information processing chain is motion capture. Motion tracking needs to capture all significant movements with sufficient accuracy.

Often players cannot see the conductor very well. There are other musicians blocking the view or they just do not have perfect sight. Therefore conductors should express their meaning using highly visible motion

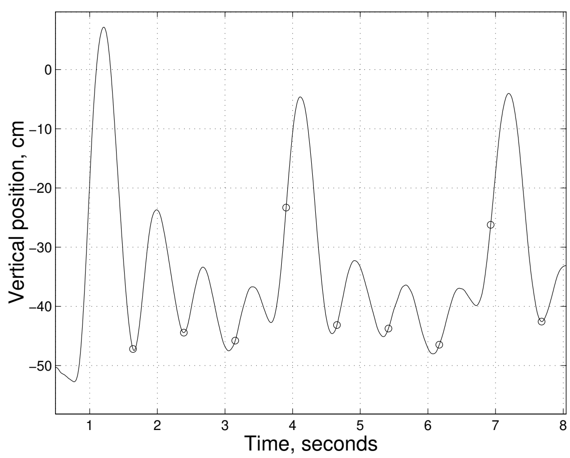

Figure 3: Measured baton tip movement. Little circles represent completely even division of rhythm for two bars. The first and last point have been selected manually and the rest are interpolated directly.

[6]. This implies that hand, baton and body movements should contain most of the expressive information.

We have used magnetic tracker for motion tracking. The tracking equipment consists of a magnetic pulse generator and multiple small sensors. This is demonstrated in figure 4. The most important sensors are in conductors hands. For reference point (origin) we use a sensor behind conductor's neck. This setup produces a rich set of data to be analyzed. Most other projects have only captured baton tip motion with either camera or accelerometers ([1], [2] [4] and [5])

3 Extracting rhythm from the movements

Following the rhythm is the most important task of conductor follower. As can be seen in figure 3 vertical movement can be used for rhythm detection well. The most simple method is to update playing tempo and

Figure 4: Conductor in data dress suit.

adjust score position whenever down-beat is observed (used by Mathews [1]). Reaction time is by definition rather long since tempo cannot be adjusted between down-beats. Mathews's system updated tempo and score position abruptly as new down-beat was detected. Better way to use down-beats as a reference can be based on oscillators. The oscillator does not react instantly to down-beats. It makes a tempo change curve to catch the rhythm. Toiviainen [7] has created oscillators that can be used to track the rhythm of played music note-on events as control information. Replacing the note-on times with down-beat times makes this an attractive conductor following technique.

More information can be extracted using more motion data. Predicting the next down-beat is a key to fast reaction. Brecht [4] used the zero crossings of velocity and acceleration for basic data. This was fed to an artificial neural network. Our approach is similar, but we use delayed position and velocity values as ANN input. Since score position and conductor motion phase can be constantly compared the rhythm can be adjusted with each new motion sample, not only at zero crossings. If radical tempo changes need to be made the rhythm can be adjusted even before the down-beat.

3.1 Requirements of the phase estimator

The aim of phase detector is to transform the motion parameters (position, velocity, acceleration etc.) to phase angle. This phase angle estimates how long time it will take to get to the bottom of the beat. The score position increases monotonously as the piece is played. It cannot be directly compared with the cyclic motion phase estimate. The score position of the follower is mapped to phase (see 5). The difference between these two phases is the estimated phase error that needs to be corrected.

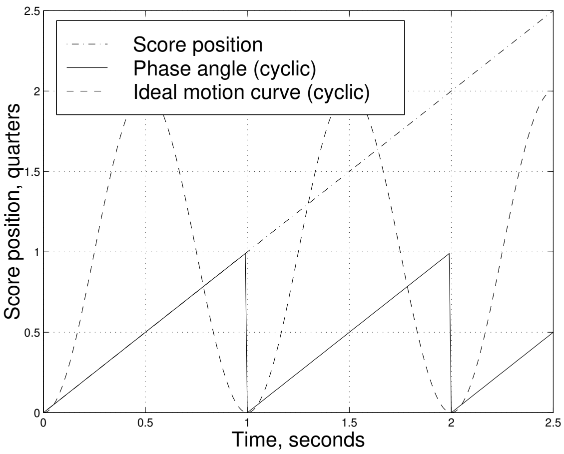

Figure 5: How the linearly growing score position is mapped to cyclic motion phase angle.

3.2 Multi layer perceptron

ANNs do not produce the sawtooth output needed very well. The discontinuous jump from one extreme to another is difficult for most nets. Our solution is using sine and cosine of the phase as network outputs (fig 6). These two values imply the current phase angle.

We use multi layer perceptron (MLP) ANN to detect beat phase. The network input consists of current position and delayed velocity values.

The MLP is trained with backpropagation learning rule [8]. Current experience shows this method is reliable and robust in tracking beats. The same net can track tempos between 40 and 180 beats per minute and movement amplitudes between 10 cm and 1.5 metres without any parameter adjustments. The net can be trained anew for each conductor to match their style.

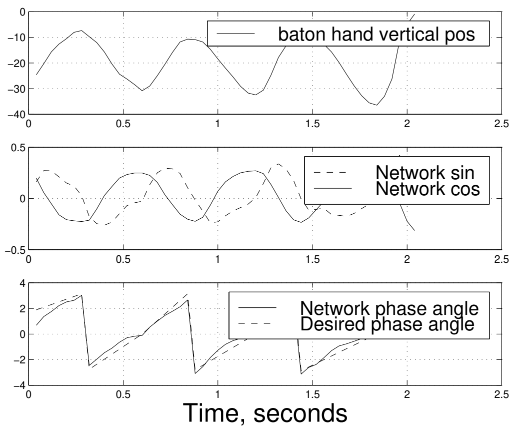

Figure 6: MLP results. Uneven motion prevents accurate tracking. One cycle of output should match one beat with raising zero-crossings matching down-beats.

The problematic feature of MLP is its lack of error measure. The user of the net must device a way to detect how well the net is performing. In out scheme the noise in network output is monitored to detect MLP misbehaviour.

3.3 Heuristics for rhythm control

Conductors do not always time their movements with high precision. Even when tempo is constant the down-beats may occur at non-uniform intervals. Besides, natural gestures always vary in form. This prevents exact beat phase estimation (see fig. 6). Extra processing is needed to control rhythm.

We use heuristics to smooth MLP output and add special behaviour.

The follower should react fast when the conductor means to change tempo, but ignore small unintentional deviations in conducted rhythm. The noise in MLP output should have minimum effect on the virtual musicians tempo.

The basic tempo estimation is taken from the down-beat intervals. The beat phase estimate is used to fine tune the base tempo.

The MLP output is considered to be the correct beat phase. The current music beat phase can be calculated from the current score position. The difference between these two is the following error that should be corrected. The most musical way to do tempo correction is still unknown. We use a variant of correction function Toiviainen has used for oscillator tempo correction [7].

The tempo of the follower is constantly corrected for phase errors. At each new motion sample the error is calculated and a new correction function (fig 7) is initialized. The correction is the sum of latest correction functions values. When the network output is noisy the output jumps around the true motion phase angle as in fig 6. On some samples, the network demands tempo acceleration and during others deceleration. Since the correction fuctions overlap eacj other the tempo corrections will mostly cancel each other and the output rhythm is stable.

If the conductor really is changing the tempo the error between estimated beat phase and played beat phase will persist. In this case the sum of correction functions will rapidly increase and the follower catches up with the conductor.

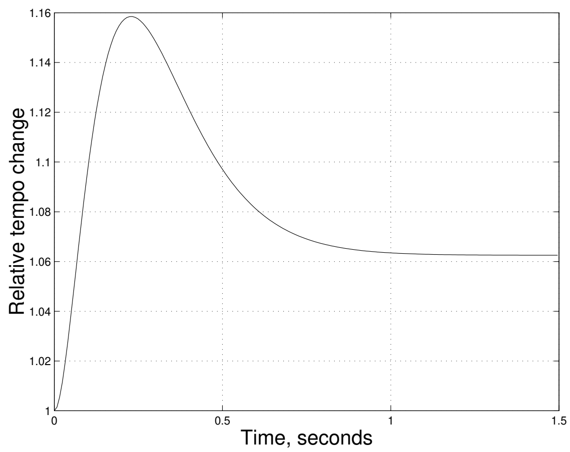

Figure 7: The used correction function. Temporary tempo overshoot will hopefully catch conductor's rhythm. The correction is time-distributed to make tempohcanges smooth.

Digital filtering techniques (PID and Kalman filters) could also be used to correct the rhythm. We feel that adding extra behaviour (stops etc.) is easier when the tempo changes are done with explicit change curves.

The music may demand special behaviour. A temporary stop is a fine example (fermata). The playing of music must stop completely for a short time and resume later to normal. It is very difficult to detect these cases quickly enough by only looking at the conductor. A list of musical hints about the piece needs to be given to the virtual (and real) musician to warn about the potential for fermata.

The amount of musical intelligence of the virtual musician is a function of quality and number of heuristic rules governing its behaviour. The heuristics are based on a simplified set of normal conducting gestures. The behaviour includes starts, fermatas, accelerandos (accelerating tempo), ritardandos (decelerating tempo), abrupt tempo changes and premature endings of the piece.

4 Conclusions

The follower gives a conductor powerful tempo and dynamics control of a virtual band.

Currently the follower performs mostly satisfactorily, but when the conductor makes gestures that conflict with the assumptions the follower may get badly lost. The follower assumes motion that matches the original training data and the working principles of the heuristics. Any new gestures are likely to cause major difficulties for the follower.

It can react faster than real musicians to tempo changes. At the same time the rhythm it generates can be unmusical. The gesture vocabulary is limited to tempo and dynamics control.

The conductor follower has been developed as part of the Digital Interactive Virtual Acoustics (DIVA) project [9]. The project integrates conductor following, orchestra animation, virtual acoustics and physical modeling of musical instruments into one artificial reality experience.

References

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

[2] D. Keane and P. Gross. The MIDI Baton II. In *Proceedings of the International Computer Music Conference*, San Francisco, 1989. Computer Music Asociation.

[3] P. Carosi and G. Bertini. Light Baton: A System for conducting computer music performance. In *Proceedings of the International Computer Music Conference*, San Francisco, 1992. Computer Music Asociation.

[4] B. Brecht and G. Garnett. Conductor Follower. In *Proceedings of the International Computer Music Conference*, San Francisco, 1995. Computer Music Asociation.

[5] F. Tobey and I. Fujinaga. Extraction of Conducting Gestures in 3D Space. In *Proceedings of the International Computer Music Conference*, San Francisco, 1996. Computer Music Asociation.

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

[7] P. Toiviainen. An interactive MIDI accompanist. *Computer Music Journal*, (In print), 1998.

[8] J. Schalkoff. *Pattern Recognition: Statistical, Structural, and Neural Approaches.* J. Wiley, 1992.

[9] V. Välimäki L. Savioja J. Huopaniemi T. Huotilainen T. Takala, R. Hänninen and M. Karjalainen. An Integrated system for Virtual Audio Reality. In *Preprint 4229 (M-4), Copenhagen: 100th AES Convention of Acoustic Engineering Society: 46*, Copenhagen, sweden, 1996.