

# Interacting with a Virtual Conductor

Pieter Bos, Dennis Reidsma, Zsófia Ruttkay, Anton Nijholt

HMI, Dept. of CS, University of Twente,  
PO Box 217, 7500AE Enschede, The Netherlands

anijholt@ewi.utwente.nl  
<http://hmi.ewi.utwente.nl/>

**Abstract.** This paper presents a virtual embodied agent that can conduct musicians in a live performance. The virtual conductor conducts music specified by a MIDI file and uses input from a microphone to react to the tempo of the musicians. The current implementation of the virtual conductor can interact with musicians, leading and following them while they are playing music. Different time signatures and dynamic markings in music are supported.

## 1 Introduction

Recordings of orchestral music are said to be the interpretation of the conductor in front of the ensemble. A human conductor uses words, gestures, gaze, head movements and facial expressions to make musicians play together in the right tempo, phrasing, style and dynamics, according to his interpretation of the music. She also interacts with musicians: The musicians react to the gestures of the conductor, and the conductor in turn reacts to the music played by the musicians. So far, no other known virtual conductor can conduct musicians interactively.

In this paper an implementation of a Virtual Conductor is presented that is capable of conducting musicians in a live performance. The audio analysis of the music played by the (human) musicians and the animation of the virtual conductor are discussed, as well as the algorithms that are used to establish the two-directional interaction between conductor and musicians in patterns of leading and following. Furthermore a short outline of planned evaluations is given.

## 2 Related Work

Wang *et al.* describe a virtual conductor that synthesizes conducting gestures using kernel based hidden Markov models [1]. The system is trained by capturing data from a real conductor, extracting the beat from her movements. It can then conduct similar music in the same meter and tempo with style variations. The resulting conductor, however, is not interactive in the sense described in the introduction. It contains no beat tracking or tempo following modules (the beats in music have to be marked by a

human) and there is no model for the interaction between conductor and musicians. Also no evaluation of this virtual conductor has been given. Ruttkay *et al.* synthesized conductor movements to demonstrate the capabilities of a high-level language to describe gestures [2]. This system does not react to music, although it has the possibility to adjust the conducting movements dynamically.

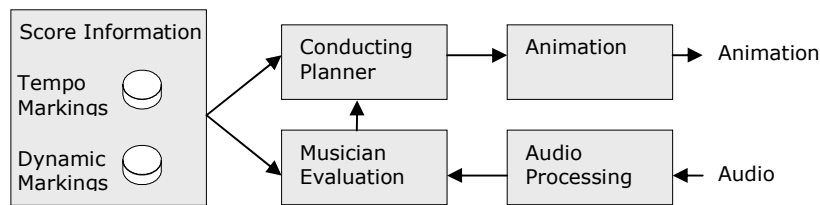
Many systems have been made that try to follow a human conductor. They use, for example, a special baton [3], a jacket equipped with sensors [4] or webcams [5] to track conducting movements. Strategies to recognize gestures vary from detecting simple up and down movements [3] through a more elaborate system that can detect detailed conducting movements [4] to one that allows extra system-specific movements to control music [5]. Most systems are built to control the playback of music (MIDI or audio file) that is altered in response to conducting slower or faster, conducting a subgroup of instruments or conducting with bigger or smaller gestures.

Automatic accompaniment systems were first presented in 1984, most notably by Dannenberg [6] and Vercoe [7]. These systems followed MIDI instruments and adapted an accompaniment to match what was played. More recently, Raphael [8] has researched a self-learning system which follows real instruments and can provide accompaniments that would not be playable by human performers. The main difference with the virtual conductor is that such systems follow musicians instead of attempting to explicitly lead them.

For an overview of related work in tracking tempo and beat, another important requirement for a virtual conductor, the reader is referred to the qualitative and the quantitative reviews of tempo trackers presented in [9] and [10], respectively.

### 3 Functions and Architecture of the Virtual Conductor

A virtual conductor capable of leading, and reacting to, a live performance has to be able to perform several tasks in real time. The conductor should possess knowledge of the music to be conducted, should be able to translate this knowledge to gestures and to produce these gestures. The conductor should extract features from music and react to them, based on information of the knowledge of the score. The reactions should be tailored to elicit the desired response from the musicians.



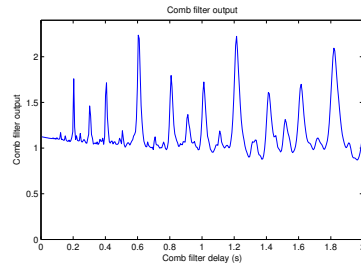
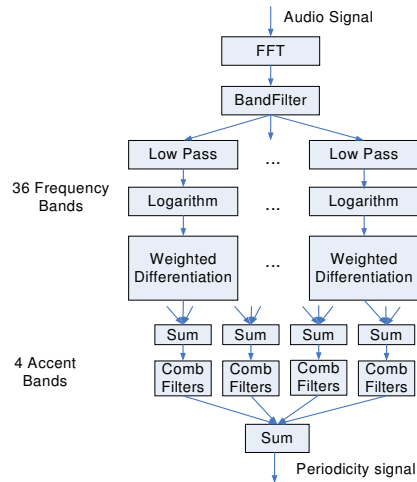
**Fig. 1.** Architecture overview of the Virtual Conductor

Figure 1 shows a schematic overview of the architecture of our implementation of the Virtual Conductor. The audio from the human musicians is first processed by the

Audio Processor, to detect volume and tempo. Then the Musician Evaluation compares the music with the original score (currently stored in MIDI) to determine the conducting style (lead, follow, dynamic indications, required corrective feedback to musicians, etc). The Conducting Planner generates the appropriate conducting movements based on the score and the Musician Evaluation. These are then animated. Each of these elements is discussed in more detail in the following sections.

### 3.1 Beat and Tempo Tracking

To enable the virtual conductor to detect the tempo of music from an audio signal, a beat detector has been implemented. The beat detector is based on the beat detectors of Scheirer [11] and Klapuri [12]. A schematic overview of the beat detector is presented in Figure 2. The first stage of the beat detector consists of an accentuation detector in several frequency bands. Then a bank of comb filter resonators is used to detect periodicity in these ‘accent bands’, as Klapuri calls them. As a last step, the correct tempo is extracted from this signal.



**Fig. 2.** Schematic overview of the Beat detector

**Fig. 3.** Periodicity signal

To detect periodicity in these accent bands, a bank of comb filters is applied. Each filter has its own delay: delays of up to 2 seconds are used, with 11.5 ms steps. The output from one of these filters is a measure of the periodicity of the music at that delay. The periodicity signal, with a clear pattern of peaks, for a fragment of music with a strong beat is shown in Figure 3. The tempo of this music fragment is around 98 bpm, which corresponds to the largest peak shown. We define a peak as a local maximum in the graph that is above 70% of the outputs of all the comb filters. The peaks will form a pattern with an equal interval, which is detected. Peaks outside that pattern are ignored. In the case of the virtual conductor an estimate of the played

tempo is already known, so the peak closest to the conducted tempo is selected as the current detected tempo. Accuracy is measured as the difference between the maximum and minimum of the comb filter outputs, multiplied by the number of peaks detected in the pattern.

A considerable latency is introduced by the sound card, audio processing and movement planning. It turned out that in the current setup the latency was not high enough to unduly disturb the musicians. However, we also wrote a calibration method where someone taps along with the virtual conductor to determine the average latency. This latency could be used as an offset to decrease its impact on the interaction.

### 3.2 Interacting with the Tempo of Musicians

If an ensemble is playing too slow or too fast, a (human) conductor should lead them back to the correct tempo. She can choose to lead strictly or more leniently, but completely ignoring the musicians' tempo and conducting like a metronome set at the right tempo will not work. A conductor must incorporate some sense of the actual tempo at which the musicians play in her conducting, or else she will lose control. A naïve strategy for a Virtual Conductor could be to use the conducting tempo  $t_c$  defined in formula 1 as a weighted average of the correct tempo  $t_o$  and the detected tempo  $t_d$ .

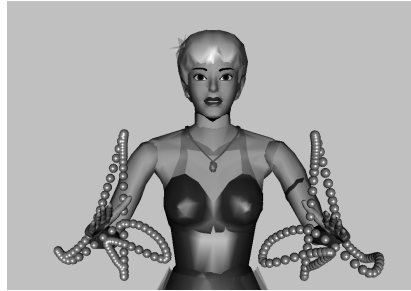
$$t_c = (1-\lambda) t_o + \lambda t_d \quad (1)$$

If the musicians play too slowly, the virtual conductor will conduct *a little bit* faster than they are playing. When the musicians follow him, he will conduct faster yet, till the correct tempo is reached again. The ratio  $\lambda$  determines how strict the conductor is. However, informal tests showed that this way of correcting feels restrictive at high values of  $\lambda$  and that the conductor does not lead enough at low values of  $\lambda$ . Our solution to this problem has been to make  $\lambda$  adaptive over time. When the tempo of the musicians deviates from the correct one,  $\lambda$  is initialised to a low value  $\lambda_L$ . Then over the period of  $n$  beats,  $\lambda$  is increased to a higher value  $\lambda_H$ . This ensures that the conductor can effectively lead the musicians: first the system makes sure that musicians and conductor are in a synchronized tempo, and then the tempo is *gradually* corrected till the musicians are playing at the right tempo again. Different settings of the parameters result in a conductor which leads and follows differently. Experiments will have to show what values are acceptable for the different parameters in which situations. Care has to be taken that the conductor stays in control, yet does not annoy the musicians with too strict a tempo.

### 3.3 Conducting Gestures

Based on extensive discussions with a human conductor, basic conducting gestures (1-, 2-, 3- and 4-beat patterns) have been defined using inverse kinematics and hermite splines, with adjustable amplitude to allow for conducting with larger or

smaller gestures. The appropriately modified conducting gestures are animated with the animation framework developed in our group, in the chosen conducting tempo  $t_c$ .



**Fig. 4.** A screenshot of the virtual conductor application, with the path of the 4-beat pattern

## 4 Evaluation

A pre-test has been done with four human musicians. They could play music reliably with the virtual conductor after a few attempts. Improvements to the conductor are being made based on this pre-test. An evaluation plan consisting of several experiments has been designed. The evaluations will be performed on the current version of the virtual conductor with small groups of real musicians. A few short pieces of music will be conducted in several variations: slow, fast, changing tempo, variations in leading parameters, etcetera, based on dynamic markings (defined in the internal score representation) that are not always available to the musicians. The reactions of the musicians and the characteristics of their performance in different situations will be analysed and used to extend and improve our Virtual Conductor system.

## 5 Conclusions and Future Work

A Virtual Conductor that incorporates expert knowledge from a professional conductor has been designed and implemented. To our knowledge, it is the first virtual conductor that can conduct different meters and tempos as well as tempo variations and at the same time is also able to *interact* with the human musicians that it conducts. Currently it is able to lead musicians through tempo changes and to correct musicians if they play too slowly or too fast. The current version will be evaluated soon and extended further in the coming months.

Future additions to the conductor will partially depend on the results of the evaluation. One expected extension is a score following algorithm, to be used instead of the current, less accurate, beat detector. A good score following algorithm may be

able to detect rhythmic mistakes and wrong notes, giving more opportunities for feedback from the conductor. Such an algorithm should be adapted to or designed specifically for the purpose of the conductor: unlike with usual applications of score following, an estimation of the location in the music is already known from the conducting plan.

The gesture repertoire of the conductor will be extended to allow the conductor to indicate more cues, to respond better to volume and tempo changes and to make the conductor appear more lifelike. In a longer term, this would include getting the attention of musicians, conducting more clearly when the musicians do not play a stable tempo and indicating legato and staccato. Indicating cues and gestures to specific musicians rather than to a group of musicians would be an important addition. This would need a much more detailed (individual) audio analysis as well as a good implementation of models of eye contact: no trivial challenge.

### Acknowledgements

Thanks go to the “human conductor” Daphne Wassink, for her comments and valuable input on the virtual conductor, and the musicians who participated in the first evaluation tests.

### References

1. Wang, T., Zheng, N., Li, Y., Xu, Y. and Shum, H. Learning kernel-based HMMs for dynamic sequence synthesis. Veloso, M. and Kambhampati, S. (eds), *Graphical Models* 65:206-221, 2003
2. Ruttkay, Zs., Huang, A. and Eliëns, A. The Conductor: Gestures for embodied agents with logic programming, in *Proc. of the 2nd Hungarian Computer Graphics Conference*, Budapest, pp. 9-16, 2003
3. Borchers, J., Lee, E., Samminger, W. and Mühlhäuser, M. Personal orchestra: a real-time audio/video system for interactive conducting, *Multimedia Systems*, 9:458-465, 2004
4. Marrin Nakra, T. Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture. Ph.D. Thesis, Media Laboratory. Cambridge, MA, Mass. Inst. of Technology, 2000
5. Murphy, D., Andersen, T.H. and Jensen, K. Conducting Audio Files via Computer Vision, in *GW03*, pp. 529-540, 2003
6. Dannenberg, R. and Mukaino, H. New Techniques for Enhanced Quality of Computer Accompaniment, in *Proc. of the International Computer Music Conference*, Computer Music Association, pp. 243-249, 1988
7. Vercoe, B. The synthetic performer in the context of live musical performance, *Proc. Of the International Computer Music Association*, p. 185, 1984
8. Raphael C. Musical Accompaniment Systems, *Chance Magazine* 17:4, pp. 17-22, 2004
9. Gouyon, F. and Dixon, S. A Review of Automatic Rhythm Description Systems, *Computer music journal*, 29:34-54, 2005
10. Gouyon, F., Klapuri, A., Dixon, S., Alonso, M., Tzanetakis, G., Uhle, C. and Cano, P. An Experimental Comparison of Audio Tempo Induction Algorithms, *IEEE Transactions on Speech and Audio Processing*, 2006
11. Scheirer, E.D. Tempo and beat analysis of acoustic musical signals, *Journal of the Acoustical Society of America*, 103:558-601, 1998
12. Klapuri, A., Eronen, A. and Astola, J. Analysis of the meter of acoustic musical signals, *IEEE transactions on Speech and Audio Processing*, 2006