

Temporal Interaction Between an Artificial Orchestra Conductor and Human Musicians

DENNIS REIDSMA, ANTON NIJHOLT, and PIETER BOS
Human Media Interaction, University of Twente

The Virtual Conductor project concerns the development of the first properly interactive virtual orchestra conductor—a Virtual Human that can conduct a piece of music through interaction with musicians, leading and following them while they are playing. This article describes our motivation for developing such a system; related work in the areas of ambient entertainment and coordinated timing, automatic music processing, virtual humans, and conducting; the design, implementation, and evaluation of the Virtual Conductor, and, finally, contains a discussion of the resulting system and expected developments in the (still ongoing) Virtual Conductor project.

Categories and Subject Descriptors: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—*Animations; artificial, augmented, and virtual realities; audio input/output; evaluation/methodology*; H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.5.5 [Information Interfaces and Presentation]: Sound and Music Computing—*Methodologies and techniques; modeling; signal analysis, synthesis, and processing; systems*; J.5 [Computer Applications]: Arts and Humanities—*Music; performing arts*

General Terms: Design, Experimentation, Human Factors

Additional Key Words and Phrases: Embodied agents, virtual conductor, continuous interaction

ACM Reference Format:

Reidsma, D., Nijholt, A., and Bos, P. 2008. Temporal interaction between an artificial orchestra conductor and human musicians. *ACM Comput. Entertain.* 6, 4, Article 53 (December 2008), 22 pages. DOI = 10.1145/1461999.1462005 <http://doi.acm.org/10.1145/1461999.1462005>

1. INTRODUCTION

One interest of our research group is in developing ambient entertainment technologies and applications that interact in a coordinated way with human partners using a multitude of different sensors, observing many characteristics

This research was supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie).

Author's address: D. Reidsma, Human Media Interaction, University of Twente, PO Box 217, 7500AE Enschede, The Netherlands; email: dennisr@ewi.utwente.nl.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2008 ACM 1544-3574/2008/12-ART53 \$5.00 DOI = 10.1145/1461999.1462005 <http://doi.acm.org/10.1145/1461999.1462005>



Fig. 1. The Virtual Conductor during a demo performance. Photo with permission from: Stenden Hogeschool and Henk Postma.

of the partner, and using many different channels of expression (such as sound, visuals, speech, and embodiment with gestures and facial expressions).

The Virtual Conductor project concerns the development of the first properly interactive virtual orchestra conductor—a Virtual Human (VH) that can conduct a piece of music through interaction with musicians, leading and following them while they are playing (see Figure 1). Its *observations* consist of different manners of musical processing of the incoming signal from a microphone. The *forms of its expressions* are defined by the possible conducting patterns (1, 2, 3, and 4 beat measures) and by the timing, speed, amplitude, and smoothness with which those patterns are expressed. The *interaction* is focused on the tempo of the musicians and includes a correction module that interactively corrects the tempo when the musicians are playing too slowly or too quickly.

This article describes our motivation for developing such a system; related work in the areas of ambient entertainment and coordinated timing, automatic music processing, virtual humans, and conducting; the design, implementation and evaluation of the Virtual Conductor; and, finally, contains a discussion of the resulting system, general applicability of the ideas and technology developed in this project, and expected developments in the (still ongoing) Virtual Conductor project.

2. MOTIVATION

Music fascinates. Music inspires. Music entertains. It constitutes quite a major part of our entertainment without the need for computers. And, more and more, it is becoming a theme in computer-based entertainment. Games in which interaction with or through music plays a central role are on the rise (see, for example, games such as *Guitar Hero*, *Dance Dance Revolution*, *Donkey Konga*, and many, many more). However, for many of those games the interaction through music is mostly one-way: the player must follow a rhythm or riff presented by the computer to achieve a set goal. When a group of people make music, interaction is inherently simultaneous and two-way. Both partners in a musical cooperation are alert to what the other is doing and

adapt their own performance to mesh [Schögler 1998]. In the Virtual Conductor project presented in this article a major element is the mutual interaction between system and musician with respect to tempo and timing. Some ideas concerning temporal coordination in interaction have been worked out preliminarily in our paper “Mutually Coordinated Anticipatory Multimodal Interaction” [Nijholt et al. 2008], in relation to topics such as synchrony. Here we just note the strong positive relation found in literature between synchrony and positive affect, or between synchrony and a positive evaluation of the interaction in human-human interaction [Crown 1991; Ramseyer and Tschacher 2008; Nagaoka et al. 2007], but also in human-computer interaction [Suzuki et al. 2003; Bailenson and Yee 2005; Robins et al. 2005]. Given the literature, it seems a reasonable assumption that implementation of modules for synchrony can add to the enjoyment and engagement of users of computational entertainment applications. The Virtual Conductor can be seen as one of the first ambient entertainment applications that take a step in the direction of harnessing interactional synchrony for improving the enjoyment and engagement of the user.

A Virtual Conductor system can be used in several ways. An edutainment application of such technology could be in teaching student conductors. As a reflective tool, the system could show good examples, as well as examples of typical conducting mistakes or allow the student conductor to visualize different ways of conducting a passage to see what it looks like. In combination with the complement of this artificial conductor, namely an artificial orchestra such as the one on display in the Vienna House of Music, a system could be envisioned that detects the student’s mistakes and graphically shows them to the student in combination with a better way of conducting. As a platform for experimentation, the system could be used to experiment with the effect of controlled variations in conducting on the music played by an ensemble.

We can also envision this conductor developed further as a rehearsal conductor. The time in which a human conductor can work with a certain ensemble is often limited; if a Virtual Conductor could be used to rehearse the more technical aspects of a piece of music, this would leave the human conductor more time to work on the creative and expressive musical aspects of a performance.

Finally, a Virtual Conductor could also be made available through the Internet to provide the casually interested layman with easy and engaging access to knowledge about, and some do-it-yourself experimentation with, conducting.

3. RELATED WORK

The work on the Virtual Conductor builds on a number of different fields. This section presents relevant related work for those fields. We discuss work related to the Virtual Conductor as a Virtual Human, specifically in the context of music interaction. After that, we present other work on conducting, technology and applications. Most of that work concerns a scenario that “complements” our project: technology to follow a human conductor and have a virtual orchestra react to his or her conducting. Finally, on the perception side, we sketch the very active area of automatic music processing, as the Virtual Conductor

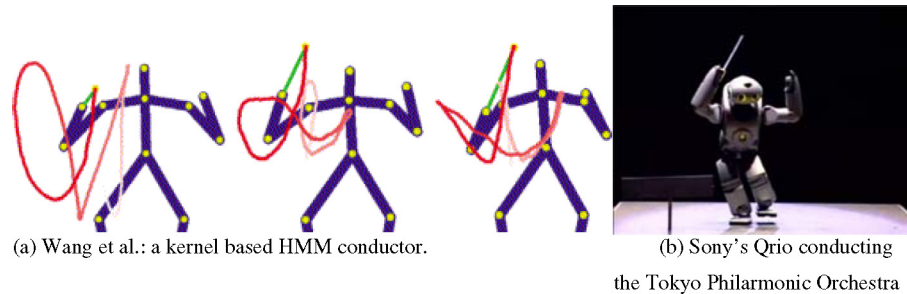


Fig. 2. Examples of other virtual conductors. Fig. 2(a): with permission from Elsevier; Fig. 2(b): with permission from Sony Inc.

uses (adapted versions of) state-of-the-art algorithms for interpreting the music played by the musicians.

3.1 Virtual Humans

To our knowledge, our system is the first *interactive* virtual conductor. However, a few other systems have been built in the past in which conducting movements were synthesized. Ruttkay et al. [2003] synthesize conductor movements to demonstrate a scripting language for humanoid animations. Their focus is on showing how to specify compound nonverbal behavior by combining smaller units in parallel and sequential constructions. Wang et al. [2003] describe a virtual conductor that learns the shape of conducting gestures from examples by human conductors using a kernel-based hidden Markov model (KHMM). Their conductor is used as an example to show that KHMMs can be used to synthesize gestures (see Figure 2(a)). Input to the algorithm is a combination of movements from a human conductor and a synchronized recording of music. Loudness, pitch, and beat are used to describe the music, positions, and movement of several joints of the conductor to describe the gestures. Their virtual conductor, trained with this data, can conduct music similar to that on which it was trained, that is, similar in time and tempo. Basic movements are used and style variations are shown. The conductor does not have audio tempo tracking; the music is semi-automatically analyzed using the movements from the real conductor to track beats. The conductor cannot interact with musicians; it can only synthesize an animation from an annotated audio file. The authors suggest that tempo changes can be handled by blending multiple trained models. This however has not been done yet.

Finally, movie files of the Honda robot Asimo conducting the Detroit Symphony Orchestra and the Sony Qrio robot conducting the Tokyo Philharmonic Orchestra in a performance of Beethoven's Fifth Symphony can be found on the Internet (see Figure 2(b)). We could not find publications describing how this was achieved. Other work by Sony researchers, on dance interaction with the same Qrio robot, shows how the authors experimented with a generative model of entrainment between Qrio and the user in order to achieve engaging harmonious interaction [Tanaka and Suzuki 2004].

Leaving aside conducting for a moment, we can see a lot of work with virtual humans and entertainment/edutainment using music. Many examples can be found of embodied agents reacting to music. Our virtual dancer system [Reidsma et al. 2006] lets a Virtual Human move in time to music using a beat tracker, interacting with a human dancer through robust computer vision. There are other, similar dancers, such as Cindy by Goto [2001] or the dancer of Shiratori et al. [2006] which makes use of the structure of music to plan and select its dance moves. Mancini et al. [2007] describe Greta, an embodied conversational agent capable of showing emotions. Greta adapts her facial expression to music using a system that detects emotion in music. DiPaola and Arya [2006] do the same, but also allow the face to be deformed in a nonrealistic (artistic) manner. Embodied agents have also been used as intelligent tutors for teaching the violin [Yin et al. 2005] or as a peer for children in their musical development [Jansen et al. 2006]. There are many more examples of these types of virtual humans interacting with or through music.

3.2 Conductor Following

Animating conducting gestures for virtual humans is not the only application of technology in the context of conducting. An important category of systems, often called “virtual conductors” are the ones that complement our situation, namely, conductor *following* systems. Such a system consists of some way to measure the movements of a conductor, gesture recognition to extract information from these movements, and often also a virtual orchestra, the performance of which can be altered through variation in the conducting gestures. Many systems use some sort of sensor that a conductor has to wear. This can be an electronic conducting baton [Lee et al. 2006; Ilmonen and Takala 1999], a sensor glove [Lee et al. 1992], or, for example, a jacket measuring the conductor’s movements in detail [Nakra 2000]. Murphy et al. [2003] and Segen [1999] use a camera to follow a normal conducting baton; Kolesnik and Wanderly [2004] use a camera to track a colored glove worn by the conductor. All of these systems have some way of recognizing naive or professional conducting gestures and adapting the playback of a prerecorded piece of music in response to detected changes in the conducting gestures. Lately, several game development companies have announced upcoming game titles based on this concept using the Wiimote controller.

3.3 Automatic Music Processing

In the Virtual Conductor project we work with (adapted versions of) state-of-the-art algorithms for interpreting the music played by the musicians. Two basic types of automatic music processing algorithms are reviewed here: algorithms that only use audio input to follow music and algorithms that also use information from a score. The algorithms not requiring a score are generally called beat-tracking or tempo-tracking algorithms, or transcription systems. Algorithms of the other type, which require a score, are called score following or score-aligning algorithms. A short summary will be given of some of these

algorithms and of their features and performance. For an interactive Virtual Conductor only the real time variants are of use.

3.3.1 Beat-Tracking Algorithms. There are many beat-tracking algorithms. For an overview of the field, we refer to the paper of Gouyon and Dixon [2005], in which the authors use a single framework for a qualitative comparison of many automatic rhythm description systems, including beat trackers, tempo induction algorithms and automated rhythm transcription. Each beat detector is described as three functional units: feature extraction, pulse induction, and pulse tracking. The separate functional units are compared. This division is used by other authors as well [Alonso et al. 2004; Goto 2001]. Beat-tracking algorithms generally work without prior knowledge of the musical piece being performed. However, they can be adapted to work with score information. For example, the relation between different metrical levels, which can be obtained from the score, is often used in beat detectors. Although audio features such as chord changes [Goto 2001] and inter-onset intervals [Dixon 2003] have been proposed, most beat-detector systems use some form of onset time or accents as features [Alonso et al. 2004; Dixon 2003; Goto 2001]. As pulse induction, most algorithms use either autocorrelation or a bank of comb filters. Pulse-tracking can be done with cross-correlation between the expected pulses and the detected pulses [Alonso et al. 2004], by probabilistic modeling [Klapuri et al. 2006] or be derived directly from the pulse induction step [Scheirer 2006].

An extensive quantitative comparison of 11 different tempo-induction algorithms is presented by Gouyon et al. [2006]. The algorithms are run on a data set of 12 hours and 36 minutes of audio. The data set consists of over 2000 short loops, 698 short ballroom dancing pieces, and 465 song excerpts from 9 different genres. The songs were hand-annotated. The ground truth of the loops was known beforehand. Accuracy was measured in two ways: the number of songs that were correctly identified within 4% accuracy, called accuracy 1, and the number of songs that were correctly identified plus the songs identified having a tempo that is an integer multiple of the real tempo, accuracy 2. The algorithm by Klapuri et al. [2006] was the “winner” with a score of 85.01% for accuracy 2 and 67.29% for accuracy 1. This algorithm was the most robust when noise was added to the audio files as well. The algorithm was adapted by Seppänen et al. [2006] to run on devices with limited computing power by optimizing and simplifying the feature detection and simplifying the music model. Their evaluation suggests only a minor performance loss.

3.3.2 Score-Following. Score-following algorithms, or online-tracking algorithms, use knowledge about the musical score to follow a musical performance. Some of these algorithms only work on input data in MIDI format, instead of audio [Pardo and Birmingham 2005; Schwartz et al. 2004], requiring an existing automated transcription system, or the use of MIDI instruments, and therefore making them unsuitable for our purposes. Raphael [2004] describes a score-following algorithm that works on polyphonic audio recordings. The algorithm works on chord changes and searches through a tree with the different options to determine the tempo of the music being played. It was tested

on orchestral classical music and worked accurately for at least a few minutes in most pieces before losing track of the music. The algorithm produces errors when no chord changes occur, on long tones. Dannenberg and Hu [2003] describe a score-following system that works on audio recordings. The recording is split into short segments, for each of which a feature vector is calculated. The best performing feature was the chroma vector containing the spectral energy in every pitch-class (C, C\#, D, . . . , B). Chroma vectors for the score are calculated by summing the notes from a MIDI file directly into a chroma vector, or by synthesizing the MIDI and calculating the vectors from the resulting audio stream. The stream of chroma vectors of audio and score are both normalized. A similarity matrix between audio and score is created using the Euclidean distance between two chroma vectors as a similarity measure. The mapping from recording to score is found through dynamic time-warping, tracing a highest similarity path through the matrix from the end of the music to the beginning. Because the time-warping algorithm has to start at the end of the music rather than the beginning, the algorithm cannot work in real-time. Dixon [2005] adapted the dynamic time-warping algorithm for real-time use. His online time-warping algorithm works by predicting the current location in the similarity matrix and calculating the shortest path back to the start of the music. Only the part of the matrix close to the prediction is calculated to give the algorithm linear instead of quadratic time and space efficiency. This algorithm proves to be effective. Concluding, we can say that state-of-the-art score-following algorithms currently align music to a score well enough to use with a Virtual Conductor, although some problems remain.

4. GLOBAL DESIGN

We have designed, and implemented, a Virtual Conductor application that consists of a Virtual Human (VH) that can conduct a piece of music through interaction with musicians, leading and following them while they are playing (see Figure 1). The design of a Virtual Human application, or indeed any kind of ambient entertainment system, contains three main elements. In the first place, such a system needs sensors with which to observe the human interaction partner(s), and, to interpret the observations correctly, background knowledge of possible activities that those humans can perform in the context of the application. For a Virtual Conductor this largely concerns music analysis algorithms. In the second place, a system needs to be able to communicate information, goals, and intentions to the user, in this case through producing appropriate verbal and nonverbal communicative (conducting) expressions for the embodiment of the Virtual Human. In the third place, these two come together in the patterns of interaction and the feedback loops between production and perception through the reactions of the human interaction partner(s). For a Virtual Conductor, this feedback loop contains elements such as the decision to signal something to the musicians based on how they are playing, the execution of music-conducting expressions, the way those signals are taken up by the human musicians playing the music, and perception by the conductor of the (altered) performance of the music. This global structure, shown in Figure 3,

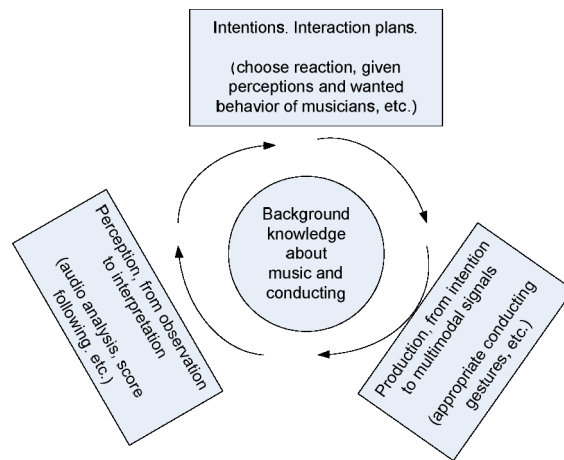


Fig. 3. A global design for the Virtual Conductor.

is explained in more detail below. Detailed descriptions of different modules in the resulting system are given in Section 5. Clearly, we cannot develop a virtual conductor without building upon knowledge of how human conductors do their work; a lot of the information in the following sections is based on interviews with an informant who shared her experience as a choir conductor with us.

4.1 Music Analysis and Background Knowledge

A human conductor knows the piece of music that is being performed. She knows the voices that the different instruments should play, as well as the interpretation with which she wants the piece to be played (timing, volume dynamics, timbre, etc.). During a performance or rehearsal she hears the musicians playing. She knows exactly *where* in the score the musicians are currently playing, she hears every sour note or precision error, but she also hears when the ensemble is doing very well indeed.

For our Virtual Conductor, the background knowledge about the piece of music contains only the time of the music (number of counts per measure), the notes of the different voices, the global tempo and tempo changes, and the volume indications. Analysis of the music as it is being played is done on the audio signal recorded through a single microphone. The main analyses performed are beat-tracking, score-following (marking where in the score the musicians are playing), and rudimentary volume analysis. This allows the Virtual Conductor to detect deviations in tempo and volume. Timbre and expression have not yet been implemented in the Virtual Conductor system.

4.2 Producing Conducting Expressions

A human conductor will lead an ensemble through mostly nonverbal interaction. The basic beat of the music is indicated using so-called “beat patterns,” with different patterns for different time signatures (cf., Section 5.1). Style and volume can be indicated by altering the shape and dynamics with which the

beat patterns are performed. Using additional expressions such as left-handed gestures, facial expressions or gaze, a human conductor can indicate the same or other aspects such as entrance cues or timbre. The additional expressions are not part of the work presented here.

The Virtual Conductor is based upon an existing framework for virtual humans, developed at HMI,¹ which is also capable of making gestures. In order to be able to conduct a small ensemble, its gesture repertoire needs to be extended with at least the basic beat patterns. Therefore, the beat patterns of human conductors have been analyzed in order to reproduce them as animations on the Virtual Human. The resulting animations are parameterizable in such a way that they convey exactly the information that needs to be conveyed (volume, tempo, and tempo changes). Because the Virtual Conductor needs to react adaptively to how the musicians take up its signals (see also the next section), a module was implemented to change the precise timing of the gestures on-the-fly (cf., Section 5.4).

4.3 Interaction between Musicians and Conductor

Being able to make the right conducting gestures for a given piece of music is not enough to be a conductor. Conducting an ensemble is not a one-way process, but a highly contingent activity with many mutual dependencies between the behavior of the conductor and the ensemble. For example, 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. Also, a conductor should take into account the extent to which the musicians take up signals.

Interaction in the Virtual Conductor system focuses on the temporal aspects as one of the most interesting areas for initial study, as such aspects have hardly been considered in earlier interactive virtual humans or ambient entertainment systems. If the musicians play too slowly, the Virtual Conductor will conduct a little bit faster than they are playing. When the musicians follow, it will conduct faster yet, until the correct tempo is reached again. When everything goes right, the musicians adapt their tempo to the gestures of the conductor, and the conductor reciprocally adapts its conducting to the tempo of the musicians.

5. DETAILED DESIGN AND IMPLEMENTATION

The global design described in the previous section contains several elements that are worth discussing in depth. Since they cannot all be presented in a single article, we will only highlight a selection in this section. The first three following sections detail how the implementation of the conducting gestures is based upon an analysis of how human conductors do their work. Then we discuss the interactive tempo-correction algorithm that allows the Virtual Conductor

¹See <http://hmi.ewi.utwente.nl/topic/Animation> for more information.

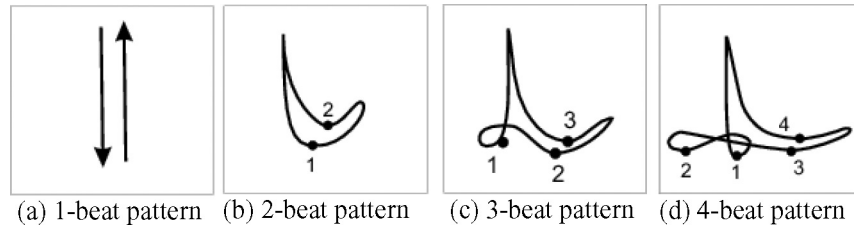


Fig. 4. Four different beat patterns.

to lead the ensemble through tempo changes and corrections. This section concludes with a discussion of the implementation and evaluation of the three audio analysis algorithms incorporated in the Virtual Conductor.

5.1 Literature on Human Conducting

To teach a computer how to conduct an ensemble, we first must learn how a human conductor does their job. We look at literature on conducting and we discuss an interview held with an expert conductor who kindly provided her expertise throughout the project. Most of the information in this section is adapted from the work² of Carse [1935], Prausnitz [1983], and Rudolph [1995]; a good source for more information is the historical overview of conducting handbooks by Galkin [1988].

5.1.1 Basic Conducting Gestures. The basic conducting gesture is the *beat pattern*. The most used beat patterns are the 1, 2, 3, and 4 beat patterns (see Figure 4). A very thorough description of variations of beat patterns used in different styles and cultures, currently and throughout history, is given by Galkin [1988]. Prausnitz [1983] mentions that, for any beat pattern, the preparation (occurring before the actual beat point and even during the upbeat) is more important than the beat itself, as it tells the musicians when the next beat will be and in what tempo. Only a good preparation in the beat pattern allows the conducting of a smooth change in tempo.

5.1.2 Style and Expressiveness. A human conductor can communicate the expression with which she wants the music to be played in many different ways. Left-hand gestures supplementing the beat pattern conducted with the other hand, facial expressions, and the style with which the basic beat pattern is conducted all provide ways to communicate musical expression and phrasing to the ensemble. For example, the beat pattern can be conducted in a very smooth way for legato passages, or in a sharp and angular way for staccato passages.

A few studies look at expressive conducting by human conductors and its impact on musicians. Poggi [2002] analyzed the meaning of different gaze, head, and face movements of a conductor in video recordings and made a start with

²In a later stage of the project, we also looked at video recordings of performances and rehearsals of human conductors with their orchestras. A preliminary report on the lessons learned from these recordings is given elsewhere [Ter Maat et al. 2008].

structuring these meanings as well as the expressions used to denote them in a ‘lexicon of the conductor’s face’. Fuelberth [2004] found that the use of different left-hand shapes in a video recording of a conductor had a strong effect on vocal tension of choral singers asked to sing along with the recorded conducting gestures. Skadsem [1997] looked at different ways of indicating dynamic markings to musicians by letting them sing with a video recording of a conductor, with a choir presented through headphones. Different ways of giving dynamic instructions (verbal, written, gestural instructions, and volume changes in the recorded choir) were clearly found to impact the singers with different strengths. The conductor (following the Nakra [2000] system, which tracks the movements of a human conductor; cf., Section 3.2) was used to perform an analysis of muscle tension in six human conductors while they were conducting. Several detailed observations have been made about how humans conduct; most correspond to the directions given in conducting handbooks.

5.2 Interview with a Human Conductor

When the first prototype of the Virtual Conductor system was ready, it was shown to our expert. This prototype was able to conduct the basic beat patterns and to change the tempo of the beat patterns. During an interview, the expert phrased several requirements for a Virtual Conductor on two dimensions: the movements displayed by the Virtual Conductor and the way in which the conductor should interactively correct the tempo when the ensemble plays too slow or too fast.

5.2.1 *Conducting Movements.* The basic (starting) pose of a conductor is with the arms slightly spread, and slightly forward. The shoulders should not be used to conduct, unless they are necessary for expressive movements. The hands should never stop moving in a conducting gesture. The conducting movements should be as fluid as possible. Every single beat consists of a preparation and the moment of the beat itself. The preparation is what tells the musicians when the beat will be, and its timing is therefore more important than the timing of the beat itself. A conductor can conduct with only the right hand. If the left hand is not being used it can go to a resting position, which is upper-arm vertically, lower-arm horizontally resting against the body of the conductor. If the size of the movements changes, the movements should be placed higher, closer to the face of the conductor. If the conductor wants to indicate pianissimo or even softer, the conducting movements may be indicated only with wrist or finger movements. The right-hand movements can be slightly bigger than the left-hand movements because the left hand has only a supportive function, but the downward movements should end at the same point for both hands.

5.2.2 *Following and Leading the Ensemble.* If the ensemble starts to deviate in tempo or lose precision, a conductor should conduct more clearly and bigger. The conductor should draw the attention of musicians, by leaning forward and conducting more towards the musicians as well. If the musicians play well, the conductor can choose to conduct only with one hand, so she can conduct with

two hands only when more attention from the musicians is required. Snapping fingers or tapping a baton on a stand can work to draw attention, but should be used sparingly, or the musicians will grow too accustomed to this.

To correct the tempo of musicians, a conductor should first follow the musicians, then lead them back to the correct tempo.³ Care should be taken that enough time is spent on following the musicians, or they will not respond to the tempo correction in time and the conductor will no longer be synchronized with the musicians. Just changing the conducted tempo will not work to correct musicians. The musicians should be prepared beforehand that the tempo will change. A conductor should change the preparation of a beat to the new tempo, then change the conducted tempo after that beat. This should preferably be done on the first beat of a bar. Care should be taken to keep each separate bar as constant as possible.

5.3 Implementation of Basic Conducting

For the implementation of the first interactive Virtual Conductor, we chose to work on the four basic beat patterns, dynamic (volume-related) gestures, well-prepared tempo changes, and accents. Unusual or irregular beat patterns, expressive styles, additional left-hand gestures, gaze behavior, and facial expressions have been left for future work (Ter Maat et al. [2008] discuss some preliminary work on, among other things, additional left-hand gestures).

5.3.1 Conducting Gestures. The four basic beat patterns should be well-formed, without undesired accentuation. The patterns, and the different beats in the patterns, should be clearly identifiable. The gestures should be adaptable in amplitude (to indicate accents and different dynamic levels), and in tempo and timing (to execute well-prepared tempo changes and corrections). For implementation of the beat patterns, we chose to use the HMI animation framework. This framework supports parameterized inverse kinematics, allowing us to define an animation as a 3D path in space that should be traced by the hand of the Virtual Human. This mechanism allows almost as much precision as motion-capture methods, but is more amenable to adaptive parameterization. Figure 5 shows how this mechanism is used to define smooth basic conducting gestures. The beat points especially need a precise specification of the path and the deceleration/acceleration in order to obtain clear and concise conducting gestures.

5.3.2 Planning Conducting Behavior. The conducting behavior during a performance of the Virtual Conductor is based on a combination of two sources of information. The first is the score of the music as it should be played; the second is the analysis of the music as the ensemble is actually playing. Information about the score is obtained from MIDI files, and includes the meter (beats per bar), the notes of the different voices, indications of gradual as well as sudden volume, and tempo changes and accented notes. The motion planning

³Another way of getting musicians to play faster is to conduct in the same tempo, but to conduct a beat slightly before the musicians play it. The musicians will instantly know they are playing too slowly and will try to adjust. The conductor can now just follow this and the tempo will be corrected.



Fig. 5. The Virtual Conductor tracing a 3-beat pattern.

used in the Virtual Conductor is very simple. When a new bar starts, the next movement (beat pattern) is loaded. The timing of the movement is adapted to allow for prepared tempo changes. When a tempo change is indicated in the score, the *preparation* of the beat where the new tempo starts is the point where adaptation of the beat pattern starts. The analysis of the music audio, discussed in more detail below, results in information about tempo and volume (deviations) of the music as it is being played by the ensemble. When unplanned tempo changes occur, the beat pattern as it was planned may need to be adapted in timing and in amplitude (cf., next section). The animation engine allows for both gradual and sudden changes in the parameters of a planned animation that is already being executed.

5.4 Implementation of Tempo Correction

Clearly, if an ensemble is playing too slowly or too quickly, a Virtual Conductor should lead them back to the correct tempo. It 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 its conducting, or else lose control. On the other hand, a conductor who simply conducts in the tempo that the musicians are playing in at any particular time is not fulfilling its role either.

When there is a mismatch between the tempo indicated in the score and the tempo at which the ensemble is playing, the Virtual Conductor needs to conduct a tempo correction that brings the ensemble back to the intended tempo without making the musicians lose track of the conductor. During the development of the Virtual Conductor, we experimented with different solutions to this problem. The first approach was to conduct at a tempo t_c between the intended tempo t_i and detected tempo t_d , as expressed by Eq. (1), in which λ defines how strict or lenient the conductor is. Under the assumption that the musicians would change their tempo t_d to be closer to the conducting tempo t_c , this would lead

the ensemble back to the correct tempo.

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

Informal testing with several individual musicians and feedback from the interview with the expert showed that this algorithm might work, but would be rather inflexible and unresponsive. An improved version of the algorithm was defined by making λ change during the course of a tempo correction. When the ensemble deviated too much in tempo, the Virtual Conductor would first follow their tempo in order to keep a grip on the ensemble, then lead them back to the tempo; see Eq. (2).

$$t_c = \lambda_a(b) t_i + (1 - \lambda_a(b)) t_d \quad (2)$$

$$\lambda_a(b) = \begin{cases} \left(1 - \frac{b}{b_{\max}}\right) \lambda_{\min} + \frac{b}{b_{\max}} \lambda_{\max} & b < b_{\max} \\ \lambda_{\max} & b \geq b_{\max} \end{cases} \quad (3)$$

In Eq. (2), λ_a is initially very low (meaning that the conductor mostly follows the musicians) and then becomes higher with b , the number of beats since the start of the tempo correction. This means that during the tempo correction the conductor gradually starts to conduct more strictly, to lead the ensemble back to the correct tempo. In effect, $\lambda_a(b)$ changes linearly from its minimum value λ_{\min} to its maximum value λ_{\max} over b_{\max} beats; see Eq. (3). Because tempo corrections initiated halfway a bar are very confusing to musicians, tempo corrections (changes to λ_a) are only allowed at the start of a new bar. Also, the new conducting tempo t_c , defined by Eqs. (2) and (3), is prepared in the same way as ordinary tempo changes. Tests with musicians showed that this correction algorithm could effectively be used to correct the tempo of musicians, bringing them back to a stable tempo over the course of several bars.

5.5 Music Analysis

The interactive tempo correction described above, as well as any interactive conducting behavior still to be developed, can only function if there is information available about the actual music as it is performed by the musicians. For this, we implemented several audio algorithms, of which the beat detector and the score follower are summarized below.

5.5.1 The Beat Detector. The beat detector is a partial implementation of the beat detector of Klapuri et al. [2006]. This beat detector consists of four stages: an accentuation detector, a periodicity detector, a period selector, and a phase selector. Of the original algorithm, the accentuation and periodicity detector have been directly implemented, augmented with a simplified period and phase-selection algorithm. The precise details of our algorithm have been reported by Bos [2007]; here we just note that the algorithm, among other things, makes use of the fact that in the context of the Virtual Conductor, we can expect that the tempo to be detected is relatively close to the tempo currently conducted. The beat detector was evaluated with the collection from the ISMIR beat-detector contest [Gouyon et al. 2006]. It turned out that this

implementation, as expected, performs worse than the complete version of Klauri's algorithm, but still comparable to state-of-the-art algorithms.

5.5.2 The Score Follower. The real-time online score follower implemented for the Virtual Conductor is a combination of several properties of state-of-the-art algorithms described in Section 3.3.2. We use the online time-warping algorithm of Dixon [2005] on the chroma vector features suggested by Dannenberg and Hu [2003]. The chroma vector of a musical signal is a 12-element vector representing the energy present in the audio signal for each of the 12 tones in Western music (C, C#, D, ..., A#, B). The chroma vectors as calculated by Dannenberg and Hu [2003] suffer from one problem: the resolution of the Fourier transform used to calculate them is linear, whereas the musical scale is logarithmic. This means that for low notes there is too little detail, while for high notes there is far too much detail. We solved this problem by replacing the FFT with the Constant Q transform [Brown 1991], which was developed specifically to address this issue. This way we obtained a chroma vector with a constant quality over all octaves, providing better detail and less noise and improving the overall quality of our feature vectors. The score follower was presented with polyphonic music of differing degrees of complexity: from chamber music with four instruments to a full symphony orchestra with choir and soloists. Informal evaluation of the results suggested that our implementation performed adequately for our purposes [Bos 2007].

6. EVALUATION

The Virtual Conductor has conducted several ad hoc music ensembles for demonstration purposes. Also, besides there having been a lot of informal testing and try-out sessions, two main evaluations with the Virtual Conductor and an ensemble were carried out. An early prototype of the conductor was tested in several sessions with one group of four musicians. The final version of the conductor was tested in two sessions with two different groups of about eight musicians. Full details on all evaluations are reported by Bos [2007]. Here we just summarize the main results.

6.1 Evaluation Setup

Both evaluations consisted of several different sessions aimed at finding out about different aspects of ensemble playing with a Virtual Conductor. First, the ensemble was asked to play several pieces of similar difficulty several times with and without the conductor. This was to draw some general conclusions about the difference between playing with and without a conductor. Second, the ensemble was given a simple piece to learn by heart, after which they were conducted playing this piece, with unknown (for the musicians) dynamics and tempo markings. This was to find out how well the musicians could follow the instructions of the conductor when they could not read those instructions in the music at the same time. Finally, a set of sessions was conducted with the aim of finding out how well the conductor can correct tempo mistakes of the ensemble. This was done in two ways. In one version, a single musician was

instructed to try to lead the rest of the ensemble in playing slightly too fast or too slow, without the other musicians being aware of this instruction. In the second version, we attempted to cause a slowdown by having a piece of music that is rather easy, with one difficult passage. Such sudden difficult passages are often a cause for slowdown in an ensemble. The evaluation sessions were concluded with debriefs during which the musicians filled out a questionnaire with a number of multiple choice and open questions.

6.2 Evaluation Results

We should note that it is near to impossible to draw any kind of quantitative conclusions from the small number of evaluations that we performed. Such conducting sessions as we performed could by no means qualify as “repeated sessions under comparable conditions”. However, we will discuss some qualitative and tentative conclusions here.

6.2.1 *Evaluation One: Tempo Interaction.* Evaluation one was with the first, naive, version of the tempo-correction algorithm described in Section 5.4. The most important conclusion to be drawn from this evaluation is that this tempo-correction algorithm did not work very well. The conductor reacted too quickly and too often on tempo deviations, sometimes even showing several different reactions within one bar. Also, the beat patterns were seen as something that should be improved. They were not clear and not smooth enough, and the dynamic markings were hard to pick up for the ensemble. The musicians were nevertheless more or less able to follow the conductor in a few sessions, and even picked up the “unexpected tempo changes” a few times.

6.2.2 *Evaluation Two: Tempo Interaction.* As a direct result of the first evaluation, the final version of the tempo-correction algorithm described in Section 5.4 was implemented and the beat pattern animations were improved in cooperation with our expert. The second evaluation was carried out using the final version of the Virtual Conductor. The difference with the first evaluation was clearly visible. Both ensembles used in this evaluation were often able to follow the tempo changes and dynamic changes indicated by the Virtual Conductor, even when these were not notated in their sheet music. Also, we could observe several successful attempts by the conductor to correct the tempo of the musicians in situations where the ensemble in the first evaluation always broke down. Too big and unexpected tempo changes still resulted in a breakdown of the performance though.

The improved beat pattern animations and the new tempo-correction algorithm clearly worked much better than the first version. The musicians could reliably play with the conductor with very little practice and could follow the dynamics and tempo changes indicated by the conductor. The tempo-correction algorithm worked this time, resulting in several instances where a faulty tempo of the ensemble was successfully corrected.

6.2.3 *Feedback from the Musicians and Conductors.* Both evaluation sessions were concluded with debriefs during which the musicians gave feedback

about the system. The system has also been shown to our informant conductor. Finally, there have been informal demonstration and testing sessions with ensembles where the conductors of the ensembles, who watched the sessions, commented on the system afterwards. The reactions were predominantly positive, with the most reservations being expressed about the possibility of introducing the *artistic* dimension of conducting as an automatic component in the system. The latter suggests that attempting to develop the system further into a *performing* conductor that leads actual concerts may only make sense for its novelty value.

On the other hand, the musicians clearly liked playing under the direction of the Virtual Conductor, indicating that they could envision the system as a rehearsal aid for practicing with an ensemble that either does not have a (human) conductor, or for practicing more than just the times during which the human conductor is available. Some human conductors remarked that this system could be used to train the technical parts of a piece with an ensemble, or with subgroups of the ensemble, leaving the human conductor more time to work on the artistic aspects of a piece.

An unexpected element of feedback came from the human conductors who watched the sessions of our Virtual Conductor. They were very interested in the possibility to apply controlled variations to the conducting behavior of the system to see what the impact of such variations would be on the behavior of the ensemble, both for correct conducting behavior (for example, a well-executed tempo change) and for incorrect conducting behavior (for example, a sudden change in tempo that was not prepared correctly). It turns out that there is not much literature on such controlled experiments; but see Fuelberth [2004] and Skadsem [1997]. The Virtual Conductor could be used in this way to perform experimental investigations into the interaction between conductors and their ensembles, leading to a better understanding of the conducting process.

7. CONCLUSION

A Virtual Conductor has been researched, designed, and implemented that can conduct human musicians in a live performance. The conductor can lead musicians through tempo, dynamics, and meter changes. During test sessions, the musicians reacted to the gestures of the conductor. It can interact with musicians in a basic way, correcting their tempo gracefully when they start playing faster or slower than is intended. Feedback from the musicians who participated in the tests shows that the musicians enjoyed playing with the virtual conductor and could see many uses for it, for example as a rehearsal conductor when a human one is not available, or for playing along with when practicing at home.

Several audio algorithms have been implemented for tracking the music as played by the ensemble. Among those, the beat detector can track the tempo of musicians and the score follower can track where musicians are in a score in real-time. The possibilities of these audio algorithms reach further than what is currently used for feedback in the Virtual Conductor, and will be very useful for future extensions of the system.

While this article presents the results of the first stages of the Virtual Conductor project, the system is still being extended in several directions. We have been working on a detailed analysis of a corpus of performances and rehearsals, which we recorded on video, in order to find out more about the intentions and signals used in conducting and about the structure of rehearsals. The repertoire of expressions that the Virtual Conductor can use is continuously being extended. We are working on implementing rehearsal modules that allow the Virtual Conductor to plan a structured series of rehearsal sessions, adapting later sessions in response to the progress achieved in earlier ones. Also, the planning and scheduling modules used to produce the conducting animations have been extensively redesigned in order to allow a much more flexible and responsive behavior generation that can handle several expressive channels at the same time. This will allow us to combine the basic beat patterns with an extensive repertoire of left-hand gestures, facial expressions, gaze behavior, and other nonverbal expressions.

In this article, several possible and seemingly viable applications of the system were discussed, given the system as it has been developed up to now, and the feedback of musicians and conductors on the result. The system can, for example, be used as a rehearsal aid in ensembles, as a tool to enrich the education of student conductors, and as a platform for controlled experimentation with the interaction between conducting behavior and an ensemble.

8. FINAL THOUGHTS: COORDINATED TIMING AND ENTERTAINMENT

A human conductor interacts with the musicians in many modalities on many different dimensions. For all of these dimensions, a prerequisite is that the conductor and the musicians are actually performing on the same time line. Therefore, we chose the coordinated timing in the interaction as one of the focal points for the interaction between Virtual Conductor and human musicians. Note that this coordinated timing is different from the more usual timing-related interaction in games such as *Donkey Konga*, where the player must closely follow the rhythm presented by the system, but the system does not adapt its own timing in direct feedback response.

Thinking about the issues related to the development of such coordinated interaction—and the interactions between system and musicians that we observed—inspired us to look at coordinated interaction in a more theoretically motivated way. It turns out to be very similar to the concept of *nonverbal synchrony*. According to the literature, there seems to be a strong relation between synchrony and positive affect and positive evaluation of the interaction partner [Crown 1991; Ramseyer and Tschacher 2008; Nagaoka et al. 2007; Suzuki et al. 2003; Bailenson and Yee 2005; Robins et al. 2005]. Does that mean that we can harness coordinated interaction to improve the enjoyment and engagement of the user? And if so, what steps do we need to take to achieve this? In the Virtual Conductor system, the kind of synchrony that can be achieved is mediated through the music, conducted by the conductor and played by the ensemble. The behavior of the users is rhythmically constrained by the score, and the temporal aspects of the users' behavior can be modeled using a beat

tracker and a score follower. For other applications this is clearly much more difficult. What are the beat patterns in gaming? Can they be modeled computationally? Even if the temporal dynamics of gaming can be modeled, this is not enough to develop nonverbal synchrony-like interaction. Our experience with implementing the Virtual Conductor clearly showed that the system cannot adapt in a purely reactive way. Does this hold for other applications as well? Is the only way to achieve mutually coordinated interaction in entertainment applications to have the system not only detect the temporal patterns of the user's past behavior, but also *predict* the temporal patterns of the short-term future behavior of the user, as in the Virtual Conductor system?

These ideas and questions will certainly guide our choices in future developments of the Virtual Conductor system, as well as other interactive entertainment applications. We will therefore conclude this discussion with a short presentation of some related literature, showing that the combination of synchrony and entertainment is relevant and is starting to be addressed by a few other people as well. For example, Weinberg and Driscoll [2006] built a robotic drummer that interacts with human musicians on rhythmic dimensions. Their focus is on the musical expressiveness and the collaborative music-making. Michalowski et al. [2007] built a robot that dances rhythmically in response to movements of the user, synchronizing to his or her rhythm. Tanaka and Suzuki [2004] explicitly modeled two-way entrainment for their dancing robot, Qrio, in order to achieve more engagement from the user. A core concept in their interaction model is the repetition of *sympathy* and *variation to it*: "We consider that there are two concepts that should be basic units designing the interaction model. One is sympathy between the two (human and robot), and the other is a variation to it. Our basic hypothesis is that by the repetition of the sympathy and variation, long-term interaction can be realized." Finally, Tomida et al. [2007] attempt to achieve "entertaining interaction" between two humans by trying to elicit entrainment quite directly. In their MiXer installation, the authors aim for implicit synchronization between two human users of the system: the "rhythms" of a user are sensed using bio sensors (heart rate, perspiration, etc.) and the rhythms are presented to the other user using "bio feedback" (visualization of those rhythms). The other user may then tap a button synchronized to the biofeedback display; the assumption is that an engaging type of "entrainment" between the users will occur in response to this process.

Although entertainment through entrainment and coordinated interaction were not a goal in themselves during the initial development of our Virtual Conductor system, it is nevertheless clear that there is a strong relation. Ideas about coordinated interaction are starting to find a place in research on interactive entertainment and the techniques that we need to develop in order to achieve the proper coordinated interaction between Virtual Conductor and musicians may become highly relevant to the kind of entertainment applications described above.

ACKNOWLEDGMENTS

Daphne Wassink provided us with her much-needed expertise as a real conductor. Musicians from the Musica Silvestra Orchestra participated in various

rounds of testing. The conductors of both the Musica Silvestra Orchestra of the University of Twente and the CHN Orchestra of the Stenden Hogeschool graciously allowed our Virtual Conductor to step into their place for real performances. Rik Bos provided our Virtual Conductor with proper attire for evening performances. To all of them, our thanks.

REFERENCES

- ALONSO, M., DAVID, B., AND RICHARD, G. 2004. Tempo and beat estimation of musical signals. In *Proceedings International Conference on Music Information Retrieval*. 158–163.
- BAILENSEN, J. AND YEE, N. 2005. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science* 16, 10 (Oct.), 814–819.
- BOS, P. 2007. Interacting with a virtual conductor. M. S. thesis, Human Media Interaction, University of Twente, NL.
- BROWN, J. 1991. Calculation of a constant q spectral transform. *J. Acoustical Society of America* 89, 1, 425–434.
- CARSE, A. 1935. *Orchestral Conducting*. Augener Ltd.
- CROWN, C. L. 1991. Coordinated interpersonal timing of vision and voice as a function of interpersonal attraction. *J. Language Social Psychology* 10, 1, 29–46.
- DANNENBERG, R. AND HU, N. 2003. Polyphonic audio matching for score following and intelligent audio editors. In *Proceedings of the Interactional Computer Music Conference*. International Computer Music Association, 27–33.
- DIPAOLA, S. AND ARYA, A. 2006. Emotional remapping of music to facial animation. In *Proceedings of the ACM SIGGRAPH Symposium on Videogames* (Sandbox '06). ACM, New York, 143–149.
- DIXON, S. 2003. On the analysis of musical expression in audio signals. In *Storage and Retrieval for Media Databases*, SPIE vol. 5021. 122–132.
- DIXON, S. 2005. Live tracking of musical performances using on-line time warping. In *Proceedings of the 8th International Conference on Digital Audio Effects*. 92–97.
- FUELBERTH, R. J. V. 2004. The effect of various left-hand conducting gestures on perceptions of anticipated vocal tension in singers. *Int. J. Research Choral Singing* 2, 1 (Jan.), 27–38.
- GALKIN, E. W. 1988. *A History of Orchestral Conducting: In Theory and Practice*. Stuyvesant, New York.
- GOTO, M. 2001. An audio-based real-time beat tracking system for music with or without drum-sounds. *J. New Music Research* 3, 2 (June), 159–171.
- GOUYON, F. AND DIXON, S. 2005. A review of automatic rhythm description systems. *Computer Music J.* 29, 1 (Feb.), 34–54.
- GOUYON, F., KLAPURI, A., DIXON, S., ALONSO, M., TZANETAKIS, G., UHLE, C., AND CANO, P. 2006. An experimental comparison of audio tempo induction algorithms. *IEEE Trans. Speech Audio Processing* 14, 5 (Sept.), 1832–1844.
- ILMONEN, T. AND TAKALA, T. 1999. Conductor following with artificial neural networks. In *Proceedings of the International Computer Music Conference (ICMC, Beijing)*, 367–370.
- JANSEN, L., VAN DIJK, B., AND RETRA, J. 2006. Musical multimodal child computer interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children (IDC)*. ACM, New York, 163–164.
- KLAPURI, A., ERONEN, A., AND ASTOLA, J. 2006. Analysis of the meter of acoustic musical signals. *IEEE Trans. Speech Audio Processing* 14, 1 (Jan.), 342–355.
- KOLESNIK, P. AND WANDERLEY, M. 2004. Recognition, analysis and performance with expressive conducting gestures. In *Proceedings of the International Computer Music Conference (ICMC)*.
- LEE, E., GRÜLL, I., KIEL, H., AND BORCHERS, J. 2006. CONGA: A framework for adaptive conducting gesture analysis. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME, IRCAM, Centre Pompidou, Paris)*, 260–265.
- LEE, M., GARNETT, G., AND WESSEL, D. 1992. An adaptive conductor follower. In *Proceedings of the International Computer Music Conference*. International Computer Music Association, 454–455.
- MANCINI, M., BRESIN, R., AND PELACHAUD, C. 2007. A virtual head driven by music expressivity. *IEEE Trans. Audio, Speech, and Language Processing* 15, 6, 1833–1841.

- MICHALOWSKI, M. P., SABANOVIC, S., AND KOZIMA, H. 2007. A dancing robot for rhythmic social interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, ACM, New York, 89–96.
- MURPHY, D., ANDERSEN, T. H., AND JENSEN, K. 2003. Conducting audio files via computer vision. In *Proceedings of the GW03*. 529–540.
- NAGAOKA, C., KOMORI, M., AND YOSHIKAWA, S. 2007. Embodied synchrony in conversation. In *Conversational Informatics: An Engineering Approach*, L. Nishida (ed.), Wiley, 331–351.
- NAKRA, T. M. 2000. Inside the conductor's jacket. Ph.D. dissertation, MIT Media Laboratory, Cambridge, MA.
- NIJHOLT, A., REIDSMA, D., VAN WELBERGEN, H., OP DEN AKKER, H., AND RUTTKAY, Z. M. 2008. Mutually coordinated anticipatory multimodal interaction. In *Proceedings of the COST*. Springer, Berlin.
- PARDO, B. AND BIRMINGHAM, W. P. 2005. Modeling form for on-line following of musical performances. In *AAAI*. 1018–1023.
- POGGI, I. 2002. The lexicon of the conductor's face. In *Language, Vision and Music*. John Benjamins, 271–284.
- PRAUSNITZ, F. 1983. *Score and Podium: A Complete Guide to Conducting*. W. W. Norton, New York.
- RAMSEYER, F. AND TSCHACHER, W. 2008. Synchrony in dyadic psychotherapy sessions. In *Simultaneity: Temporal Structures and Observer Perspectives*. World Scientific, Singapore, Ch. 18.
- RAPHAEL, C. 2004. A hybrid graphical model for aligning polyphonic audio with musical scores. In *Proceedings of the International Conference on Music Information Retrieval*.
- REIDSMA, D., VAN WELBERGEN, H., POPPE, R., BOS, P., AND NIJHOLT, A. 2006. Towards bidirectional dancing interaction. In *Lecture Notes in Computer Science* 4161, R. Harper et al. (eds.) Springer, Berlin, 1–12.
- ROBINS, B., DAUTENHAHN, K., NEHANIV, C. L., MIRZA, N. A., FRANCOIS, D., AND OLSSON, L. 2005. Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. In *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*.
- RUDOLPH, M. 1995. *The Grammar of Conducting: A Comprehensive Guide to Baton Technique and Interpretation* (3rd ed.), Schirmer.
- RUTTKAY, Z. M., HUANG, A., AND ELIËNS, A. 2003. The conductor: Gestures for embodied agents with logic programming. In *Proceedings of the 2nd Hungarian Computer Graphics Conference*. 9–16.
- SCHEIRER, E. D. 1998. Tempo and beat analysis of acoustic musical signals. *J. Acoustical Society America* 103, 1 (Jan.), 558–601.
- SCHÖGLER, B. 1998. Music as a tool in communications research. *Nordic J. Music Therapy* 7, 1, 40–49.
- SCHWARZ, D., ORIO, N., AND SCHNELL, N. 2004. Robust polyphonic midi score following with hidden markov models. In *Proceedings of the International Computer Music Conference (ICMC)*.
- SEGEN, J. 1999. Visual conductor. In *Proceedings of the ACM SIGGRAPH '99 Conference: Conference abstracts and applications*. ACM, New York, 180.
- SEPPÄNEN, J., ERONEN, A., AND HIIPAKKA, J. 2006. Joint beat and tatum tracking from music signals. In *Proceedings of the 7th International Conference on Music Information Retrieval*, R. Dannenberg et al. (eds.), 23–28.
- SHIRATORI, T., NAKAZAWA, A., AND IKEUCHI, K. 2006. Dancing-to-music character animation. *Computer Graphics Forum* 25, 3 (Sept.), 449–458.
- SKADSEM, J. A. 1997. Effect of conductor verbalization, dynamic markings, conductor gesture, and choir dynamic level on singers' dynamic responses. *J. Research in Music Education* 45, 4, 509–520.
- SUZUKI, N., TAKEUCHI, Y., ISHII, K., AND OKADA, M. 2003. Effects of echoic mimicry using hummed sounds on human-computer interaction. *Speech Communication* 40, 4, 559–573.
- TANAKA, H. AND SUZUKI, H. 2004. Dance interaction with qrio: A case study for non-boring interaction by using an entrainment ensemble model. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN)*, IEEE, 419–424.
- TER MAAT, M., EBBERS, R. M., REIDSMA, D., AND NIJHOLT, A. 2008. Beyond the beat: Modeling intentions in a virtual conductor. In *Proceedings of the INTETAIN*. ACM, New York.

- TOMIDA, T., ISHIHARA, A., UEKI, A., TOMARI, Y., FUKUSHIMA, K., AND INAKAGE, M. 2007. MiXer: The communication entertainment content by using “entrainment phenomenon” and “bio-feedback”. In *Advances in Computer Entertainment Technology*, M. Inakage et al. (eds.), ACM, New York, 286–287.
- WANG, T.-S., ZHENG, N.-N., LI, Y., XU, Y.-Q., AND SHUM, H.-Y. 2003. Learning kernel-based HMMs for dynamic sequence synthesis. *Graphical Models* 65, 4, 206–221.
- WEINBERG, G. AND DRISCOLL, S. 2006. Robot-human interaction with an anthropomorphic percussionist. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, New York, 1229–1232.
- YIN, J., WANG, Y., AND HSU, D. 2005. Digital violin tutor: An integrated system for beginning violin learners. In *Proceedings of the 13th Annual ACM International Conference on Multimedia*. ACM, New York, 976–985.

Received March 2008; revised ; accepted May 2008