

Beyond the Beat: Modelling Intentions in a Virtual Conductor

Mark ter Maat

Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, the Netherlands
+31 53 489 3740
mark.termaat@gmail.com

Dennis Reidsma

Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, the Netherlands
+31 53 489 3740
dennissr@ewi.utwente.nl

Rob M. Ebbers

Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, the Netherlands
+31 53 489 3740
rob.ebbers@gmail.com

Anton Nijholt

Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, the Netherlands
+31 53 489 3740
anijholt@ewi.utwente.nl

ABSTRACT

We describe our research on designing and implementing a Virtual Conductor. That is, a virtual human (embodied agent) that acts like a human conductor in its interaction with a real, human orchestra. We reported previously on a first version that used a digital musical score to lead an orchestra. This conductor was able to conduct the beat with a certain tempo and dynamics, and to correct the tempo if necessary, using advanced audio analysis. We observed this Virtual Conductor at work during various performances for which he was invited. These performances made us aware of shortcomings. Therefore we took a closer look at the interaction between conductors and musicians in practice, both during performances and during rehearsals, and based on this study we introduced conducting gestures that display the intentions of a conductor and developed rehearsal modules. Apart from the literature on conducting we took into account videos of human conductors and interviewed human conductors. In addition we introduced principles from conversational analysis in the new design of our Virtual Conductor.

Categories and Subject Descriptors: H5.1 [Information Interfaces and Presentation]: Multimedia Information Systems - *Animations, Artificial, augmented, and virtual realities, Audio input/output*; H5.2 [Information Interfaces and Presentation]: User Interfaces; H5.5 [Information Interfaces and Presentation]: Sound and Music Computing - *Methodologies and techniques, Modeling, Signal analysis, synthesis, and processing, Systems*

General Terms: Design, Experimentation

Keywords: Intelligent agents, virtual humans, audio analysis, entertainment computing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. The Second International Conference on Intelligent Technologies for Interactive Entertainment (ICST INTETAIN '08). January 8–10, 2008, Cancun, Mexico. Copyright 2008 ICST. ISBN 978-963-9799-13-4.

1. INTRODUCTION

A conductor is a person who directs a musical performance by way of visible gestures. The musicians in the ensemble or orchestra respond to these gestures while playing, giving the conductor control over the live performance in terms of tempo, dynamics, phrasing and other dimensions of musical expression. A conductor is also in charge of the rehearsals that lead to a successful concert performance.

Apart from the technical skills required for conductors, conducting is also considered an art. There are no complete and absolute rules on how to conduct correctly, and it is difficult to say whether one conductor is better than another in the small and exclusive group of top conductors who travel all around the world to conduct major symphony orchestras.

Several systems have been created that are related to conducting and automating part of the conducting process. These include systems that synthesize conducting gestures, systems that try to follow a human conductor and automatic accompaniment systems for musicians.

We have created a virtual embodied agent that can conduct musicians in a live performance. Using a score and sound input from a microphone, this Virtual Conductor can interact with the musicians, both leading and following their tempo while they are playing music (Bos et al. [3]).

The first Virtual Conductor (see Figure 1) had some successful performances, but as may be expected, it also became clear that there is still a big distance to go. A human conductor does not just indicate the beat and correct the musicians when they play too fast or too slow, but also indicates through conducting gestures his intentions about the play style of the orchestra or particular instruments, e.g. by raising his left hand or pointing at a group of instruments. In order to have successful performances it is also necessary to have successful rehearsals. For that reason it became clear that we also should look at the role of a Virtual Conductor during rehearsals. How can we make a rehearsal successful? Clearly, a rehearsal allows the orchestra to get acquainted with the intentions a conductor has and with the way he expresses them in his conducting behavior.

In this paper, an overview of the latest work on the Virtual Conductor is presented. In particular we look at our research on intention signalling and rehearsal requirements. The results of this research provided us with guidelines for a modular and flexible design of a new version of the Virtual Conductor which was implemented and evaluated.

In Section 2 we have a few more words about the original Virtual Conductor and its characteristics. The need for a more sophisticated version is discussed and translated into research goals. The sources we used to study conductor behavior are mentioned. Section 3 and Section 4 are about modelling the interaction between conductor and orchestra. A Virtual Conductor needs to be provided with a conducting schedule (to be obtained from the score and a conductor's intentions) from which it can be animated. The architecture of the Virtual Conductor should however also allow real-time deviations from such an enriched score based on the playing of the orchestra (e.g., to give corrective feedback) and it should allow for all kinds of conducting behavior related to rehearsals rather than to performances. Section 3 is about intentions, Section 4 is about rehearsals. In both sections we discuss theoretical observations as well as observations obtained from studying video material and holding interviews with conductors. Section 5 discusses the audio sensors needed for the new version, focussing on the new loudness sensor. Design considerations and the actual design and the implementation of the Virtual Conductor based on the observations in Section 3 and 4 are presented in Section 6 of this paper. In the final sections we discuss performed evaluations, future research and conclusions.

2. A VIRTUAL CONDUCTOR PROJECT

The original Virtual Conductor system (Bos [2], Bos et al. [3]) is a virtual human that is able to conduct the beat from a midi-file in the correct tempo and with the correct dynamics. It hears the music played by the musicians, analyses their tempo and tries to correct that tempo if necessary. To do this, first the input of the orchestra is processed and analyzed. This analysis is compared to the original score to determine what to do next. A planner generates the corresponding conducting movements, which are subsequently animated.

The audio processing consists of two modules, a beat and tempo detector and a score follower. The 'beat detector' is based on the work of Scheirer [18] and Klapuri [12]. First, accents in certain frequency bands of the audio are detected, which results in a graph with a peak for every accent. This is then passed to a bank of comb filters to find periodicity in the graph. The output of this is a graph with all possible musical periods. The correct period is the peak with the highest value. The 'score follower' is used to match the original score to the music played by the orchestra to find out how the orchestra progresses through the score. This is done by creating chroma vectors of both the score and the music and matching those two vectors with a dynamic time warping algorithm (Bos [3]).

But knowing that the ensemble is playing too slow or too fast is not enough, the conductor should be able to rectify the problem. To get the orchestra back on track, the conducting tempo t_c is defined as:

$$t_c = (1 - \lambda)t_0 + \lambda t_d$$

where t_0 is the correct tempo and t_d the detected tempo. The ratio λ determines how strict the conductor is. When correcting the tempo of the musicians the conductor starts with a low value of λ and gradually raises this value, until the orchestra is playing in the right tempo again.

Conducting the tempo is done by beating a 1-, 2-, 3- or 4-beat pattern (depending on the score and the tempo). The animations of these patterns are defined using inverse kinematics and hermite splines, where the amplitude of the movement is adjustable to the preferred required dynamics.

While the system was evaluated and works well for what it was designed to do, there is room for improvement. What this conductor does is indicating the beat, the tempo and dynamics. A human conductor does a lot more on several levels. For example, he gives signals to indicate the entrance of a certain section on a difficult spot, he indicates the style in which the musical piece should be played by conducting in a certain way (e.g. jumpy, heavy, etc), and he knows what to pay attention to in a particular orchestra.

For these reasons we researched possible extensions. The aim was to add a repertoire of signals and gestures for communicating intentions concerning e.g. the play style, and to create functionality for rehearsal techniques. To be able to embed these extensions in the system, its architecture had to be modified. The new architecture, discussed in detail in Section 6, basically contains three modules:

- A perception module that is responsible for analyzing the score and the music played by the orchestra to extract useful information such as the beat, the dynamics, etc.

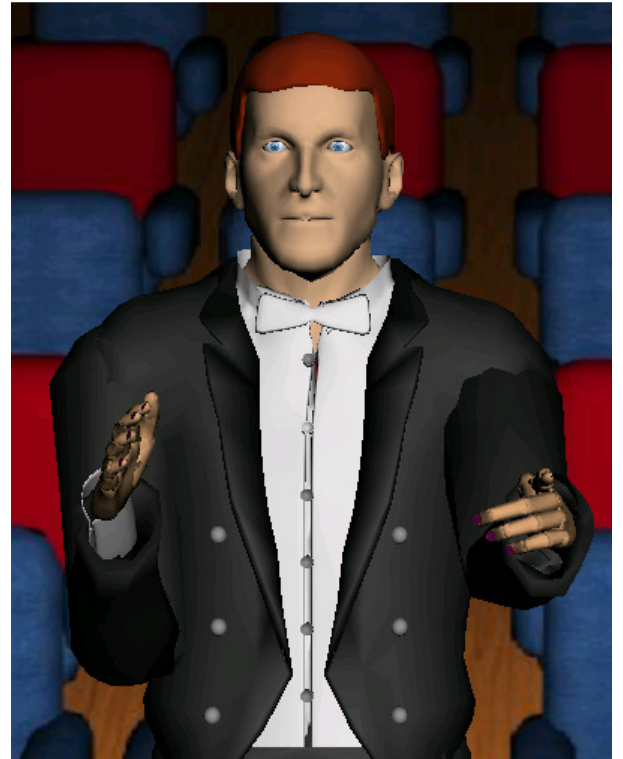


Figure 1: The Virtual Conductor

- A cognition module that, based on this extracted information, generates intentions (determines what the conductor wants), converts these intentions to actual gestures, and puts these gestures in a conducting schedule.
- An animation module, which takes the gestures from the schedule and animates them.

As mentioned, the first version of the conductor only indicates the beat and does not give other signals to the orchestra. To add the possibility of conducting other signals the intentions and corresponding signals of conductors were studied. The important question is: what messages does a conductor want to communicate to the musicians? Intentions and corresponding signals have been collected and have been made part of the cognition module (see Section 3).

Again, as mentioned earlier, it is also natural to expand the current Virtual Conductor with functionality for rehearsing. The Virtual Conductor should be able to lead a rehearsal and provide some kind of useful feedback to the musicians about their playing. For this goal, a study was done on orchestral rehearsals. How do conductors rehearse and what are adequate strategies for rehearsing? As in the case of displaying intentions, the cognition module and the overall architecture of the Virtual Conductor have been designed with the flexibility to incorporate these strategies.

Two important sources of information, besides the literature and our experiences with the first version of the system, were the interviews we held with several human conductors and a set of video recordings of rehearsals and performances we annotated.

Three conductors of amateur orchestras were interviewed and asked about their personal experiences with rehearsals. All three conductors worked with amateur orchestras most of the time, and these interviews were mainly used to get a good overview of intentions and rehearsal structures.

Four videos of human conductors were studied. We made two recordings of an amateur accordion orchestra and an amateur symphony orchestra (both conductors were also interviewed). We also studied recordings of two professional orchestras. One recording was of Carl Maria Giulini rehearsing Bruckner's 9th symphony with the Stuttgart Symphony Orchestra [9], and the other one was a video of Carlos Kleiber conducting the Vienna Philharmonic Orchestra [13].

Apart from the findings that became part of our modules and overall design of the Virtual Conductor, we also concluded that we should add a 'loudness' (or dynamics) sensor to our already existing sensors (among other things, a beat detector and score follower). This newly added sensor will be discussed in Section 5, before we introduce the overall design in Section 6.

3. CONDUCTOR-ORCHESTRA INTERACTION

In this section we investigate the conductor-orchestra interaction. First we compare existing approaches in literature on conversational analysis with observations on conducting behavior. After that we present our findings on intentions, and signals to convey them, as observed in the recorded conducting sessions.

3.1 Conversational Analysis

3.1.1 Continuous Interaction

An interesting point when comparing conductor-orchestra interaction with conversations is that the turn-taking paradigm of Sacks et al. [17] does not apply at all. The communication between the conductor and the orchestra is never turn-based. The conductor cannot wait for the orchestra to finish with something before he acts, but always listens and acts at the same time.

This corresponds with the theory where listeners do not only listen, but participate in the conversation at the same time. For example, O'Connell et al. criticize the turn-taking theory and the studies that seem to support that theory, and instead they suggest a new paradigm of conversation, where 'all the time belongs to all the interlocutors' [15]. The reason for this is that 'communication by eye, gesture, and touch can be continuous and therefore simultaneous on the part of the interlocutors'. Bavelas et al. [1] do something similar by studying the active role of 'listeners', which they do not call listeners but co-narrators.

Another interesting point is that the conductor can only use gestures and facial expressions to make clear what he means. This sounds very limited, but surprisingly an orchestra almost always knows exactly what the conductor means. This may be related to the fact that both parties share a lot of information; both the conductor and the orchestra know what to play (the notes, play style, volume). The conductor's task is to modify the details and keep the orchestra on track.

A last point that we want to stress in this continuous interaction is that when the conductor tries to keep the orchestra in the right tempo, instead of reacting to the past or the present, the conductor adapts his behavior to anticipations of the future. It tries to predict what the orchestra will do the next few seconds and will try to influence this behavior with (corrective) feedback.

It is clear now that the interaction is not as straightforward as it seems, with a lot of interesting aspects to cope with. To investigate the interaction, two theories from the field of conversational analysis were chosen, applied on the conductor-orchestra interaction and studied for interesting features.

3.1.2 Grice's Maxims on Cooperation

The theory of Grice's Maxims [11] defines assumptions of a listener about the speaker, such as the assumption that the speaker's acts are relevant (so if they seem not relevant the speaker must mean something different) and the assumption that the speaker speaks the truth. These assumptions are grouped in four principles: the Principle of Quantity, the Principle of Quality, the Principle of Relevance and the Principle of Manner.

To fit this model on the conductor-orchestra interaction, the orchestra can be seen as the listener while the conductor is the speaker. The orchestra makes assumptions about the conductors communication-acts (its gestures and facial expressions), so obviously these assumptions should fit the maxims of Grice as well.

Looking at conductor-orchestra interaction in the light of Grice's Maxims resulted in a list of recommendations for the design of the Virtual Conductor. The most salient recommendations are:

- According to the principle of quantity the contribution of the speaker should be as informative as required, but

not too informative. To regulate this, signals should receive a priority value to indicate how important it is to display that particular signal. When at a certain time there are too many signals to display, this could result in an overload of information for the musicians and only the signals with the highest priority should be conducted. The priority value also affects overlapping signals. When two signals overlap in time and make use of the same body parts of the Virtual Conductor then the signals should be either rescheduled or the signal with the lowest priority should be replaced by the signal with the highest priority.

- Only act on detected sensor events if the conductor is certain enough about that sensor event. This follows from the maxim ‘Do not say that for which you lack adequate evidence’ from the principle of quality.
- The principle of relevance states that the speaker’s contributions (in this case the conductor’s) should be relevant. This means that when giving a signal, it should be early enough to give the musicians time to react, but late enough to be relevant to the signalled event.
- A target and the meaning of the signal should be clear and unambiguous. This can be explained by the maxim ‘Avoid ambiguity’ from the principle of manner.

3.1.3 Mind Markers

The other theory that was used to analyze the conductor-orchestra interaction comes from Poggi’s work on Mind Markers [16]. This theory states that in a conversation, a speaker sends, next to his normal speech act, signals about his state of mind. These signals can be single signals or gestures, but also signal modifiers. For example, a person can look up to indicate that he is thinking (a single signal), but he can also speak with a falling intonation to indicate that he is sure about what he is saying (modifying the speech signal). The information in the mind markers is about the beliefs, the goals and the emotions of the speaker (Belief Markers, Goal Markers and Emotional Markers) and can help the listener to better understand the speaker.

An orchestra can use the similar meta-information to better understand the conductor (presuming that a conductor actually sends these markers). Therefore we investigate how the Virtual Conductor could send these markers as well.

Belief Markers are not very useful here, because the conductor should conduct everything with certainty, as if everything he conducts is true. So, in the eyes of the musicians, every signal they get should be treated as the truth because the conductor believes it is true.

Goal Markers are more relevant for the Virtual Conductor. Just like in a conversation where goals can be sent on different levels (word level, phrase level, sentence level), with the conductor its signals can have goals on different levels (and the Virtual Conductor should send these goals on the same levels). Poggi describes the Goal Markers on the lowest level as goals of the single signals. At this level, goals can be compared with the communicative intentions of the Virtual Conductor. For example, a speech act (‘sit!’) has a specific goal, and the same holds for a single conducting gesture (which can mean ‘play louder!’). On the next level Poggi’s Goal Markers appear as relational markers and topic-comment markers, which are no longer the same as the

Virtual Conductor’s intentions. Relational markers appear in a signal that tells something about another signal. For example, a pointing gesture can define the target of another signal. Topic-comment markers tell the difference between the signals taken for granted (the topic) and the new and important signals (the comment), where the comments usually are conducted in a more distinct manner. Goal Markers on the two highest levels deal with phrases and the whole composition, and are mostly about the play style of a certain passage or the piece. When the conductor wants to send a goal marker on this level, usually these markers appear in all of his signals of the corresponding passage. For example, when a certain phrase should be played very light, then the conductor will conduct that phrase very light.

The last type of marker is the Emotional Marker. The conductor has two types of emotions: emotions he wants to show to indicate a certain play style (e.g. look angry when the music should be played aggressive), and emotions he really has (e.g. looking angry because the orchestra is playing the wrong notes). Information about his emotion is transmitted through Emotional Markers in everything he does. This way the orchestra knows if they should play aggressive or if the conductor is really angry at them. Because the focus of the Virtual Conductor is currently on the more technical aspects of conducting, emotion and personality will not yet be used in the system.

The above study leads to a list of concrete recommendations for the design of the Virtual Conductor. The most important recommendations are:

- When conducting signals that are taken for granted (the topic, for example the beat pattern when the musicians easily play the right tempo), these signals should be conducted less distinct and notable. When a new and important signal appears (the comment, for example an entrance cue), this signal should be conducted very clearly. In contrast with the (less clear) beat pattern the cue will appear more pronounced, drawing the attention of the musicians to that signal.
- To send belief markers, the Virtual Conductor should show confidence. This is done by letting the Virtual Conductor take a confident pose and by creating animations that show confidence. To do this, these animations should be clear, powerful and without hesitation.
- From the analysis of relational markers it was found that signals can be linked, where one signal would tell something about another signal. An example of this is indicating a forte (‘play loud’) with the left hand and look at a certain section at the same time. The gaze defines a target for the forte, thereby linking the signals. In the design this should also be possible.

3.2 Intention Analysis in a Video Corpus

While in the previous section we could fall back on a long tradition of linguistics and conversational analysis, hardly anything is known, despite existing handbooks, about the way conductors express their intentions. For that reason we studied the video material and conducted interviews (cf. Section 2). For this study, we used the two videos we recorded, plus the video of Carlos Kleiber. From these three videos, a total of six fragments were cut: three short fragments of about two minutes from Kleiber



Figure 2: Some snapshots from videos used in the intention analysis

and three fragments of about 5 minutes from the videotaped conductors.

These fragments were then annotated using Elan [8] with the different signals given by the conductor. Different tiers were used to indicate what body part the conductor used for a certain signal: right hand/baton, left hand, eyes, head, body, and facial expression. For every signal, the following data was written down: the description of the signal, the intention with which the conductor performed that signal and the body part that was used to conduct the signal.

A total of 49 different intentions were found in the study. In Figure 2, three of them are shown. The most common intentions are ‘Accent’, ‘This is for you’ (to indicate that the current behavior is for the target musicians), ‘Play soft’, and ‘Play to climax or accent’ (this indicates a build-up in the music towards a local climax or accent). These intentions were seen 59, 53, 34 and 29 times. Some intentions were only seen once or twice in all videos, for example ‘Play dreamy’ and ‘This signal is for the orchestra (and not for soloist)’.

The main goal of the annotation was to retrieve a list of most intentions plus possible corresponding signals, not to collect all possible signals. Therefore we counted for each annotation round how many new intentions and signals were found that were not yet encountered in the earlier rounds. Table 1 lists the number of new intentions and new signals for each annotation round. It can be seen that the number of new intentions found after the first annotation round is very low, while the number of new signals stays high. This suggests that the number of intentions used by any conductor is limited, while the number of signals is huge and varies from person to person. This also suggested that we would not find many new intentions by annotating yet more fragments.

Table 1: The number of additional intentions and signals found in the annotated fragments in each annotation round.

Video	# Intentions	# Signals
1: Fragment 1, conductor 1	32	63
2: Fragment 2, conductor 1	9	34
3: Fragment 1, conductor 2	2	40
4: Fragments of Kleiber	6	52

The annotation exercise resulted in a list of intentions for a conductor, together with a list of possible signals for each intention. Very striking was the difference between a conductor of an amateur orchestra and a conductor of a professional orchestra. The conductor of the professional orchestra almost never strikes the beat because the orchestra is good enough to keep the correct tempo by themselves. Most of his signals are meant to indicate the play style. The conductor of the amateur orchestra has to

continually strike the beat or else the orchestra will speed up or slow down. Because it is easier to start a conductor that strikes the beat continually with the necessary signals instead of a conductor that only strikes the play style, and because it is easier to get a group of amateur musicians for testing, the system will be created for an amateur orchestra.

4. REHEARSING

The goal of a rehearsal is to improve the performance of a particular piece of music by the orchestra. The conductor leads the rehearsal according to a certain rehearsal strategy. In the first subsection we study the existing literature on rehearsing. In the second subsection we study our videotapes and interviews again, this time looking for information about rehearsing.

4.1 The Rehearsal Process in Literature

In the literature on conducting techniques very little seems to have been written about the process of rehearsing. The focus of most of the literature seems to be on baton technique, analyzing, preparing and interpreting scores, and technical knowledge about the orchestra and its individual instruments.

Although the rehearsal strategies of famous conductors are not widely documented, some rehearsal characteristics and anecdotes can be found in individual biographies of conductors and historical overviews of conducting and conductors. These conductors usually work with first-class orchestras that ought to have no or very little problems with the technical difficulties of playing orchestral music, such as playing the correct notes at the right time. With amateur orchestras, the target for our Virtual Conductor, the technical side of playing music is much more important, and during rehearsals a much larger portion of the available rehearsal time is devoted to these aspects.

However, despite the apparent lack of literature about rehearsing, we were able to find some information about rehearsing and orchestral rehearsals.

For a first rehearsal of a work with any orchestra, Bowen [4] and Dolmetsch [7] suggest a complete run-through, however rough, stopping only in the case of absolute train-wrecks. The conductor should keep mental notes of problem areas, and return to them after the first pass is complete.

Carse [5] gives suggestions for working with orchestras consisting of beginning players that can’t be expected to be able to play through a work the first time they play it together. When trying to correct their errors, the order in which these faults should be addressed is suggested as follows:

1. Get the time right, i.e., play correct note-values.
2. Identify and correct wrong notes.
3. Correct the worst of the faulty intonation.
4. Work on the ensemble (unanimous movement of parts, attack, etc.).

Bowen [4] also remarks on other behavior of the conductor during rehearsals: make jokes, self-depreciation never hurts, be clear, speak loudly, lead and govern with your eyes, be direct, a simple and eloquent baton is more than adequate. This author also suggests using section rehearsals (sectionals), even in first-class orchestras, because they allow close and detailed attention to every note in every section of the orchestra. Another suggestion made is using “rehearsal-only” tempi: take fast passages slowly, and vice versa. Using “rehearsal-only” dynamics, playing loud

passages very quietly, is beneficial for articulation and voice-leading and is far less exhausting for the musicians' hearing, arms and embouchure. It can also reveal problems obscured by mere volume.

In the design of the new rehearsal capabilities of the Virtual Conductor we will focus on the suggestions about the technical side of playing music, i.e. when to interrupt the playing and the priority with which certain errors should be corrected. The suggestions about the social behavior of the conductor are less useful to the Virtual Conductor at this stage, because it lacks the functionality to exhibit this behavior.

4.2 Rehearsal Study in a Video Corpus

The three interviews and videos of rehearsals mentioned in Section 2 were studied for relevant information about rehearsing. In the videos, a huge difference between the rehearsals of an amateur orchestra and a professional orchestra can be observed. The typical rehearsal of an amateur orchestra is, especially in the beginning of the rehearsal process, mainly focussed on purely technical matters like playing the right notes at the right time. The conductors rehearsing with professional orchestras mainly focus on their personal interpretation of the piece they are playing. This fits with the distinction already noticed in the literature study.

Studying the videos resulted in several observations about rehearsals and rehearsing. What follows is the list of observations that are most relevant to the Virtual Conductor's rehearsal functionality and that have guided us in our design.

- When rehearsing a new piece with an ensemble, the best starting strategy is to just start playing it from the beginning and then try to play through the whole piece, stopping only when the ensemble is falling apart and several musicians completely have lost track of the music. The new version of the Virtual Conductor must therefore be able to conduct more lenient than the previous version, sometimes following rather than leading the musicians when they have trouble keeping the tempo. In case the ensemble falls apart, the new Virtual Conductor must be able to interrupt the playing.
- When rehearsing a piece that has been rehearsed before, stop the playing when a difficult passage was incorrectly played. Repeat that passage again until it is played correctly. Then go back a little before the difficult passage and start playing from there again. This interrupting-repeating cycle is typical of the process of rehearsing and needs to be incorporated in a Virtual Conductor with rehearsal capabilities.
- Technically difficult passages can be rehearsed more slowly than the prescribed tempo. That way, the passage is not just easier to play technically, but it is also easier to play all the parts together and in sync with each other. Also, the sensors of the Virtual Conductor are able to give more detailed information about the playing in a slower tempo.
- When the musicians make different types of errors at the same time, which is very likely to happen, it is usually best to start correcting wrong notes and rhythms before correcting dynamics and out-of-tune playing.
- A conductor must be aware of possible false entries, i.e., musicians starting to play their parts too early or too late, especially in polyphonic passages. Musicians are not always aware of these mistakes themselves because of all the voices

playing simultaneously. For the Virtual Conductor, this would mean a major extension of the existing score follower.

- Musicians sometimes do not play notes for their full duration, especially when playing long notes or notes at the end of a phrase. They stop playing the note too early because they often are more focused on the beginning of a next note than on the end. The Virtual Conductor could potentially be aware of this and correct the musician.
- It can be very beneficial to rehearse certain difficult passages with only a section instead of the whole orchestra, although from a pedagogical point of view it is not desirable to let individual players play a passage alone. A section rehearsal can also be regarded as a separate rehearsal, but with a smaller number of players than a regular rehearsal. No separate functionality for section rehearsals will be incorporated in the Virtual Conductor.
- The conductor can stop playing to tell something to the orchestra, but he can also say things while the playing is still in progress, e.g. when something minor went wrong on which the conductor doesn't want to spend time repeating it. In the case of the Virtual Conductor, spoken feedback might cause problems with the sound analysis as it is also being picked up by the microphone.
- The interviewed conductors both remarked that during rehearsals, they sometimes give unexpected signals (e.g. tempo wise) to check and improve the attention of the musicians. However, until the Virtual Conductor has been thoroughly tested, it is undesirable to introduce such elements of uncertainty that could confuse the players and, as a result of this, the Virtual Conductor.

5. MUSIC ANALYSIS AND DIAGNOSTICS

In the previous sections it has become clear that in a design where we want to take into account intentions and rehearsing functionality, we need analysis (how are they playing) and diagnostics (how does it compare to the ideal situation) of several aspects of the playing of the orchestra.

The previous version of the Virtual Conductor already had capabilities of assessing the notes that are played and their tempo. These capabilities proved to be accurate enough for now, so we decided to reuse them in the new Virtual Conductor.

Intonation, or determining exact pitch, was not a part of the original Virtual Conductor and we decided not to address this problem as it is very hard to determine exact pitches in music with more than one voice. However, it could be a research goal for a future version of the Virtual Conductor.

The first version of the Virtual Conductor did not have functionality for detecting loudness. For a proper assessment of the dynamics that are played by the orchestra, the Virtual Conductor needs a sophisticated measurement of loudness. This section reports our findings developing such a module.

Several approaches for measuring loudness were examined, implemented in MATLAB and tested. Based on the results from these tests, one of these approaches of calculating loudness was chosen and used in the Virtual Conductor.

Loudness is a subjective measure, and true perceived loudness varies from person to person. However, several psychoacoustic

models exist that can be used to calculate perceived loudness. These models ideally don't just take into account the physical intensity, but also spectral and temporal effects, just like a human ear would do.

Skovenborg and Nielsen compared twelve different models of loudness perception for long-term loudness with durations of 10-15 seconds per sound segment [19]. Langner et al. describe three different loudness models to calculate three loudness curves which are used in note onset detection [14]. Loudness curves are time-loudness graphs which are obtained by continually calculating the loudness of the sound signal over a short period, e.g. 50 ms.

Because these two sources seem to disagree on which loudness model is the best one to use for measuring the loudness of music, some testing was done to determine which loudness model is the best model for calculating loudness in the Virtual Conductor.

Three different loudness models were considered, which are, from simple to complex:

1. Calculating sound pressure level (SPL). (Unweighted equivalent continuous sound level $L_{eq}(\text{Lin})$; Skovenborg and Nielsen [19]). Although correlation between this method of measuring loudness and perceived loudness is generally considered weak, this time-efficient way of calculating loudness might be good enough for our purposes.
2. Calculating the frequency weighted equivalent continuous sound level, $L_{eq}(W)$, as described by Skovenborg and Nielsen [19]. Several different frequency weightings W exist. We used their A-weighting, which is widely adopted in environmental noise measurements (see e.g. Zwicker et al. [21]).
3. Using Zwicker's model as described in [21] and specified in the DIN45631/ISO-532B standard to calculate the loudness in sound. This is a multi-band loudness model which divides the frequency range of the signal in 25 critical bands of $1/3^{\text{rd}}$ octave to model the human auditory system more realistically than single-band models.

Since dynamics in music are relative and not absolute, and dynamic markings such as *piano* or *forte* do not correspond to fixed loudness values, we are not interested in correct absolute loudness values, but in loudness values that can correctly

distinguish between soft and loud playing. The $L_{eq}(\text{Lin})$ and $L_{eq}(A)$ values were not converted to the logarithmic dB-scale for easier comparison to the values of Zwicker's model.

The three loudness models were implemented in MATLAB and tested with several fragments of orchestral and choral music, using both overlapping and non-overlapping time periods of 50 ms. Figure 3 shows a graphical comparison of the loudness curves of ten bars of a recording of the first movement of Mozart's 40th symphony, which features 6 bars of *piano*, 2 bars of *crescendo* to *forte* and finally 2 bars of *forte*.

From these tests it was concluded that for now, simply calculating $L_{eq}(\text{Lin})$ using non-overlapping time periods of 50 ms is sufficient for our needs. The difference between using overlapping or non-overlapping time periods is small, because the loudness doesn't fluctuate much in the short time period of 50 ms.

The resulting loudness module has been made part of the architecture of the new Virtual Conductor. It calculates the sound pressure level for every 50 ms. The module has to be calibrated every time before an actual rehearsal or performance starts because of differences in the composition of the ensemble, acoustics and recording levels. This is done by letting the ensemble play three short fragments: one *pianissimo*, one *mezzo piano/mezzo forte*, and one *fortissimo*. The loudness values obtained from these calibrations can then be used for proper assessment of the dynamical playing of the ensemble during the rehearsal of performance.

6. DESIGN OF THE VIRTUAL CONDUCTOR

Gratch et al [10] give an extensive overview of what is needed to build a virtual human and they lift out three key areas. The first is the area of emotions and personality, which is about the impact of infusing behavior with emotions. However, as explained before, the Virtual Conductor's current focus is on the technical aspects of conducting, and adding emotions and personality have not been considered yet. The second area is that of human figure animation. The animation of the conductor is done by the inverse kinematics player we have previously developed (Welbergen et al. [20]), which uses functions as animations to calculate the positions and rotations of the body parts of the virtual agent at a certain time. The last area is that of face-to-face communication and is exactly what the Virtual Conductor is about: getting input and deciding what to do or say next. Gratch et al. give a list of features such a module must have and by way of example describe the architecture of the BEAT-toolkit [6]. This toolkit annotates input, generates possible behaviors, uses some filters to select certain behavior and finally performs this behavior.

This architecture inspired the architecture of the Virtual Conductor (see Figure 4), which contains three major parts, i.e. perception, cognition and animation. Obviously, the first part analyzes the input, the second part creates the behavior and puts this in some kind of conducting schedule, which is then executed by the animator.

6.1 Perception

The perception module basically is a set of interlinked sensors which all have a different function. These functions can be filters, such as a Fast Fourier Transformation filter, or data analysis modules, such as a beat detector to find the tempo of the music

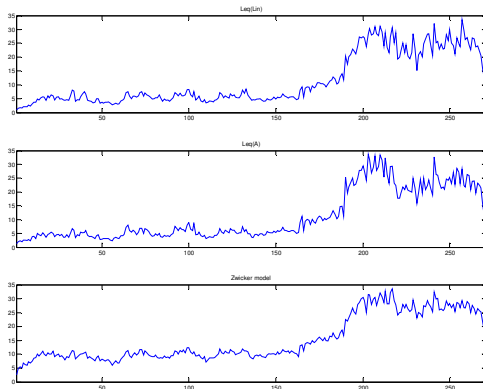


Figure 3: Graphical comparison of three loudness models

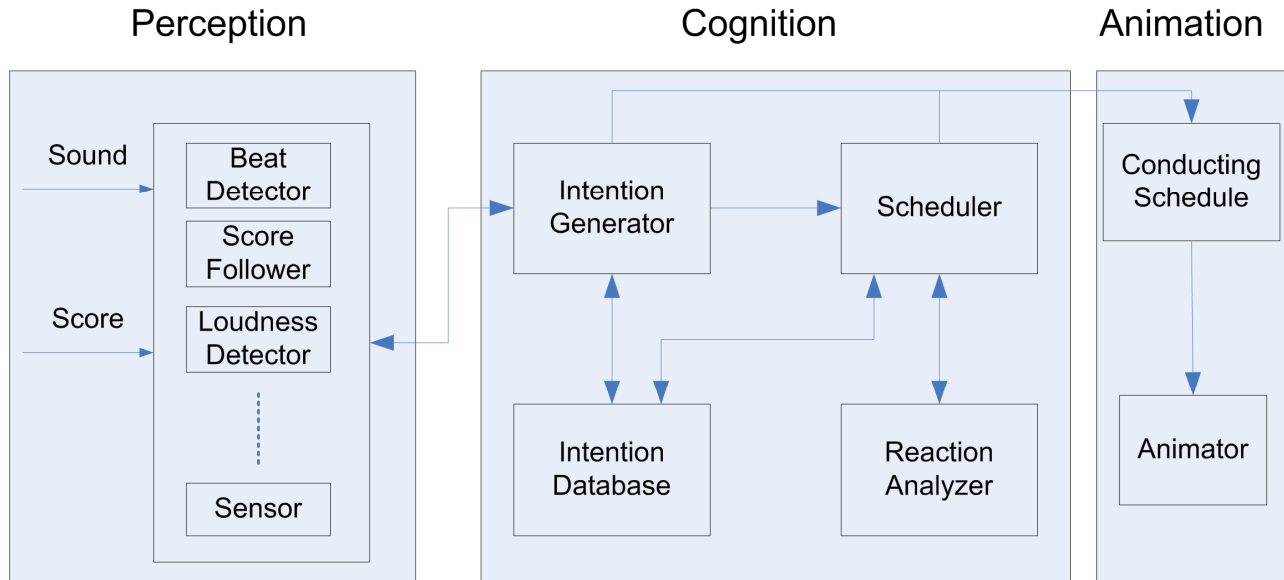


Figure 4: Global architecture of the Virtual Conductor

played by the orchestra. Filters are mainly used to modify the data; the data analysis modules provide the useful data for the rest of the system. The previous data analysis modules were the Beat Detector, a music signal presence detector, a Chord Detector (which was created but not actively used in the first version of the Virtual Conductor), a score follower and a midi file reader. More information about these modules can be found in [2]. In addition to these modules, a new loudness detector was implemented based on the research described in Section 5.

All these sensor modules are used to retrieve information about the played music. This information can then be related to the original score and the rehearsal process, where the following events can be detected:

- Wrong notes can be detected by comparing the data from the score follower and the chord detector.
- The timing and correct lengths of notes can be checked by using the data from the score follower and the chord detector. Sloppy ensemble playing can also be detected this way.
- Using the data from loudness module, the dynamics of the played music is evaluated: the correct playing of dynamic markings such as *piano* or *forte*, and execution of crescendos and decrescendos.
- The beat detector is used to check the tempo and detect a gradual slowing down or speeding up by the orchestra.

6.2 Cognition

The appropriate reactions to the sensor information coming in from the music analysis modules are determined by the cognition module. In the original system the cognition module was very straightforward. Halfway every bar the current tempo of the orchestra was checked and compared with the required tempo. Based on this information the new tempo was selected and the beat pattern of the next bar, with the new tempo, and an amplitude based on the required dynamics, was sent to the animator.

The task of the new cognition module (see Figure 4) is to continuously update and fill a schedule of conducting behavior by

evaluating the orchestra's current behavior, trying to predict how the orchestra will behave in the near future and create a conducting plan to fit that behavior.

The Intention Generator consists of several modules that continuously analyze the current behavior of the orchestra and generate intentions for the conductor. For example, the tempo analyzer checks the current tempo of the musicians and the required tempo from the score, and modifies the schedule to get the musicians back on the right tempo. Other examples are the loudness analyzer, the mistakes checker (which tries to identify mistakes made by the orchestra) and the reaction analyzer (which analyzes how the orchestra reacts to given signals). The Intention Generator can modify the schedule directly or delegate this to the Scheduler by sending it the intention.

Another important part of the Intention Generator is the rehearsal module which creates intentions that try to correct errors. When trying this, there is a priority in the order in which these errors should be corrected; wrong notes and bad timing have more priority than badly played tempos or wrong dynamics. The exact functions of this module are all based on the findings in Section 4.

This module can also decide to stop the playing when too much goes wrong and a certain threshold of errors has been reached. It is not always desirable to immediately stop the playing when something wrong has been detected, i.e., when an orchestra plays a piece for the first time, it is best not to stop at all, unless the ensemble completely falls apart.

When the Virtual Conductor stops the musicians, it must tell them what went wrong and where it went wrong. After that, it gives a bar number, somewhere before the place of the mistake from where to start playing again. This feedback is presented as text on the screen. When repeated rehearsing of a particular passage doesn't help, the Virtual Conductor can decide to conduct that passage in a slower tempo. The module keeps track of the phase the rehearsal is in and what went wrong on previous passages.

The Scheduler takes the intention as an input, uses an Intention Database to find all possible signals to execute the intention and

tries to fit one of these signals in the schedule. If necessary the schedule has to be rearranged to make room for every signal. This rescheduling is based on the priority of the signals. This priority (based on the study of Grice's Maxims) indicates how important it is to conduct that particular signal. Before a signal is actually added to the schedule, it is first sent to the Reaction Analyzer.

This module knows how the orchestra reacts on certain signals and can approve or disapprove certain signals (because they tend to work or not to work). Also they can adapt the signal itself, for example by raising the amplitude to make it more visible (the current orchestra might need that with that type of signals).

The intention database is a simple mapping between the intentions of the Virtual Conductor and the possible corresponding signals. This database also contains extra information about these intentions and signals. The database will eventually be filled with the intentions and signals found in section 3.2, but currently only several of them are implemented, namely 'entrance cue', 'louder', 'softer', 'ask attention', 'look around' and 'co-conduct'.

When running the Virtual Conductor, the Cognition module is used twice. First, the complete score is read and processed offline, intentions are generated and the schedule is filled with the information needed to conduct the complete piece. Second, when the musicians start playing, the cognition module is used to process the data online. It then becomes a reactive system which updates the schedule real-time using the input from the musicians.

6.3 Animation

In the end, everything comes together in the Conducting Schedule. This schedule is like an agenda of what the conductor has to do in the future. The cognition module continually modifies the schedule by adding, removing and rearranging the signals of the conductor. At the same time, a thread continuously reads the schedule to find out whether a new signal should start and sends this to the animation package if necessary.

There are two ways to modify the conducting schedule. The first is to directly add, change or remove a signal. The other option is to send an intention to the scheduler which tries to find the best fitting signal and will then change the conducting schedule.

Signals in the schedule contain a reference to their intention and the animation that belonging to it. They are also tied to an abstract time. These abstract times are meant to easily synchronize multiple animations by attaching them to the same abstract time. In a separate data structure the real times that correspond to the abstract times are stored. When the tempo changes, these times are changed accordingly, without modifying the signals directly.

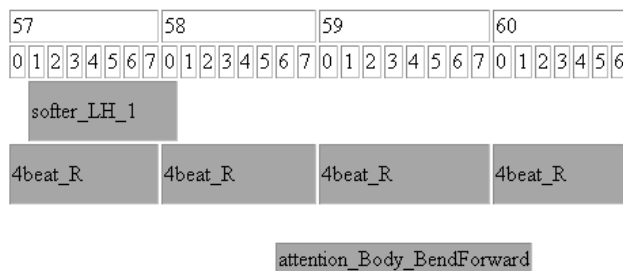


Figure 5: A fragment of a conducting schedule

A fragment of an actual conducting schedule can be found in Figure 5. In this fragment a 4-beat pattern is continuously conducted with the right hand. In bar 57 the conductor signals a 'softer' with his left hand and halfway bar 58 the conductor bends a bit forward to ask for attention (in this particular fragment the tempo was increasing).

When a signal is ready to be animated it is sent to the animator. This animator uses inverse kinematics to play the animations [20]. Each animation consists of a set of functions for all positions and rotations of the hands, the feet, the head and all joints of the Virtual Conductor. The inverse kinematics player uses these functions to calculate at any time the exact pose of the conductor. Because the schedule can contain multiple animations at the same time, extra functionality was added to squash multiple animations together into one consistent animation.

7. EVALUATION

To evaluate the performance of the improved Virtual Conductor, we have tested it in practice. Some tests have been carried out already, while some tests will be conducted in the near future.

The evaluation of the new intentions/signal-system was done with three different tests. The first was used to test the new system as a whole by letting four musicians play with the conductor. In the second test the conductor was demonstrated to a human conductor for professional feedback. During the third evaluation a total of nine musicians played with the Virtual Conductor. The goal of this experiment was to test if the extra signals of the conductor with his left hand and his head really helped the players.

The results of these evaluations were encouraging. All musicians had to get used to the Virtual Conductor, but the longer they played the better they understood its gestures. Sudden tempo changes were a problem because the new tempo was not indicated in the upbeat prior to the new tempo, but sudden loudness changes were picked up very fast when the conductor indicated this with his left hand. Another problem was that the musicians had difficulty following the beat pattern of the conductor when it conducted with very small hand signals. Nevertheless, all in all the new version of the Virtual Conductor could more robustly lead an orchestra than the previous version, while providing more information through its conducting gestures.

For testing the rehearsal functionality, we plan to organize at least one fully staged rehearsal with a small orchestra in which the musicians go through a rehearsal of a piece of music they have never played before, led completely by the Virtual Conductor.

We plan to have two different groups of musicians, each group playing some music they have never played before. To determine whether the Virtual Conductor has any effect on the improvement of the performance of the musicians, one group will rehearse some music with the Virtual Conductor, while the other group will rehearse without and vice versa.

After this rehearsal, the musicians who participated will be asked about their experiences. We are interested in their opinions about the realism of the signals the Virtual Conductor gives them and its rehearsal capabilities. Do they think the Virtual Conductor is a useful tool for rehearsing music with an orchestra?

Also, some real conductors will be invited to be present at the rehearsal and will be asked to comment about the performance of their virtual colleague. Special attention will be given to their

opinions about the rehearsal strategy the Virtual Conductor uses. Do they find it realistic? What would they do different and why?

The reactions of the musicians during the rehearsal and the opinions of both the musicians and conductors after the rehearsal will be analysed and used to improve the Virtual Conductor.

8. CONCLUSION

In this paper we described our research on conductor-orchestra interaction. The interaction was compared with theories from the field of conversational agents and videos of real conductors were studied for rehearsal techniques, intentions and corresponding gestures and facial expressions. Using this information, a model for rehearsing was created and a list of intentions (what does a conductor want) was produced with every intention having a list of possible signals, i.e., ways of showing this intention.

Then the architecture of the Virtual Conductor was extended with a model for rehearsing, making it capable of planning and executing a rehearsal session, and to make it possible to fill a conducting schedule with intentions to be executed by the virtual agent. Finally, the new architecture was implemented and an evaluation method was designed. Some of the evaluations have already been performed; the full evaluations will be reported on more extensively when all evaluations are finished.

The new Virtual Conductor can do a lot more than the first version. It can conduct a lot more signals, it contains a conducting schedule in which it plans its moves, and it can easily be extended with a lot more intentions and signals. Also it is almost capable of leading a rehearsal in which it can efficiently rehearse a musical piece with musicians. Currently it is still mainly focused on the technical aspect of conducting, such as keeping the tempo and the dynamics correct and helping musicians with entrances, but the basics of a conductor are laid. In the future it might also deal with more abstract features like the correct balance of the orchestra, the tone colours, the expression of the music, and the personal interpretation of the conductor. But as we said, the fundamentals are present and the Virtual Conductor has a lot of potential for more improvement building on these fundamentals.

9. ACKNOWLEDGEMENTS

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

10. REFERENCES

- [1] Bavelas, J. B., Johnson, T., and Coates, L. Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6), 941–952, 2000.
- [2] Bos, P. *Interacting with a Virtual Conductor*. MSc Thesis, Human Media Interaction, University of Twente, 2007.
- [3] Bos, P., Reidsma, D., Ruttkay, Z., and Nijholt, A. Interacting with a Virtual Conductor. *Proc. of 5th International Conference on Entertainment Computing*, Cambridge, UK, 25–30, 2006.
- [4] Bowen, J. (ed.). *The Cambridge Companion to Conducting*. Cambridge University Press, 2003.
- [5] Carse, A. *Orchestral conducting*. Augener, London, 1935.
- [6] Cassell, J., Vilhjalmsson, H., and T. Bickmore, T. BEAT: The Behavior Expression Animation Toolkit. In *Proc. SIGGRAPH '01*, ACM Press, New York, 477–486, 2001.
- [7] Dolmetsch, R. *The Art of Orchestral Conducting*. Bosworth & Co, London, 1942.
- [8] ELAN, *Eudico Linguistic Annotator*. Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, <http://www.lat-mpi.eu/tools/elan/>
- [9] Giuliani, C.M., Radio-Sinfonieorchester Stuttgart. Anton Bruckner: Symphony No. 9 in D minor, Arthaus-Musik DVD 101 065, recorded 1997.
- [10] Gratch, J., Rickel, J., André, E., Cassel, J., Petajan, E., and Badler, N.I. Creating interactive virtual humans: some assembly required. *IEEE Intelligent Systems*, 17(4), 54–63, 2002.
- [11] Grice, H.P. Logic and conversation. *Syntax and Semantics*, 3: *Speech Acts*, 41–58, 1975.
- [12] Klapuri, A., Eronen, A., and Astola, J. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Speech and Audio Processing*, 2006.
- [13] Kleiber, C., Vienna Philharmonic Orchestra. Johannes Brahms: Symphony No. 2 in D Major, Philips DVD 0701619, recorded 1991.
- [14] Langner, J., Kopeiz, R., and Feiten, B. Perception and representation of multiple tempo hierarchies in musical performance and composition: Perspectives from a new theoretical approach. In *Controlling creative processes in music*. Frankfurt am Main: Peter Lang, 1998.
- [15] O'Connell, D. C., Kowal, S., and Kaltenbacher, E. Turn-taking: a critical analysis of the research tradition. *Journal of Psycholinguistic Research*, 19(6), 345–373, 1990.
- [16] Poggi, I. Mind markers. *Gestures. Meaning and use*, 2002.
- [17] Sacks, H., Schegloff, E. A., and Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735, 1974.
- [18] Scheirer, E.E. Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, 103(1), 558–601, 1998.
- [19] Skovenborg, E. and S. Nielsen. Evaluation of different loudness models with music and speech material. In *Proceedings of the 117th AES Convention, San Francisco*, 2004.
- [20] Welbergen, H. van, A. Nijholt, D. Reidsma, and J. Zwiers, Presenting in Virtual Worlds: Towards an Architecture for a 3D Presenter explaining 2D-Presented Information. *IEEE Intelligent Systems*, 21(5), 47–53, 2006.
- [21] Zwicker E. and H. Fastl. *Psychoacoustics*. Berlin: Springer, 1990.