Improving Orchestral Conducting Systems in Public Spaces: Examining the Temporal Characteristics and Conceptual Models of Conducting Gestures

Eric Lee, Marius Wolf, Jan Borchers
Media Computing Group
RWTH Aachen University
52056 Aachen, Germany
{eric, wolf, borchers}@cs.rwth-aachen.de

ABSTRACT
Designing interactive conducting exhibits for public spaces poses unique challenges, primarily because the conceptual model of conducting music varies amongst users. In a user study, we compared how conductors and non-conductors place their beats when conducting to a fixed orchestral recording of *Radetzky March*, and found significant differences between these two groups. Conductors lead the actual music beat with their gestures by an average of 150 ms, compared to 50 ms for non-conductors; non-conductors also vary their placement of the beat 50% more than conductors. Furthermore, we found differences in how users conceptually mapped their gestures to the music, such as conducting to the musical rhythm rather than to the beat. We are incorporating these results into an upcoming conducting system for public spaces to increase its usability; we believe they also apply to a more general class of musical gestures such as dance.

ACM Classification Keywords
H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – augmented reality, audio output, evaluation/methodology; H.5.2 [Information Interfaces and Presentation]: User Interfaces – evaluation/methodology, user-centered design.

Author Keywords
conducting; gestures; music interfaces; exhibits; empirical study; conceptual models.

INTRODUCTION
Gesture-based interaction techniques are an increasingly popular part of current research in human computer interaction [21]. Gesture interaction has been shown in popular movies such as *Minority Report* [6], and has also begun to appear in mainstream commercial software such as Lionhead's role-playing game *Black & White* [16] and Apple's motion graphics application *Motion* [1]. Gesture-based interaction techniques are especially promising for multimedia: conducting and dance, for example, predate computers for gestural interaction with music.

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.

Conducting as an interaction metaphor for computer music has received much attention in recent years: Mathew's *Radio Baton* [18], Morita et al.'s conducting system [19], Marrin's *Conductor's Jacket* [17], and our own previous system *You're the Conductor* [13] are just some examples. As part of our research in this area, we aim to build an *adaptive conducting system* that adjusts to the user's conducting ability; this system would react precisely to a professional conductor's baton gestures, but still be forgiving to the potentially erratic gestures of an untrained conductor. Such a system would offer increased usability in a public space, such as a museum, where people with varying conducting skills will use it. It would also be a useful training tool for student conductors to help them navigate the learning curve, providing the equivalent of training wheels for bicycles. However, to build such a system we first need a solid understanding of conducting, in particular how to systematically evaluate the amount of conducting training a user has, the various conceptual models of conducting that may differ from user to user, and the factors influencing these models.

While qualitatively evaluating our previous conducting systems for public spaces [3, 13], we observed a variety of usability breakdowns, which we believe to be a result of differing conceptual models. For example, we observed some users conducting to the musical *rhythm* (musical pattern formed by the dominant melody/percussion) rather than to the *beats* (consistently spaced intervals to count time); since these systems change the tempo in response to beats, conducting to the rhythm results in erratic tempo changes, confusing the user. We also frequently observed the "spiral of death", where users, in response to a slowdown of musical tempo, slowed down their conducting, which caused a further slowdown of the music tempo, and so on. We hypothesized this phenomenon to be a result of the user conducting to or behind the beat (as if playing an instrument along with the orchestra), rather than ahead of it as conductors are taught to do. Conductors, on the other hand, frequently complained that their control of the orchestra was not as "tight" as with a real orchestra.

These types of usability breakdowns motivated us to study more carefully the temporal relationship between users' conducting gestures and the beat of a musical piece; for example, while conductors are taught to conduct ahead of the beat, do non-conductors naturally conduct behind it? Does musical ability, such as expertise playing an instrument, affect this temporal relationship between gesture and music beats? We will show how the results from our study can be used to improve the usability of current conducting systems and to design the adaptive conducting system described above.

RELATED WORK

While there is a large body of research on conducting systems, most of these systems are designed for interpreting movements of either professional conductors [11, 14, 19, 20] or non-conductors [2, 3, 13], but not both.

Harrer performed a series of studies with the famous German conductor Herbert von Karajan in the 1970's [10], where he measured the reaction of Karajan and one of his students to music. He measured and recorded their ECG (electrocardiogram), breathing, and GSR (galvanic skin response). The discussion of his findings is brief: both Karajan and his student produced similar readings that could be traced to the structure of the music. There is no analysis beyond these readings, nor did Harrer collect readings from, or compare them with, any other people.

Morita et al. created a system that follows a human conductor using a CCD (charge-coupled device) camera and sensor glove [19]. They measured a conductor's movements, qualitatively analyzed the position, velocity, and acceleration of his movements, and mapped these parameters to music tempo and dynamics. They did not analyze movements from non-conductors, and their analysis was limited to spatial characteristics of the gestures.

Usa and Mochida discussed various aspects of conducting, including beat timing, in the presentation of their *Multimodal Conducting Simulator* [24]. According to their findings, how much a conductor leads the beat with their gestures depends on their expertise and cultural background. They experimentally determined that Japanese conductors feel "satisfied" leading the beat by 100 ms for music with a tempo of 50 bpm (beats per minute) and 0 ms for a tempo of 110 bpm. They did not elaborate on these results, nor did they include non-conductors in their analysis.

Marrin compared data from student and professional conductors measured using her *Conductor's Jacket* [17]. This data includes measurements of muscle tension and respiration. She was primarily interested in mapping expressive features to sections in the music score, rather than obtaining measurements on how movements map to rhythm and beats.

Research on *beat induction* aims to computationally model the cognitive task of tapping to the beat while listening to music. While there is a large body of current research in computer music and music psychology on this topic [7, 22], they do not examine conducting specifically, where the aim is to guide the beat in addition to finding it. Moreover, for our work, we are more interested in *where* people place the beat than *how* they find it.

Thus, our work is unique in the following ways:

- it compares professionally trained conductors to non-conductors
- it analyzes the temporal characteristics of conducting gestures (placement and timing of the beats) as opposed to their spatial characteristics (shape, velocity, acceleration)
- it provides quantitative results in addition to a qualitative analysis
- it examines users' conceptual models of conducting (how they mentally map gestures to music tempo)

STUDY SCOPE AND OBJECTIVES

For this work, we had the following objectives:

1. determine a set of parameters distilled from conducting gestures that can be used to distinguish between conductors and non-conductors, and can possibly be used to determine to what degree the user is a trained conductor
2. quantitatively measure where conductors and non-conductors place their perception of the beat relative to the actual beat of the music
3. qualitatively understand what factor(s) effect where users place their beats for a given piece (e.g., familiarity with the music piece, musical ability, etc.)
4. better understand both conductors' and non-conductors' conceptual models of conducting

Based on preliminary interviews, we determined that conducting gestures vary widely from conductor to conductor. We observed a similar situation with non-conductors using our systems. Therefore, a study of the spatial properties of conducting gestures (e.g., shape, velocity, acceleration) would not have helped us meet our objectives. Moreover, we received one comment from a conductor who claimed that professional conductors "probably have very consistent timing of the beat points". Thus, we chose to examine the *temporal* properties of users' gestures, such as how the timing of the beat points is related to the music beats. How to extract beats from gestures is a topic [11, 14, 20, 23] outside the scope of this article. We instructed our users to conduct in a simple up-down motion; their beats are marked by the lower turning point of the baton for these gestures.

It is important to emphasize that our intention is not to judge how *well* a person can conduct—this type of evaluation is well beyond our capabilities as conducting system designers; moreover, it is questionable whether or not such an evaluation can be performed systematically given the widely differing conducting styles amongst conductors. What we hope to achieve is a measurement of *how much conducting training* a person has undergone, and adapt the system to their level of ability.

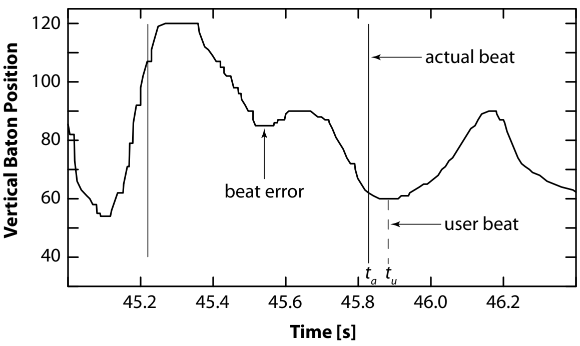

Figure 1. Sample y vs. t plot of a non-conductor showing where he has placed his beats (t_u) relative to the actual beats (t_a). A beat error occurs at around time $t = 45.5$ seconds.

A system that is able to adapt to a user's conducting ability would also require a good understanding of their conceptual model of conducting, such as whether users conduct ahead or behind the music beat, or whether they conduct to the rhythm or to the beat. Also, it would be interesting to see if these conceptual models can be influenced by introducing a simple metaphor that could, for example, be given as part of instructions to a music exhibit in a museum.

To meet our objectives, we needed to observe and collect data on conductors and non-conductors in a controlled environment; thus, we decided to analyze conducting behavior using a fixed recording that does not change in tempo or volume in response to user input. By using this "passive" system, we ensured our results would not be adversely influenced by our previously observed usability breakdowns. We assume that our findings apply to an "active" system where the tempo and volume change in response to user input, and plan to verify this assumption in future work.

HYPOTHESES

Definitions

We start by defining some of the gesture parameters that we will use to measure conducting ability. Fig. 1 shows a sample plot of a user's vertical baton position over time. As the user was instructed to conduct in a simple up-down motion, the lower inflection point marks his beats (t_u). The actual beats of the music are also shown on this plot (t_a).

- *beat offset:* The time difference between where a user places his/her beat and the actual beat: $\Delta t = t_u - t_a$. A negative value occurs when the user conducts ahead of the beat; a positive value occurs when the user conducts behind the beat. The mean beat offset, $\overline{\Delta t}$, is the average of the user's beat offset over the piece.
- *beat variance:* A measure of how much a person's beat offset varies over the piece. The beat variance, σ, is the standard deviation of Δt over the piece.
- *beat error rate:* A measure of how often a user makes a beat error with his/her gestures; a beat error occurs when the user skips a beat or adds a beat that is not in the music (see Fig. 1). The mean beat error rate, $\overline{\epsilon}$, has units of errors/beat.

Based on previous qualitative evaluations of our conducting systems, we predicted the following:

H1: The ictus (lower turning point) of a conductor's gestures to a fixed recording occurs significantly ahead of a non-conductor ($\overline{\Delta t_c} < \overline{\Delta t_n}$).

H2: The ictus of a conductor's gestures to a fixed recording varies significantly less than a non-conductor ($\sigma_c < \sigma_n$).

H3: A conductor makes significantly less errors when marking the beats with his/her gestures to a fixed recording than a non-conductor ($\overline{\epsilon}_c < \overline{\epsilon}_n$).

H4: The ictus of a conductor's gestures to a fixed recording occurs consistently ahead of the music beat ($\overline{\Delta t_c} < 0$).

H5: The ictus of a non-conductor's gestures to a fixed recording occurs consistently behind the music beat ($\overline{\Delta t_n} > 0$).

H6: A non-conductor's musical experience[1](expertise with one or more musical instruments) is correlated to their $\overline{\Delta t}$, σ and $\overline{\epsilon}$ values. A person with more musical experience will have $\overline{\Delta t}$, σ, and $\overline{\epsilon}$ values closer to a conductor's.

H7: A non-conductor's conducting performance ($\overline{\Delta t}$, σ, and $\overline{\epsilon}$ values) can be improved through the use of a simple metaphor, such as "conduct as if reeling in a fish, where you pull the beat (fish) with each gesture".

Testing H1, H2, and H3 will help us meet objectives 1 and 2 (determine level of conducting training, obtain quantitative measurements on gesture beat timing). Testing H6 will help us meet objective 3 (understand what factor(s) effect gesture beat timing). Data from H3, H4, H5 and H7 will help us infer users' various conceptual models of conducting and thus meet objective 4 (better understand users' conceptual models of conducting).

METHOD

Experimental Set-up

Our user study was performed with the aid of a *Buchla Lightning II* system [5]. The *Lightning II* consists of a baton that emits an infrared signal; the emitted signal is tracked by a controller that converts it to MIDI (Musical Instrument Digital Interface) data. We wrote *GestureRecorder*, a custom software that plays back a QuickTime movie and records the current baton position to a file together with the current position in the movie. For the study, the software was run on

[1]We would actually like to test correlation with musical *ability*. Unfortunately, there are no clear standards for measuring musical ability. Metrics have been proposed in the past [4, 15, 25], with some more recent work done by Edwards [8]. For simplicity, we will use one's expertise with musical instruments as an approximate measure of musical ability and qualify this as musical experience.

Figure 2. Devices used in our user studies: 14" iBook laptop computer and a *Buchla Lightning II* baton and tracker.

a 14" iBook laptop computer with a 933 MHz G4 CPU, a 1024x768 resolution display, and 640 MB RAM (see Fig. 2).

Since we sought to obtain quantitative measurements using this set-up, we had to account for system latency; this system latency includes the output latency (time it takes for the system to render video to the display or audio to the speakers) and the input latency (time it takes for the system to receive input from the baton). We measured this system latency by simultaneously filming, using a Redlake MotionXtra HG-100K high-speed camera at 500 frames per second, the physical baton and a display showing its currently tracked position. We determined the latency to be between 90 and 100 ms and subsequently offset data collected from *GestureRecorder* by 95 ms prior to analysis.

We selected an audio and video recording of *Radetzky March* by Johann Strauss, performed by the Vienna Philharmonic and approximately 3 minutes long, as the musical piece for our user studies. We selected this piece because its mostly constant tempo and percussive nature make its beats easy to track. We had previously used this piece in one of our conducting systems, and had observed users interacting with it. Thus, we expected any differences we would observe in beat placement between conductors and non-conductors to establish a minimal difference between the two groups; non-conductors would likely have even more difficulties placing their beat compared to conductors for more difficult pieces. The tempo of this recording varies between 75 and 125 bpm (beats per minute), averaging around 100 bpm.

The "actual" beats of the piece were required for comparison. These beats were manually marked using *BeatTapper*, another software tool we implemented to play the movie, mark its beats, and fine-align them graphically with the transients (energy spikes caused by the percussion that humans perceive as beats) of the audio waveform.

Participants

23 volunteers (6 conductors and 17 non-conductors) were recruited for this user study. Conductors were between 36 and 66 years of age, and had between 10 and 45 years of professional conducting experience. The 17 non-conductors were between 19 and 53 years of age with varying musical expertise, but no conducting experience. Participants were compensated with some chocolate for their time.

Overall Procedure

We divided our studies into two stages: in the first stage we compared conductors and non-conductors, and in the second stage we compared non-conductors before and after introducing a "fishing rod" metaphor.

In the first stage of our user studies, all 6 conductors and 11 of the 17 non-conductors were first shown a 30-second clip of *Radetzky March* audio and video recording to ensure they had some idea of the piece. They were then asked to use the Buchla baton to "conduct" this recording using up-down movements; they were aware, however, that their movements did not affect the movie speed or volume. Each user was asked to conduct the entire 3 minute piece twice, and then requested to fill out a short questionnaire regarding their level of musical or conducting expertise.

The remaining 6 non-conductors participated in the second stage of our studies, and were also asked to conduct the recording twice. The first time through the piece, they were given the same instructions as in the first stage ("use up-down motions"); however, for the second time, they were instructed to use the baton like a fishing rod, imagining that they were pulling a fish out with each beat[2]. This "fishing rod" test was always done on the second trial to prevent these instructions from influencing the "regular" test; we believed that this influence would be greater than any learning effect from always doing the fishing rod test in the second trial.

RESULTS

We implemented a third software utility, *BeatVisualizer*, to simultaneously view the QuickTime movie, music beats, baton position, and graph of vertical baton path (see Fig. 3). Using this tool, we were able to visually confirm that our users marked the beats with the lower turning points of their gestures, and not the upper turning points (there was one exception, which we will discuss in the next section). Thus, the lower inflection point of a y vs. t plot marks the beats (Fig. 4). However, the gestures of non-conductors were sometimes erratic, especially in sections of the piece where the beat was more difficult to track (for example, where there was no percussion). We also found that non-conductors' movements often followed the rhythm of the piece rather than the beat, and that the size of their gestures naturally followed the volume of the music. Thus, we chose to manually mark the beats of the conducting gestures rather than processing the data automatically. To reduce the amount of data to process, we selected a part of the music 40 seconds into the piece and 40 seconds long (beats 55–121 inclusively).

[2] A professional conductor might argue that "fishing" is not the most appropriate metaphor for conducting, since it places more emphasis on the upwards movement, when in fact a strong downwards movement is desired in professional conducting. However, our hypothesis was that by asking users to conceptually think about "pulling" the music beat, they would naturally lead it rather than follow it. Since proper conducting technique cannot be taught in one or two short instructions, we did not make it a priority.

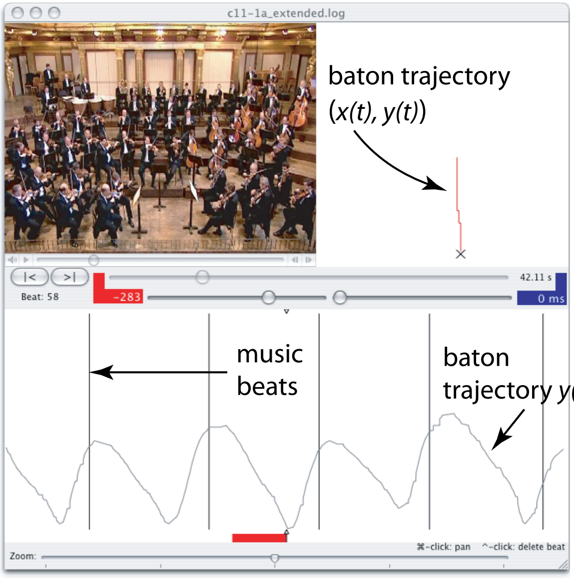

Figure 3. ***BeatVisualizer*** **program, which we wrote for visualizing users' baton gestures and marking their beats. The data shown is from a conductor.**

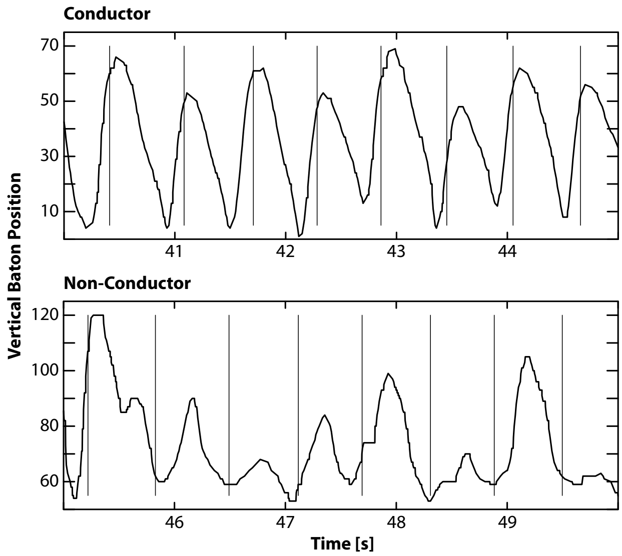

Figure 4. Sample y vs. t plot of a conductor and a non-conductor. Conductors conduct more consistently than non-conductors. The vertical lines mark the actual beats of the music.

Conductors vs. Non-conductors

We used Student's t-test (two-sample, 1-tailed, assuming unequal variances) to compare conductors and non-conductors. Fig. 5 shows a plot of the mean beat offset ($\overline{\Delta t}$), variance (σ), and error rate ($\overline{\epsilon}$) for the two groups.

The t-test found that conductors conduct on average significantly more ahead of the beat than non-conductors ($t = -6.34$, $df = 13$, $p < 0.001$). With a 95% confidence interval, conductors conduct on average 152 ± 17 ms (corresponding to about $\frac{1}{4}$ of a beat at 100 bpm) ahead of the beat while non-conductors conduct on average 52 ± 26 ms ($\frac{1}{12}$ of a beat) ahead of the beat.

The t-test found that conductors conduct, on average, significantly more consistently to their beat than non-conductors ($t = -2.38$, $df = 9$, $p < 0.02$). With a 95% confidence interval, the average beat variance is 47 ± 4 ms ($\frac{1}{12}$ of a beat) for conductors and 72 ± 21 ms ($\frac{1}{8}$ of a beat) for non-conductors.

Due to the way our mean beat error rate data was distributed within the user groups, we did not perform a t-test to compare the two groups and conclude that the error rate is not a good metric for distinguishing conductors and non-conductors.

Effect of Conducting Experience

We found no obvious correlation between a conductor's experience with conducting (number of years) and their mean beat offset, variance, and error rate.

Effect of Musical Instrument Experience

We used the results of the questionnaire users completed after participating in our study to rank our users by music expertise. The criteria we used in our ranking were: number of musical instruments, experience with each instrument in years, and self-rated level of ability. We then used this information to calculate a "musical ranking" from 0 to 1 for each non-conductor, with 0 being no musical expertise and 1 being a high level of musical expertise.

Plots of this ranking against the mean beat offset, variance, and error rate over this musical ranking are shown in Fig. 6. Based on these graphs, we can see that there is no obvious correlation between musical experience and these three parameters.

Effect of a Metaphor on Conducting

Paired plots of the data collected for the 6 non-conductors who participated in the "fishing rod" experiment are shown in Fig. 7. Using a paired Student's t-test, we found no significant difference in the three conducting parameters between regular conducting and conducting with the "fishing rod" metaphor, and conclude that this particular metaphor does not influence a person's conducting behavior.

Summary of Results

Table 1 shows a summary of the results cross-referenced with our original hypotheses.

DISCUSSION

Of the data we collected from our 23 participants, we found two outliers in our data that we subsequently discarded from the analysis. Both users were non-conductors. One participant was a little too enthusiastic in his conducting, resulting

Hypothesis	Description	Supported?
H1	Conductors conduct ahead of non-conductors.	Yes
H2	Conductors vary their beats less than non-conductors.	Yes
H3	Conductors make less beat errors than non-conductors.	No
H4	Conductors conduct ahead of the beat.	Yes
H5	Non-conductors conduct behind the beat.	No
H6	A non-conductor's musical experience influences their placement of beats.	No
H7	A non-conductor's conducting can be influenced using a fishing metaphor.	No

Table 1. Summary of results cross-referenced with hypotheses.

in erratic data that frequently left the range of the *Lightning II* tracker (and almost smashing the baton onto the iBook screen in the process). The other participant appeared to have a different mental model of synchronizing his gestures to the beats: he conducted in a "pendulum" style, swinging the baton back and forth in an arc like a pendulum and synchronizing his beats to the *upper ends* the arc rather than the lower inflection point. Since all other participants synchronized the music beats to the lower turning point of their gestures, we discarded this particular data to maintain consistency in our data set.

Our results support using a user's beat offset and variance parameters for determining whether or not the user is a conductor, but not the beat error rate. We examined more closely the data collected from the participants with the two highest beat error rates. Replaying their baton movements synchronously with the movie, we saw that they had a mental model of conducting to the musical *rhythm* of the piece rather than to the beat.

There appears to be no correlation between a conductor's mean beat offset, variance, and error rate. For non-conductors, the strongest correlation is between their mean beat variance and the square root of the mean beat error ($r_{\sigma,\sqrt{\epsilon}} = 0.91$, see Fig. 8), but no correlation between the other values. As higher beat variance means that users are having more trouble marking a consistent beat, and higher beat errors were seen to be associated with users conducting to the rhythm rather than the beat, perhaps these trends are related to a person's experience or natural ability with music. This theory would also explain why there is no such correlation for conductors. Further user tests would be required to make a conclusive statement.

Based on our results, however, we can say that a person's experience/ability to *play* a musical instrument does not influence their conducting behavior to a fixed recording; some people who have had no musical training were able to time their beats better and more consistently (relative to a conductor) than a person with over 30 years experience playing the flute and guitar at an intermediate level, or a person with 6 years experience playing the trumpet at an expert level. More study would be required to see if this beat timing and consistency is associated with other factors, such as level of familiarity with the piece or the musical quotient proposed by Edwards [8]. However, our results clearly show that there is no obvious equivalent to professional conducting training/experience that will cause a person to time his/her beats similar to a conductor.

Our results disprove our original hypothesis that non-conductors conduct consistently behind the beat. Only one user had an average beat offset behind the beat ($\overline{\Delta t} = 3$ ms). However, many users conducted behind the beat at some point during the piece, which could still explain the "spiral of death" problem we have previously observed with existing conducting systems. Moreover, we believe that users' familiarity with the piece could influence their mean beat offset, variance and error rate. The piece we chose for this study, *Radetzky March*, was well-known amongst our test group of Germans (it appeared in a popular television commercial a few years ago): only one user did not know the piece. The piece also has a strong percussion, which may help users predict where the beat is. Our results seem to support this theory. Let us define a user's *normalized beat offset*, $\widetilde{\Delta t}$, to be:

$$\widetilde{\Delta t}_i = \frac{\Delta t_i - \overline{\Delta t}}{\sigma}$$

where i is the beat number. Fig. 9 shows a plot of the normalized beat offset over time for five non-conductors, filtered with a 9-point averaging filter to reduce noise. One can notice a trend where the users are consistently conducting behind their average beat between beats 67 and 77. One explanation for this phenomenon is that they are hesitating, unsure of their placement of the beat. In fact, beats 64 to 77 correspond to a section of the piece that is not part of the main theme (less likely to be familiar) and the music has no percussion (more difficult to track the beat).

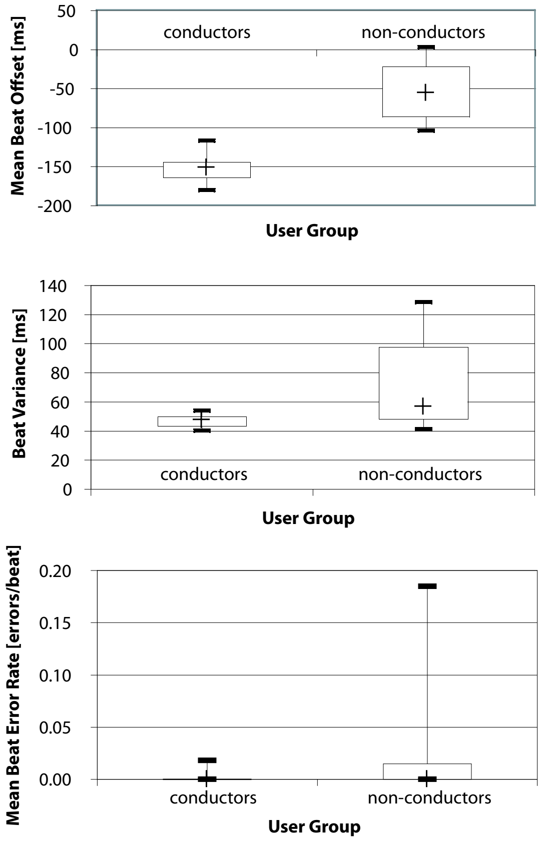

Figure 5. A comparison of conductors and non-conductors using the mean beat offset ($\overline{\Delta t}$), beat variance (σ) and mean beat error rate ($\overline{\epsilon}$). The mean beat offset and beat variance for the two groups are significantly different.

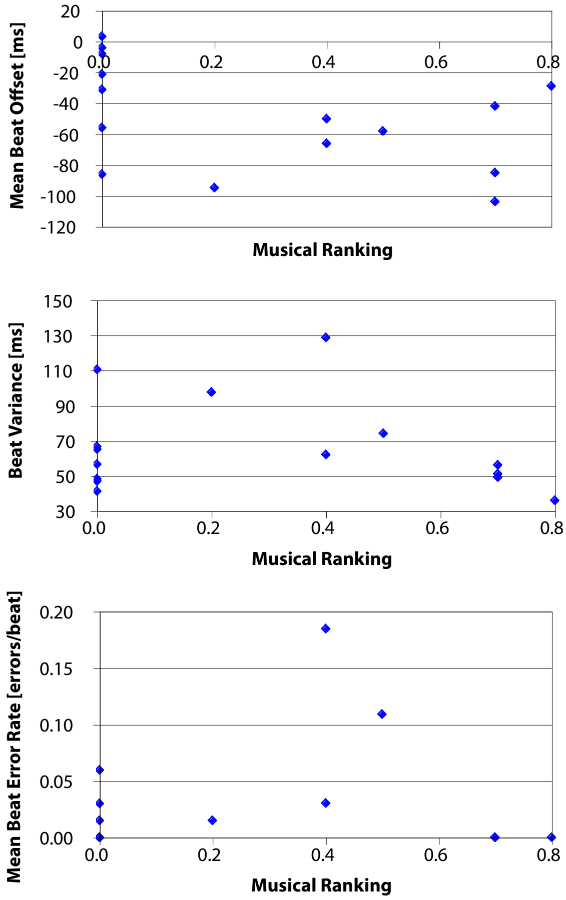

Figure 6. Effect of musical ranking (0 = no experience, 1 = lots of experience) on conducting. There does not appear to be any correlation between users' ability to play a musical instrument and their mean beat offset, variance, and error rate.

DESIGN IMPLICATIONS

The results we have obtained can be used directly to improve the usability of conducting systems. For example, we now have quantitative metrics to show that while conductors' gestures vary widely from conductor to conductor, their beats are placed consistently ahead of the music beat (and with little variance). Thus, when designing a conducting system for conductors, it is important to account for this "lead time" in the tempo following algorithm for matching a musical piece's tempo with users' gestures. This temporal aspect has not been rigorously addressed in previous literature [3, 24].

We are currently incorporating our results into the design of an upcoming conducting system for public spaces. The ability to systematically distinguish conductors from non-conductors allows us to build a system that adapts to a user's conducting ability.

The first step is to determine whether or not an arbitrary user is a conductor. We can measure the timing (beat offset and variance) of the user's placement of the beats to the music using, for example, the first 10 seconds of the piece, where we fix the tempo. Extracting beats from conducting gestures has been addressed before in previous systems [11, 14, 20, 23], although more work is required to parse arbitrary gestures. Based on these initial beat-timing measurements, we can classify the user as a conductor (mean beat offset is between roughly 130 to 170 ms, and the variance is less than roughly 50 ms for *Radetzky March*) or non-conductor (mean beat offset is less than 130 ms, or the variance is greater than 50 ms for *Radetzky March*).

Since we can depend on the precision and reliability of conductors' movements, tempo changes in response to their gestures can be instantaneous. In fact, since their placement of the beats is less likely to be random and/or unintentional, these users would benefit from having their movements "tightly-coupled" to the music. Non-conductors, on the other hand, would benefit from some averaging of the

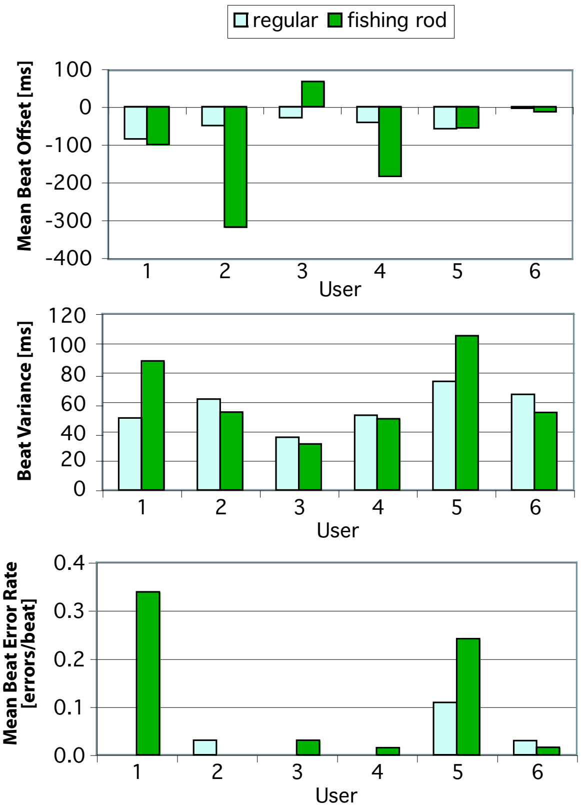

Figure 7. Paired plot of the conducting parameters for each user, before and after being instructed to conduct "fishing rod" style. The metaphor does not significantly improve one's conducting.

data collected from the gestures over a certain time window. This averaging would mitigate the effects of user errors, and the size of this time window can be a function of the variance measurement (higher variance is correlated to higher number of errors). The beat variance can also be tracked as the user continues through the piece, with the system reducing the averaging window size if it detects an improvement in the conducting, or vice-versa.

Such a system would not only be enjoyable for a wider range of users, but it would also enable us to continue our study of conducting behavior amongst conductors and non-conductors, and continue to better understand peoples' conceptual models of conducting. We also believe it can be adopted as a "training wheels" system for student conductors. By allowing them to produce pleasant results with their conducting from an early stage, we hope to offer to them a better way to navigate the learning curve.

FUTURE WORK

We have identified several areas that deserve further investigation:

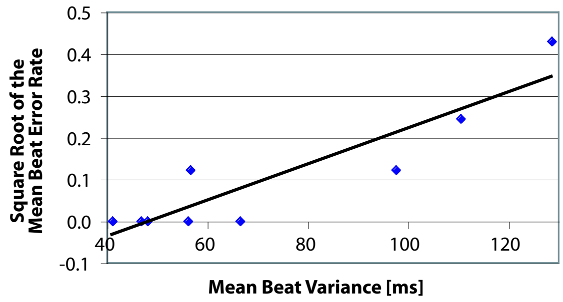

Figure 8. Correlation ($r = 0.91$) between beat variance and the square root of the mean beat error rate.

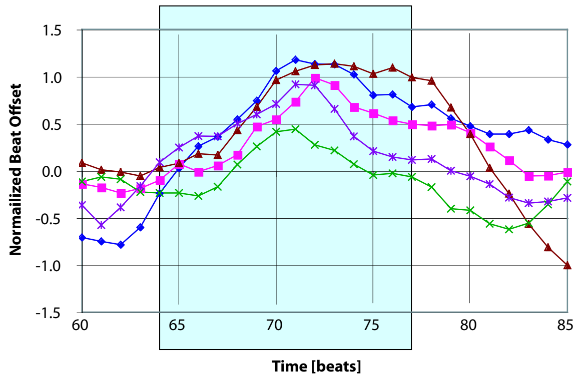

Figure 9. Plot of the normalized beat offset ($\widetilde{\Delta t_i} = \frac{\Delta t_i - \overline{\Delta t}}{\sigma}$, where i is the beat number) for five users over time. The consistent hill suggests that users are unsure with their placement of the beat and thus hesitating. The shaded region between beats 64 and 77 marks a section of the piece that is not part of the main melody, and has no percussion.

Trends amongst student conductors: We found a statistically significant distinction between conductors and non-conductors, but no obvious correlation between our measurements and conductors' level of conducting experience or non-conductor's level of musical instrument experience. We would like to continue our studies with student conductors at various stages of their education to determine if trends exist within this group that bridge the gap between conductors and non-conductors.

Trends amongst non-conductors: Unlike conductors, non-conductors had a larger variance in their measurements of beat offset and variance. We would like to continue to explore the factors that could cause such a wide range. Some possibilities are a user's level of familiarity with the musical piece, or his/her natural musical talent (which could possibly be measured using Edwards' proposed musical quotient). Previous studies have shown that children have a different beat and tempo perception than an adult [9, 12], so age may also be a factor.

Trends amongst different musical pieces: We limited this work to one musical piece. A further dimension would be to test with multiple pieces, with varying tempi and amount of percussion. Such a study would, for example, allow us to determine how tempo influences conducting ahead, from conducting 150 ms ahead of the beat independent of the tempo (absolute offset) to $\frac{1}{4}$ of a beat (relative offset). Since people often use percussion as a guide for finding the beat, it would also be useful to see how this influences users' placement of beats.

Testing with an active conducting system: We used a passive system in our user tests, where the user input does not affect the music tempo, to be able to consistently obtain quantitative measurements of beat placement, variance and error across our user groups; we assumed that our results were valid for an active system that responds to user input. To verify this assumption, we could change the beat following algorithm in our current conducting systems and then perform studies to see how it affects users' interaction with the system.

CONCLUSIONS

We presented an empirical analysis of users conducting to a fixed audio and video recording of a popular classical musical piece, *Radetzky March*. Our analysis yielded quantitative and qualitative results comparing the conducting gestures of conductors and non-conductors. Based on feedback from preliminary user interviews and inspired by previous evaluation of our own conducting systems, we chose to examine the temporal characteristics of conducting gestures, rather than their spatial properties. In particular, we measured how far users place their beats from the actual music beats, how much their beats vary, and the rate at which they incorrectly mark beats.

We found that conductors conduct on average 152 ± 17 ms ($\frac{1}{4}$ of a beat at 100 bpm) ahead of the beat with an average variance of 47 ± 4 ms ($\frac{1}{12}$ of a beat). Non-conductors conduct on average 52 ± 26 ms ($\frac{1}{12}$ of a beat) ahead of the beat with an average variance of 72 ± 21 ms ($\frac{1}{8}$ of a beat). All intervals were computed with a 95% confidence. We found that how far ahead a person conducts to the beat and how much s/he varies the beat can be used to distinguish between a trained conductor and a non-conductor.

We also discussed differences in conceptual models of conducting based on our quantitative results. Our test participants were instructed to conduct in an up-down motion, and most participants intuitively synchronized the music beat to the downwards turning point of their gestures ("foot-tapping" metaphor). However, we also observed one participant conducting like a pendulum, synchronizing his beats to the upper turning points of his gestures ("pendulum" metaphor). Furthermore, our analysis of beat error rates revealed that high error rates were caused by users conducting to the rhythm, rather than to the beat, of the music. Finally, there is a correlation between how often non-conductors incorrectly placed beats and how much they vary their beat placement ($r = 0.91$).

We aim to improve the usability of computer conducting systems using these results; our adaptive conducting system for public spaces will give a wider range of users a satisfying experience. This type of system can also help students practice conducting, using technology to smoothen their learning curve.

While our user study was centered around conducting gestures, we believe our results apply to how people temporally map gestures to music rhythm and beat in general. Dance, for example, is another area of gestural interaction with music where our work could be applied. As interactions with time-based media, such as audio and video, become more ubiquitous, we hope our results will serve as a foundation and inspire further work on creating new and better gestural interfaces to time-based media.

ACKNOWLEDGEMENTS

The authors would like to thank Thorsten Karrer for his work on the *BeatTapper* program; Jorinde Witte for her help in designing the user experiments; Rafael Ballagas, Steve Yohanan, Sidney Fels, and Teresa Marrin Nakra for their valuable feedback; and all the people who participated in our user study, in particular Gisbert Stenz who provided much assistance in our understanding of conducting.

REFERENCES

1. Apple Computer. Motion. http://www.apple.com/motion/.
2. Borchers, J. WorldBeat: designing a baton-based interface for an interactive music exhibit. *Proc. CHI 1997*. ACM, 1997, 131–138.
3. Borchers, J., Lee, E., Samminger, W., and Mühlhäuser, M. Personal Orchestra: A real-time audio/video system for interactive conducting. *ACM Multimedia Systems Journal Special Issue on Multimedia Software Engineering 9*, 5 (2004), 458–465. Errata published in next issue.
4. Boyle, J. D., and Radocy, R. E. *Measurement and Evaluation of Musical Experiences*. Schirmer Books, New York, 1987.
5. Buchla, D. Lightning II MIDI controller. http://www.buchla.com/.
6. Clarke, D. J. MIT grad directs Spielberg in the science of moviemaking. *MIT Tech Talk 47*, 1 (2002).
7. Desain, P., and Honing, H. Computational models of beat induction: The rule-based approach. *Journal of New Music Research 28*, 1 (1999), 29–42.
8. Edwards, A. D. N., Challis, B. P., Hankinson, J. C. K., and Pirie, F. L. Development of a standard test of musical ability for participants in auditory interface testing. *International Conference on Auditory Display*. 2000.
9. Flohr, J. W., and Rose, E. Young children's ability to perform a steady beat. *Music Educators National Conference Convention*. 1996.

10. Harrer, G. *Grundlagen der Musiktherapie und Musikpsychologie*. Gustav Fischer Verlag, Stuttgart, 1975.

11. Ilmonen, T., and Takala, T. Conductor following with artificial neural networks. *Proc. ICMC 1999*. ICMA, 1999, 367–370.

12. Kube, G. *Kind und Musik. Psychologische Voraussetzungen des Musikunterrichts in der Volksschule*. Kösel, München, 1958.

13. Lee, E., Nakra, T. M., and Borchers, J. You're the Conductor: A realistic interactive conducting system for children. *NIME 2004*. 2004, 68–73.

14. Lee, M., Garnett, G., and Wessel, D. An adaptive conductor follower. *Proc. ICMC 1992*. ICMA, 1992, 454–455.

15. Lehman, P. R. *Tests and Measurements in Music*. Prentice-Hall, Englewood Cliffs, New Jersey, 1968.

16. Lionhead Studios. Black & White. http://www.bwgame.com/.

17. Marrin Nakra, T. *Inside the Conductor's Jacket: Analysis, interpretation and musical synthesis of expressive gesture*. PhD thesis, Massachusetts Institute of Technology, 2000.

18. Mathews, M. V. *Current Directions in Computer Music Research*. MIT Press, Cambridge, 1991, ch. The Conductor Program and Mechanical Baton, 263–282.

19. Morita, H., Hashimoto, S., and Ohteru, S. A computer music system that follows a human conductor. *IEEE Computer 24*, 7 (1991), 44–53.

20. Murphy, D., Andersen, T. H., and Jensen, K. Conducting audio files via computer vision. *Gesture Workshop 2003*. Springer (2004), vol. 2915 of *Lecture Notes in Computer Science*, 2004, 529–540.

21. Myers, B. A. A brief history of human computer interaction technology. *ACM interactions 5*, 2 (1998), 44–54.

22. Palmer, C., and Krumhansl, C. L. Mental representations for musical meter. *Journal of Experimental Psychology - Human Perception and Performance 16*, 4 (1990), 728–741.

23. Usa, S., and Mochida, Y. A conducting recognition system on the model of musicians' process. *Journal of the Acoustical Society of Japan 19*, 4 (1998).

24. Usa, S., and Mochida, Y. A multi-modal conducting simulator. *Proc. ICMC 1998*. ICMA, 1998, 25–32.

25. Wing, H. D. *Tests of musical ability and appreciation*. Cambridge University Press, Cambridge, 1968.