MUSICAL ANALYSIS OF CONDUCTING GESTURES USING METHODS FROM COMPUTER VISION

Teresa Marrin Nakra[1], Andrea Salgian[2], Micheal Pfirrmann[2]

The College of New Jersey
[1]Department of Music
[2]Department of Computer Science
nakra@tcnj.edu, salgian@tcnj.edu, micheal.pfirrmann@gmail.com

ABSTRACT

We present technical interpretations of a noted conductor's gestures. These were enabled by Computer Vision techniques that tracked the position of the conductor's right hand from a video sequence. The output of the hand tracking system, when combined with beat and tempo data from the audio signal, provides many possibilities for analyzing the conductor's gestures and expressive techniques. We describe the stages of the video/audio processing, and present our analyses of the conductor's movements.

1. INTRODUCTION

We undertook a joint research project to investigate the gestures of a noted professional orchestra conductor, using new Computer Vision techniques that extracted the vertical component of the movements. In this paper we present our analyses of the conductor from a concert performance. We specifically used a live performance because we assumed that the motivation to perform authentically, correctly and expressively was high. There is scientific value in understanding how professional conductors convey emotion and meaning through gesture; analysis of real conducting data can reveal basic truths about how humans express non-verbal information such as emotion. Our source video features the Boston Symphony Orchestra and conductor Keith Lockhart. This footage was obtained during a performance of the orchestra at a concert in April 2006, one of many data streams collected in a research collaboration between the BSO, McGill University, Immersion Music, and TCNJ.

2. PRIOR WORK

There has been a great deal of work in sensor-based musical systems for conductors in the computer music community. The majority of these projects have featured sensor interfaces using accelerometers, gyroscopes, pressure sensors, infrared tracking devices, electric field proximity sensors, sonar arrays, RF transmitters, EM sensor arrays, and physiological monitors. Perhaps the best known of the many sensor batons is the Mathews/Boie *Radio Baton*, invented in 1987 [7]. Several other batons have featured IR tracking devices, including those by Marrin and Paradiso [6], Morita et al [8], and Lee et al [4].

Much of the prior work has featured interactive conducting applications that control playback parameters in audio and video. Our approach instead has been to analyze the moment-to-moment inflections in a professional conductor's gestures. There have been others who have done similar work; most notably Camurri et al [1], Kolesnik and Wanderley [3], and Lee et al [5]. In addition, Nakra's *Conductor's Jacket* interface [9] was used for both interactive performances and analysis.

3. VIDEO PROCESSING

We analyzed a segment of video footage from the concert and focused on the conductor's right hand, because it is standard in conducting technique for conductors to beat time with the right hand. We tracked the position of the hand and plotted its height over an 80-second recording of Mozart's Marriage of Figaro Overture.

The video sequence was edited to begin at the second statement of the opening theme, because the first several seconds of the original video featured a zoom-in that was unusable for vision tracking. Our segment begins at the moment when the zoom stopped and the image stabilized. In order to center the conductor in the image, we cropped the frames and zoomed in to maximize his image. Occasionally some occlusion occurred between the right and left hands. This was the majority factor in the occasional errors in the tracking system's output. A detailed explanation of the Computer Vision techniques used is featured in Salgian et al [11]. Figure 1 shows a frame from our source video sequence.

After extracting the hand position, we next found the beats in the audio track and aligned them with the video beats. For this, we applied the Ellis/Poliner beat tracking algorithm [2]. A frame is shown below; the final video is available at http://www.tcnj.edu/~nakra/BSO.avi. It consists of two images, set side-by-side: the left side contains the cropped and zoomed portion of the original

video, with tracking on the right hand; the right side consists of three overlaid items:

- a vertical dotted line indicating the current moment in the video recording, intersected by a red circle indicating the height of Lockhart's right hand.
- a continuous dark line indicating the vertical position of Lockhart's right hand over time.
- a series of green dots indicating the moments when the audio beat tracker determined that a beat occurred.

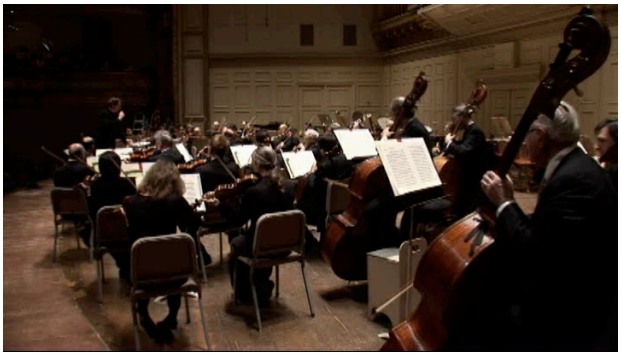

Figure 1. Frame from source video

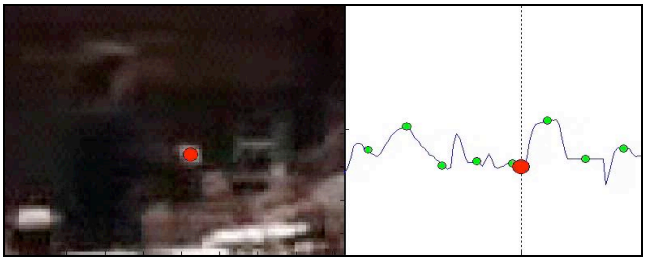

Figure 2. The hand tracker and aligned beats.

4. MUSICAL ANALYSIS

Our musical analytical goals with this project were to determine whether the conductor's right hand height correlated with the tempo of the performance, musical expression markings, and structural elements in the score. Using all the visual information provided by the various data streams in our output video sequence, we made several observations related to Lockhart's conducting technique. While these observations strictly refer to this conductor, it can be assumed that some of his features may also apply to analyses of other conductors' gestures. Our observations are listed in the subsections below.

4.1. Deviation from Pedagogical Conducting Technique

We began with an assumption that this conductor's movements would somewhat conform to standard conducting techniques, as documented in pedagogical texts such as the one by Max Rudolf [10]. The Rudolf text explains that the vertical component of a conductor's right hand is associated with beat (pulse) information. Further, Rudolf defines the two-dimensional trajectory of a 2-beat metrical pattern (featured in this Mozart work) as a downward gesture (beat 1) followed by an upward gesture (beat 2), where the moment of the beat is indicated at the bottom inflection point of both gestures.

It became clear from repeated viewings of the video that this conductor's gestures did indeed feature a regular down-up trajectory. However, the moment of the beat was not always indicated according to the method defined by Rudolf. For example, Lockhart's downbeat (beat 1) gestures appeared mostly to be placed at the lower inflection point, but the beat 2 indications were more irregular – sometimes at a lower inflection point, sometimes at an upper inflection point, sometimes at a midpoint in the gesture. We assume that this deviation from pedagogy does not indicate a problem; rather, we assume that with a professional orchestra, temporal clarity in the second beat is not required.

We were also surprised to discover that the correlation between the conductor's beats and orchestra's beats was not consistent. At times, the orchestra's beats line up with the inflection points of the conductor's gestures, but at other times they do not. This also contradicts conducting pedagogy, which presumes that the orchestra is consistently and synchronously following the conductor's gestures. Instead, our video sequence shows that the beats of the orchestra and conductor appear to be frequently out of phase with each other. As a result, we were not able to determine any regular time delay between the conductor's beats and the orchestra's response, as we had planned.

4.2. Correlations between Beat Patterns and Musical Expression

Despite the lack of consistent temporal and phase correlation between the conductor's gestures and the orchestra's beats, we did find a strong set of correlations between the conductor's right hand height and structural or expressive elements in the score. Generally, the vertical features of movement in Lockhart's right hand changed with the phrase structure of the music. Below we describe several features that were revealed by our visual analysis:

4.2.1. Height Deltas

At times, there are marked changes in the distance between the hand's upper and lower inflection points. The change can be gradual or abrupt. This phenomenon seems to relate to expressive elements in the score, particularly volume and phrasing. For example, at 0:02-0:06, hand height slowly increases, indicating increased tension to the middle of the phrase. This segment corresponds musically with the second statement of the *A* theme, mm. 18-26. Correspondingly, at 0:06-0:10, the hand height slowly decreases, indicating a new tension profile as the second

part of the A theme resolves, mm. 27-30. Later, at 0:20-0:31 (mm. 43-58), the height delta increases twice, following the contours of two consecutive phrase groupings (the second one repeats).

4.2.2. Tiered gesture "platforms"

Lockhart seems to use baselines of different heights at different times in the music. The choice of height seems to be related to the orchestration and volume indicated in the music. One gesture "platform" begins at 0:27; Lockhart seems to be indicating a new expressive idea here at the beginning of the "B" theme.

4.2.3. Smooth versus jagged beat-shapes

Sometimes the beats appear round and sinusoidal, whereas at other times their contours are jagged and abrupt. Jagged beat shapes appear prominently at 0:00-0:09, whereas round beat shapes appear elsewhere, including 0:56-1:02. We hypothesize that the shapes of the beats are intended to indicate accents (attacks) and other forms of note articulations (envelopes). Also, at 0:20 (m. 43), suddenly the beat shapes become much more even. They begin small and sinusoidal, and then as the height delta dramatically increases over several seconds, the beat shapes also become much more peaked and jagged.

4.2.4. Rate of pattern change

Sometimes a particular feature stays uniform over a passage of music, sometimes it gradually changes, and sometimes it changes abruptly. Accelerating changes seem to signal the musicians to watch for upcoming events.

4.2.5. Preparatory Inhibition

The video yields some evidence that Lockhart uses "preparatory inhibition"; this is the neurological principle that describes how humans reduce movement in advance of a major upcoming event. The use of this phenomenon by conductors has been documented in [9], where it was hypothesized that a sudden lack of information is eye-catching for the musicians and requires minimal effort from the conductor. A quick transition between information-rich and information-neutral states could be an efficient way of cueing musicians to upcoming changes.

4.3. Hand height correlation with Tempo

We also observed that the height of the conductor's right hand correlated generally with the tempo of the music; the higher his hand rose, the faster the musicians played. This is demonstrated in Figure 3, below, which demonstrates continuous tempo and hand height variations during the performance of Mozart's Overture to Marriage of Figaro, mm. 25-126. During this recording, the tempo fluctuated between 134 and 156 BPM.

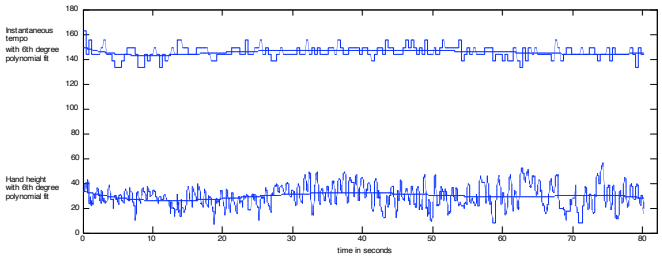

Figure 3. Comparison of tempo and hand height; the top trace shows the tempo and its 6th degree polynomial fit; the bottom trace shows the hand height and its fit.

To determine the correlation between the conductor's hand height and tempo, we compared the moment-to-moment tempo with the vertical movement of the conductor's hand. First we converted the vector of beat onset times (detected by the Ellis/Poliner algorithm, [2]) to a timeseries containing interonset intervals (time deltas between successive beats). We then converted the vector of interonset intervals to a vector of instantaneous tempo values, using this simple algorithm:

$$\text{tempo} = 60/\text{interonset interval} \qquad (1)$$

This gave us a timeseries of local tempo fluctuations, measured in beats per minute (BPM). This timeseries provided a precise, localized representation of the tempo fluctuations, beat-to-beat. (We acknowledge that there is some quantization or aliasing error here; we believe it comes from our Java implementation of the Ellis/Poliner algorithm.) Next, in order to view large-scale trends, we smoothed the tempo data by applying a 6th degree polynomial fit.

We then visually compared the tempo values with the hand height values (bottom of Figure 3), using the smoothed traces in the 6th degree polynomial fit graphs. A visual analysis reveals that both tempo and hand height share a long peak in the center of the segment. Both traces vary similarly over time. This result countered our expectations, because hand height is associated with emotional intensity in the pedagogical literature on conducting. Several conducting texts suggest that conductors should *decrease* the height of the right hand when indicating an *increase* in tempo. Our result demonstrates that the opposite might also be true.

5. CONCLUSIONS AND FUTURE WORK

In summary, our analysis revealed that while our conductor sometimes thwarted standard technique, nonetheless his right-hand movements indicated important correlations with musical elements. The features we identified in the hand height signal that correlated with musical aspects

included height deltas, tiered gesture platforms, beat shapes, rates of pattern change, and preparatory inhibition. In addition, we found a strong correlation between hand height and tempo.

The area of our study that remains most unresolved is that the conductor's beats were often not in phase with the orchestra. We assume that at the performance level of a major professional orchestra, the information that the conductor passes along is not necessarily "traffic control" and coordination, but rather the shaping of the phrases and subtleties in the expressive aspects of the performance.

There is much more work to be done in this area. Very little is known about how professional conductors' gestures actually correlate with orchestral responses. It is hoped that more research will yield interesting and useful findings on the expressive and emotional aspects of conducting gestures. We have targeted several future extensions to this work. First, we plan to compare the visual analyses with sensor data from the same concert -- in particular, we will compare Lockhart's hand height with his heart rate. Secondly, we plan to analyze video from other conductors and orchestras.

Future work in this area could have useful applications. For example, conductor-following systems can be built to interpret conducting gestures in real-time and allow conductors to control media streams in synchrony with a live orchestral performance. (Anecdotally, Lockhart mentioned that it would be fun to be able to control the fireworks or cannons on the 4th of July celebrations in Boston while conducting the Boston Pops Orchestra.) Also, we have planned with the BSO Education Office to present programs to school-age children, using scientific research on music. Finally, musical games using interfaces such as the Nintendo Wii-mote could benefit tremendously from a more sophisticated understanding of how expert conductors use gestures to convey information.

6. ACKNOWLEDGEMENTS

This work was supported by a grant from VRQ to D. Levitin/CIRMMT at McGill University. Additional data collection was supported by the Canada Research Chair in Music Perception and Cognition at McGill. We particularly thank Daniel Levitin and Stephen McAdams for their support and expertise in this collaboration. The authors would also like to thank the Boston Symphony Orchestra and conductor Keith Lockhart for generously donating their recordings and data for this research. Myran Parker-Brass, Director of Education Programs, assisted greatly with this project.

7. REFERENCES

[1] Camurri, A., et al (2000). "EyesWeb: Toward gesture and affect recognition in interactive dance and music systems." *Computer Music Journal* 24 (1), 57 – 69.

[2] Ellis, D., Poliner, G. (2007). "Identifying 'cover songs' with chroma features and dynamic programming beat tracking." *ICASSP*, 1429-1432.

[3] Kolesnik, P. and Wanderley, M. (2004). "Recognition, analysis and performance with expressive conducting gestures." *Proc. of ICMC*.

[4] Lee, E., Nakra, T. M., and Borchers, J. (2004). "You're the conductor: a realistic interactive conducting system for children." *Proc. of the Conf. on New interfaces for Musical Expression*, 2004.

[5] Lee, E., Wolf, M., and Borchers, J. (2005). "Improving orchestral conducting systems in public spaces: examining the temporal characteristics and conceptual models of conducting gestures." *Proc. of SIGCHI Conference on Human Factors in Computing Systems*, pp. 731-740.

[6] Marrin, T. and J. Paradiso. "The Digital Baton: a Versatile Performance Instrument." *ICMC 1997*.

[7] Mathews, M. *The Conductor Program and Mechanical Baton*. Cambridge: MIT Press, 1989.

[8] Morita, H., S. Hashimoto, and S. Ohteru. A computer music system that follows a human conductor. *IEEE COMPUTER*, 24(7): 44-53, 1991.

[9] Nakra, T.M. (2000). *Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture*. Ph.D. Thesis, MIT.

[10] Rudolf, M. (1995). *The Grammar of Conducting: a Comprehensive Guide to Baton Technique and Interpretation* (3rd ed.) New York: Schirmer.

[11] Salgian, A, Pfirrmann, M. and Nakra, T.M. "Follow the Beat? Understanding Conducting Gestures from Video." *3rd Int. Sym. Vis. Comp.*, Springer, 2007.