

ANALYZING A CONDUCTORS GESTURES WITH THE WIIMOTE

David Bradshaw & Kia Ng
ICSRiM - University of Leeds,
School of Computing & School of Music,
Leeds LS2 9JT,
UK

info@icsrim.org.uk
www.icsrim.org.uk

Abstract – This paper presents an overview of a system for the acquisition and analysis of a conductor’s hand gesture. Multiple wiimotes are used to capture 3D position and 3D acceleration data which describe the movements. These are then analysed and fed back to the user using sonification and visualisation. An alternative method for using wiimotes to track 3D position is implemented; data segmentation, three analysis modules and a sonification method are also described. Whilst primarily designed with conducting in mind, the system described should provide a highly useful tool in a study of all forms of musical expression through gesture.

INTRODUCTION

Conducting is a fascinating art form. Despite having no direct control over the sound being produced the conductor is able to drastically affect how a piece of music sounds solely through the use of gesture. This project focuses on providing a system to analyse and feedback information about such gestures in an informative and entertaining manor. The projects main focus is on analysing the movement of a conductor’s hands. The conductor is able to communicate a vast amount of information using hand movements alone such as tempo, dynamic or expressive information. This system aims to extract these features from the movements and feed them back to the user using several appropriate methods including sonification, visualisation and vibrations in the controller. The wiimote is used as a controller and can be used to track 3D movements in the form of both positions and accelerations. This system is not limited to conducting, it is capable of providing feedback on any movement or gesture which would provide a useful tool in any study of musical expression through gesture or other applications were gesture is used to communicate information.

BACKGROUND

This section presents an overview of existing systems and approaches related to conductor’s gesture tracking. There are many important factors to be considered when assessing such system. The technical capabilities of the system must be assessed; these include accuracy and sampling rate of the tracking. The type of sensors that are used to acquire the data is also an important consideration; these can include cameras, accelerometers, pressure, and grip sensors among others. Also human factor issues such as whether the system is intrusive in any way must be considered.

One of the earliest systems is the “GROOVE” project, (General Real-time Output Operations on Voltage-controlled Equipment) [7] that enabled users to have real

time control over synthesis. This was followed by the “Mechanical Baton” in 1989 and later the “Radio Baton” in 1991 [6]. The radio baton made use of 5 directional antennas to detect the 3D position of the tips of two batons at a sampling rate of 200Hz. The baton was used to control a live computer music performance.

In 1991, Morita, Hashimoto and Othuru produced a “Computer Music System that Follows a Human Conductor” [11]. This system combined using a data glove with a CCD camera. The data glove is used to detect left hand expressive movements whilst the CCD camera is used to track an infrared LED mounted at the end of a conductor’s baton in the right hand. The camera used has a resolution of 256 by 256 pixels and the system as a whole had a sampling rate of 30Hz. The system was the first to use camera based tracking for such an application and was used to attain both tempo and dynamics information to control a pre-recorded MIDI score.

The use of cameras proved a success and thus has been adopted in later systems such as the work by Murphy in 2003, “Tracking a Conductors Baton” [12]. This system made use of two cameras, one in front of the user and one to the side (profile) and was used to track a normal conductor’s baton visually. Image processing is used to first locate the position of the baton in *seek* mode. *Track* mode is then used to follow the baton. This data can then be used to create a 3D reconstruction of the batons position. The input images are down sampled to 160 by 120 pixel greyscale images and the system tracks at 25 frames per second. This work was later advanced by Murphy, Andersen and Jensen by implementing an audio playback system that could be dynamically controlled by the conductor’s movements. [13]

Friberg [2] uses a single camera to detect quantity, width and height of motion by computing the difference between frames. Fuzzy logic is then used to map input parameters to a set of emotional parameters which are then used to control various musical and visual applications. Using this technology Friberg later developed a system for controlling a musical score through gestures. [1]

Another camera based approach can be seen in [5] where hidden markov models are used to analyse data provided by two USB cameras.

In 1996, Marrin produced the “Digital Baton” [8]; this was an altered baton that contained a variety of sensors. 2D position was captured by tracking an infrared LED mounted at the tip of the baton. The baton also included 3D accelerometers and grip pressure sensors.

Marrin then went on to develop the “Conductors Jacket” with Picard in 1998 [9]. The jacket was modified to contain multiple sensors of many varieties that could provide vast amounts of different information. These sensors include electromyography sensors measuring muscle tensions, a heart rate sensor, a respiration sensor and also temperature and galvanic skin response sensors. This could provide information including but not limited to tempo, dynamics, articulations and accents and vibrato. Both the “Digital Baton” and the “Conductors Jacket” are used to control pre-recorded MIDI scores.

In 1999, Illmomen and Takala devised a system to track the conductor using artificial neural networks [4]. The system uses a number of 6 degree of freedom (DOF) motion sensors to measure body and hand movements. Current baton tip position, current and delayed velocities are then used as an input vector to multiple artificial neural networks. These are used to analyse the motions which can then be used to control tempo and articulation of a pre-recorded MIDI score.

As seen above there are two main approaches to conductor following, camera based and sensor based. Camera based approaches have the obvious advantage of not affecting the user in anyway as all information is gathered from a 3rd party source (i.e. a camera). However, this could also be considered a disadvantage as it limits the amount of data that can be gathered to observable positions and movements. Cameras also have the disadvantage that they cannot capture any data if the source they are tracking goes out of their field of view or their view is obstructed in any way.

Sensor based methods; involving modifications of equipment or clothing have the advantage of being able to extract data that is not visually observable such as grip pressures, heart rate and muscle tensions among others which can give much greater insight into a mood or expressive intentions. The disadvantage however is that they may be considered intrusive to the performance.

APPLICATIONS

In this section three pedagogical scenarios are discussed for the usage of the system. Whilst the scenarios described mainly focus on learning or practising various basic conducting techniques they should also be applicable to accomplished conductors wishing to compare different styles or simply practice a performance.

The first application of the system concerns recording the path of a user's gesture. Within an educational environment it is often that a teacher would like to show a student a specific style, technique or even how to conduct a specific piece of music. During the learning process the student would try to imitate the teacher until they have mastered the material. The teacher will be able to record a specific gesture, or series of gestures in which the student then imitates. The difference between the movements can then be calculated and the similarities or dissimilarities can be displayed to the user. Equally a more accomplished conductor would be able to perform a series of the same gesture and compare them in order to highlight weaknesses within his or her performance. Considerations into user body size will also be taken. This allows for scaling before the comparison process so that two gestures from different people could be compared respective to the two body sizes of the users.

Secondly, when practicing a technique, the ability to maintain consistency is the key to mastering the material. In this situation the metronome provides an invaluable tool; the user aims to master the material at a slow speed and by gradually increasing the metronome tempo is able to improve their performance until they have mastered it at performance tempo. This system should provide both tempo and consistency tracking, the tempo between beat points can be compared with a set tempo on a metronome and the consistencies between beats can be used to determine the accuracy that the student is able to match this tempo.

Finally, expression is taken into account. The gesture are broken down and analysed in order to provide feedback as to their observable gestural content. This can then be fed back to the user, thus illustrating how the performed gestures are being perceived. This allows the user to alter his or her performance in order to communicate the correct information to the observer, which could be either this system within a practise situation or within a in a live situation, the orchestra.

DESIGN AND DEVELOPMENT

This section details the design of the system; it shows the overall design as depicted in Figure 1. This section then describes data acquisition using multiple wiimotes, data segmentation using acceleration data, three analysis models and a sonification method for tempo data.

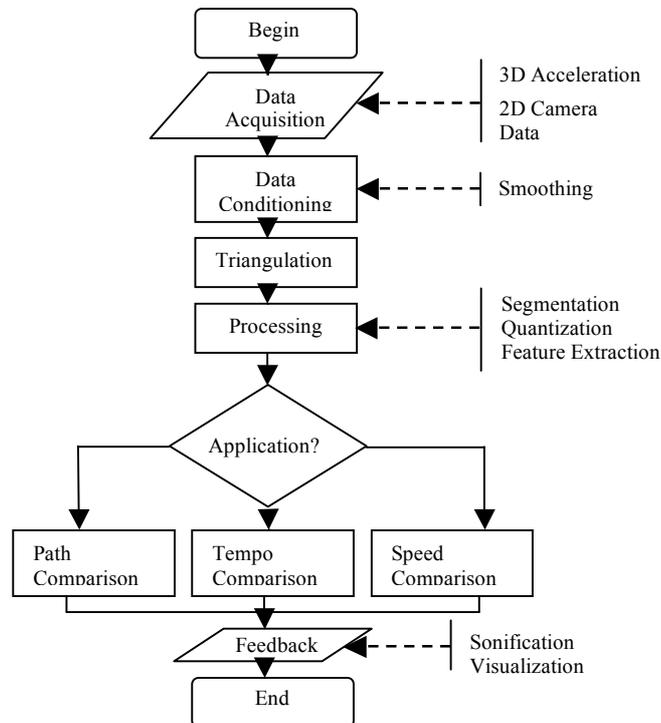


Figure 1. System Flow

Data Acquisition and Wiimote Setup

The wiimote is a commercially available game controller and was chosen as it is affordable and easily attainable. The wiimote contains both a 3D accelerometer capable of measuring up to at least $\pm 3g$ with 10% sensitivity and an infrared detecting camera with a resolution of 1024 x 768 and capable of tracking up to 4 infrared sources at any one time. The wiimote sends data via Bluetooth at a rate of 100 Hz.

The system uses 3 wiimotes; the first is modified by adding an infrared marker to it and is hand held. This provides 3D acceleration data and can provide feedback in terms of vibrations. The remaining two wiimotes are used to track the infrared marker attached to the first, providing two sets of 2D positional data from the wiimote cameras. The wiimotes are positioned 3 meters apart at a height of 136cm. The wiimotes have a horizontal field of view of approximately 41 degrees and are positioned at an angle of 49 degrees to the base line in order to provide the tracking range required whilst maximizing tracking resolution. This is an ideal setup designed to accommodate a large range of motion that could be used in conducting; however, it is possible to reduce the size of the setup for testing situations.

Data Segmentation

The data is segmented in order to aid the analysis process; segmentation is used to break the movements down into their component parts or segments which can then be analysed independently of one another. In conducting, beat points provide ideal segmentation points as they are at regular time intervals and are located evenly throughout a movement. Figure 2 illustrates a typical conducting pattern, showing the location of the beat points.

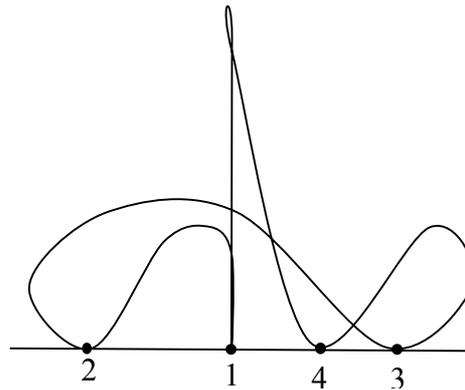


Figure 2. A typical $\frac{4}{4}$ conducting pattern

Acceleration data from the 3D accelerometer is used to detect beat points and thus calculate segmentation points. Beat points coincide with directional changes which form peaks in the acceleration data. At beat points, prominent peaks can be found on one or more axis, Figure 3 shows the accelerations on each axis for two beat points. Depending on the exercises, it can be beneficial for the user to be able to focus on a particular axis for more sensitivity. With our system, the user is able to specify which axis is to be used as input data for segmentation algorithm. It is also possible to use a sum of the axis.

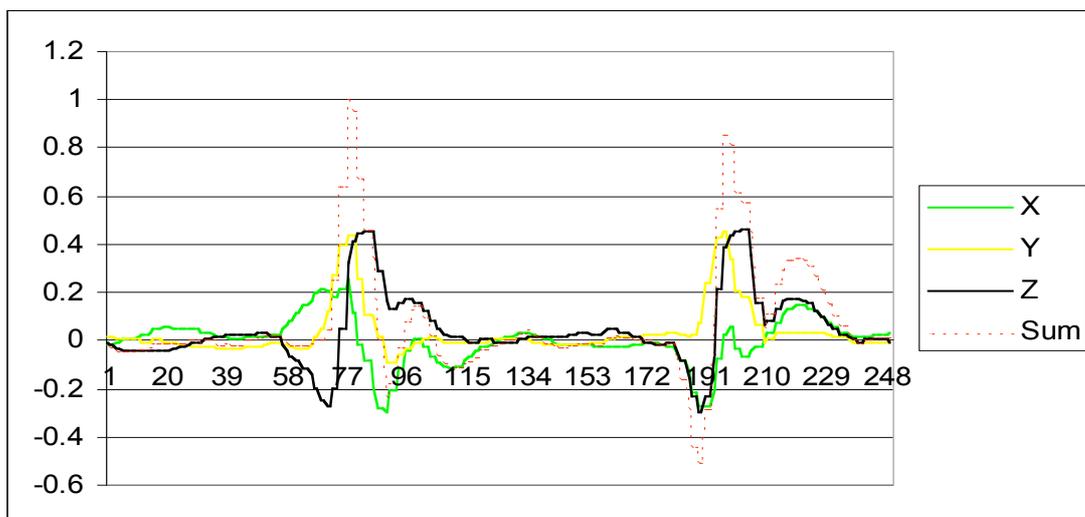


Figure 3. Shows peaks in acceleration data

The segmentation algorithm is used to locate beat points and use their position to segment the data. Firstly the input data is modulated so that all acceleration values are positive. The difference between the current and previous acceleration values is then calculated, this indicates whether the acceleration level is rising or falling. If this value changes in polarity it implies a peak has been found and thus a beat point has been located. A threshold level, which is user adjustable, is set to isolate the peaks from the data; only when the acceleration level is above the threshold level can a beat point be located. Once a beat point has been located, the algorithm is stalled so that no more beat points can be located within a certain time period; this removes any errors that occur due to counter accelerations producing secondary peaks. The stall period is determined according to the interval between the beat points; it is set to be equal to a quarter of the time interval between the previous two beat points.

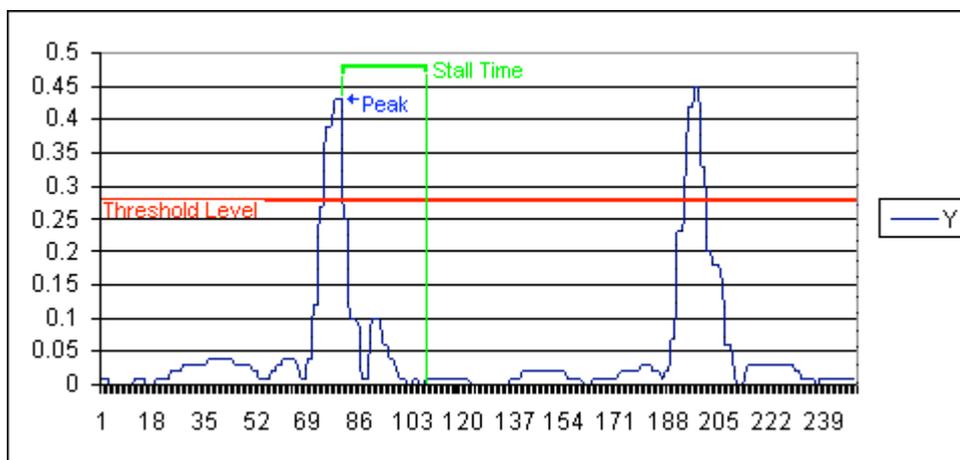


Figure 4. Shows threshold, peak and stall time

The data is consistent and very rarely are multiple peaks produced meaning there is no need for smoothing and other processing before the data can be used. This is very beneficial as it simplifies the algorithm and reduces latency. It was found that the algorithm works well when the beat points are prominent within the movement as they produce well defined peaks in the acceleration data; however, within legato passages the peaks became less well defined and thus often were not detected. A solution to this problem is to incorporate positional data; changes in direction can be detected within this data which can then be used in combination with the acceleration data in order to distinguish between accelerations caused by linear movement and accelerations caused by directional changes.

Calculating Tempo and Consistency

Instantaneous tempo is calculated from the time intervals between the segmentation points. Each value is then loaded into a distribution and a mean tempo value is calculated. The difference between the instantaneous tempo and the mean tempo is calculated allowing the standard deviation to be calculated; this shows the consistency of the beat points. The polarity of the difference calculation also gives an indication as to whether the user needs to speed up or slow down, thus enabling the user to improve their consistency rating. It is also possible to set a goal tempo; this is then set as the mean value and the same calculation is done as before. In this case the

consistency measurements indicate how accurately the user is able to stick to that tempo.

Path Comparison

Gesture path comparison is used to highlight the difference between two data sets. Firstly, the data is scaled before the comparison process in order to account for different user sizes then vector quantization is used to reduce the amount of data to contain only key data to changes in the path. As differing gesture sizes will affect the comparison process the segmentation points are used to synchronise the data, meaning that the comparison process is applied to each segment at a time. Once the data is fully prepared the difference between the data sets is calculated giving a rating indicating the similarity or difference between the two movements. The development and optimisation of the module represents the work in hand.

Expressivity tracking

Expressive tracking is a combination of both the tempo and path comparison algorithms. The quantised path data is segmented as in the path comparison algorithm; speeds and accelerations are then calculated relative to time. This data can then be assessed with regard to the tempo; this gives speed data relative to the tempo data. The speed is followed through out the segment indicating fast and slow areas. Also the consistency of the speed is also monitored. From the combination of this data it is possible to analyze the user's expressive intentions and detect musical characteristics such as legato or staccato. This is currently under development and evaluation.

Sonification

Sonification is used to present information regarding the user's movements in both a descriptive and informative manor. Sonification is used to both describe current movements and also influence the user's successive movements.

Sonification of beat points is implemented by using a short sine wave tone indicating when a beat point has been located. A metronome function is also provided which can provide a steady click if desired. The sonification of beat points allows the user to work more accurately with the metronome and keep the tempo steady. A guide tone to help improve tempo consistency can also be sonified using delays. The beat point pulse is delayed by a variable time; this is equal to the time of the previous interval between beat points. The user aims to conduct the next beat at the same time as the delayed guide tone, if successful, then their tempo is consistent. However, this approach will only sonify the consistency of adjacent time intervals; by using an average interval time value it is possible to sonify consistency over a larger period. The user is able to specify how many previous points are taken into account when computing the average. A different tone frequency is used in order to distinguish between the beat points and guide pulses.

CONCLUSION

This paper has given an overview of a system used to acquire and analyse a conductor's hand movements, aiming to provide an educational yet entertaining tool for learning and practising basic conducting techniques. Multiple wiimotes are used to track a conductor's performance and in testing have proven a success as they are able to provide an adequately accurate yet inexpensive and easily attainable controller for tracking motion. This is an ongoing project and some future directions include tracking of the second hand using a wrist mounted accelerometer to help attain greater understanding of the expressive content of the performance and even tracking balance shifts within a conductor's stance.

References

- [1] FRIBERG, A: Home Conducting – Control the overall musical expression with gestures. *Proceedings of the 2005 International Computer Music Conference*.
- [2] FRIBERG, A: A fuzzy analyzer of emotional expression in music performance and body motion. *Proceedings of Music and Music Science*, 2004.
- [3] HUNT, A Herman, T: The Importance of Interaction in Sonification. *Proceedings of the International Conference on Auditory Display, Sydney*, 2004.
- [4] ILMONEN, T and Takala, T: Conductor Following With Artificial Neural Networks. *Proceedings of the 1999 International Computer Music Conference*.
- [5] KOLESNIK, P Wanderley, M: Recognition, Analysis and Performance with Expressive Conducting Gestures. *Proceedings of the 2004 International Computer Music Conference*.
- [6] MATTHEWS, M: The Radio Baton and the Conductor Program, or: Pitch-the most important and least expressive part of music. *Computer Music Journal* 15(4), 37-46, 1991.
- [7] MATTHEWS, M and Moore, F: GROOVE-a program to compose, store and edit functions of time. *Communications of the ACM* 13(12), 715-721, 1970.
- [8] MARRIN, T and Pradiso, J: The Digital Baton: a Versatile Performance Instrument. *Proceedings of the 1997 International Computer Music Conference*.
- [9] MARRIN, T and Picard, R: The “Conductor’s Jacket”: A Device for Recording Expressive Musical Gestures. *Proceedings of the 1998 International Computer Music Conference*.
- [10] McELHERAN, B: *Conducting Technique for beginners and professionals*. Oxford University Press, 2004.
- [11] MORITA, H, Hashimoto, S and Otheru, S: A Computer Music System that Follows a Human Conductor. *Computer*, vol. 24, no. 7, 1991
- [12] MURPHY, D: Tracking a Conductor’s Baton. *Proceedings of the 12th Danish Conference on Pattern Recognition and Image Analysis*, 2003.
- [13] MURPHY, D, Andersen, T and Jensen, K: Conducting audio files via Computer Vision. *Proceedings of the 2003 International Gesture Workshop*, Genoa, Italy.