Musical Interaction with Hand Posture and Orientation: A Toolbox of Gestural Control Mechanisms

Thomas Mitchell
University of the West of England
Bristol, UK
tom.mitchel@uwe.ac.uk

Sebastian Madgwick
University of Bristol
Bristol, UK
s.madgwick@bristol.ac.uk

Imogen Heap
Megaphonic Records
London, UK
info@imogenheap.com

ABSTRACT

This paper presents a toolbox of gestural control mechanisms which are available when the input sensing apparatus is a pair of data gloves fitted with orientation sensors. The toolbox was developed in advance of a live music performance in which the mapping from gestural input to audio output was to be developed rapidly in collaboration with the performer. The paper begins with an introduction to the associated literature before introducing a range of continuous, discrete and combined control mechanisms, enabling a flexible range of mappings to be explored and modified easily. An application of the toolbox within a live music performance is then described with an evaluation of the system with ideas for future developments.

Keywords

Computer Music, Gestural Control, Data Gloves

1. INTRODUCTION

The use of hand tracking for computer interaction has formed a longstanding focus for research and investigation since the emergence of the earliest tracking devices in the 1970s [16]. A variety of approaches have been developed focusing on the acquisition and processing of hand gestures to bring our interactions with electronic devices closer to our natural interactions with non-computerised objects. The range of motion tracking technology available for this purpose can be broadly separated into two categories: methods relying on external apparatus and methods relying on wearable self-contained sensors. External apparatus is frequently required when tracking is performed using optical camera-based approaches [19]. Whereas self-contained methods generally rely on sensing devices which may be worn, often incorporating bend sensors and/or Inertial Measurement Units (IMUs) [14]. As well as enhancing conventional computer interaction, wearable motion capture technology has been widely adopted as a mechanism to enhance aspects of audio and music interaction. Notable examples include [10, 17].

The structure of a gestural musical instrument is frequently depicted with the components shown in Figure 1. In response to gestures at the system input, audio is produced at the system output. The sensing apparatus produces input data which is processed to produce control parameters before being translated into audio parameters via a mapping layer.

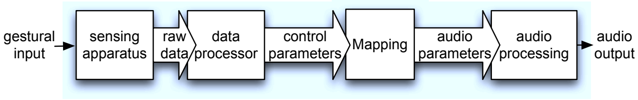

Figure 1: Gestural musical instrument structure

This paper describes a toolbox of simple control mechanisms for live music performance which are available when a data glove and Attitude Heading Reference System (AHRS) device are integrated to offer a self-contained wearable device monitoring finger flexion and hand orientation. The remainder of this paper sets out the sensing apparatus adopted herein, followed by an overview of the analysis algorithms which are used to extract gestural features from the streams of sensor data. An overview of control options afforded by the resulting data is then provided, combining divergent modes of control already extant in the literature with novel combinations to develop a flexible toolbox of control mechanisms which can be employed for live musical performance. The paper concludes with an evaluation of these control processes when incorporated within a live musical performance at the TEDGlobal 2011 conference.

2. SENSING APPARATUS

Data gloves provide a wearable, self-contained approach to capturing fine motor activity which are immune to occlusions and place minimal restrictions on the wearer's movements. Since the development of the first data glove there have been numerous examples of their utility within a variety of musical contexts including composition [8], percussion [6] and synthesis [13, 18].

Within this work, two off-the-shelf devices were employed: the 5DT 14 Ultra gloves [2] and the x-io Technologies x-IMU AHRS device [3], see Figure 2. It should be noted that the analysis and control mechanisms presented here are not limited to this hardware; any comparable devices are equally applicable.

Data Glove. The 14 Ultra device developed by Fifth Dimension Technologies incorporates 14 fibre optic bend sensors. The sensors are positioned at the metacarpophalangeal and proximal interphalangeal joints to measure finger flexion, and between the fingers and thumb to measure abduction/adduction. Frames of 12-bit samples for each bend sensor are continuously transmitted at approximately 60Hz via a wired (USB) or wireless (Bluetooth) connection.

Orientation Device. AHRS devices are self-contained units able to give an absolute measurement of orientation relative to the Earth coordinate frame. An AHRS device

NIME'12, May 21 – 23, 2012, University of Michigan, Ann Arbor.

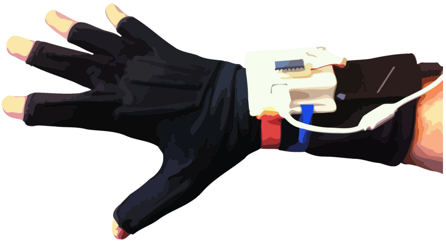

Figure 2: Data glove and orientation sensor

consists of a triple-axis gyroscope, accelerometer and magnetometer and a sensor fusion algorithm to combine the information provided by each sensor into a single estimate of orientation. The x-IMU is capable of transmitting instantaneous orientation values simultaneously and raw sensor data with rates of up to 512Hz via a wired (USB) or wireless (Bluetooth) connection.

3. A TOOLBOX OF GESTURAL CONTROL MECHANISMS

The raw data from the tracking sensors can be analysed to identify a meaningful set of gestural control mechanisms which may be subsequently mapped to control audio processing parameters. The control mechanisms which have been identified and implemented so far are categorised in to continuous, discrete or combined groups and are delineated throughout the remainder of this section.

3.1 Continuous Control

The data-glove and AHRS sensor provide a continuous flow of instantaneous finger flexion and orientation data which can be extracted directly as control parameters.

Flexion Control. To ensure parity between users it is important that calibration is performed to scale the raw sensor to a floating-point value in the range 0 to 1. To produce continuous control parameters by finger flexion, the sensor value for any chosen joint angle may be interpreted directly as a control parameter. Furthermore, sensor readings may be combined by taking the mean average flexion values for multiple sensor values.

Orientation Control. In gestural interaction, deictic or directional gestures referring to a point in space rely on the availability of orientation data. With the AHRS orientation datum set to a known direction (e.g. pointing towards the audience), the Euler angles ϕ, θ and ψ represent a set of continuous control parameters.

Positional Displacement. The raw inertial data transmitted by the AHRS device may be processed to provide an indication of relative angular rotation and linear translation using trapezoidal integration of the raw gyroscope measurements and double integration of the gravity compensated raw accelerometer values respectively.

3.2 Discrete Control Mechanisms

The control mechanisms identified above accommodate many useful modes by which audio parameters may be controlled continuously. Additionally, the continuous data streams may be analysed to identify discrete features resulting from a particular pose or gesture. Successful identification of these gestures leads to the introduction of mechanisms producing discrete control parameters which can later be mapped to control state information.

Posture Identification. When the wearer's hand assumes a particular posture, the flexion values sampled by the glove exhibit a unique pattern. Consequently, the problem of identifying a predefined set of postures from the glove data becomes a pattern recognition problem. A range of techniques are available to address this problem but for this work an artificial neural network has been shown to be reliable when gloves are removed/replaced and between different users [12].

Segmented Orientation. Division of the sensor orientation range into subregions enables the orientation to act as a discrete control mechanism. This can be computed using Euler angle segmentation or, to avoid problems associated with singularities, by defining rotation matrix segments.

Inertial Peak Detection. With access to the continuous flow of raw inertial sensor data, peak detection algorithms can be used to search for value fluctuations resulting from sharp changes in motion around or along the X, Y and Z sensor axes.

3.3 Combined Control Mechanisms

By combining the continuous and discrete control mechanisms set out above, a further range of gestural control options emerge which may combine both orientation and flexion data. These new control mechanisms are naturally more expressive/intuitive as they as they resemble 'real-world' control interfaces by incorporating the notion of metaphor [7]. Moreover, the combination of continuous and discrete control mechanisms facilitate the development of a state-based control system to enable one-to-many gestural mappings. In practice, this can be achieved by grouping audio controls into modes; this arrangement also has the added benefits of minimising the likelihood of unintentional interaction and providing space for (unmapped) ancillary/ performance gestures.

Ratcheting. Ratcheting is a combined control mechanism which enables a control value to be modified using a process that resembles the operation of a mechanical ratchet. The mechanism works by using the identification of discrete postures as an enabling mechanism for the traversal of a continuous control parameter or discrete control parameter using an estimate of relative angular displacement. For example, a clasping posture (Figure 4b), could be used to engage the addition of rotary displacement. To the wearer, this process is analogous to turning a rotary encoder. Alternatively, a two-fingered point posture may be used to enable and initialise the identification of subsequent swipe gestures using an estimation of positional displacement to produce a similar control mechanism.

Selective Orientation. Selective orientation is comparable with ratcheting, with postures used as an enabling mechanism for absolute orientation, rather than relative displacement. The continuous or segmented orientation of the hand produces a control parameter only when a specific hand posture is formed. For instance, a fist posture may be used to enable the control of an application parameter which is scaled to the θ Euler angle. This combined control mechanism produces a gestural metaphor for the act of pulling a lever: the hand only controls the 'lever' when a fist posture is assumed, at all other times the wearer is able to move freely.

Segmented Threshold Triggering. The combination of segmented orientation with inertial peak detection enables the orientation of the hand to be taken into account when a sharp change in motion is detected. The control message produced may be controlled by the angular region occupied by the hand. An obvious example of this combined control mechanism would be for the control of drums [6], where the orientation of the hand selects the drum sound and the 'strike' gesture invokes its playback. Using inspection of the rotation matrix, a bass drum could be triggered when the peak is detected if the hand points downwards, a snare drum selected if the hand points forwards and a high hat selected if the hand points upwards.

4. A MUSICAL APPLICATION

The toolbox of control mechanisms for a data glove and AHRS hand tracking apparatus was developed in anticipation of a six minute performance to be made by the composer/performer Imogen Heap at the TEDGlobal2011 conference in Edinburgh, UK [1]. Due to the narrow timeframe within which the collaborators could meet to build the system, the toolbox was conceived to accelerate the development of a gestural performance system for which only the specifics of the hardware were known in advance and the gestural mapping and audio processes were to be established later. Based on our experiences developing 'Soundgrasp' [12] it was clear that the most effective mappings emerge when the performer drives the design process. The system architecture was thus designed to produce flexible mappings which could be modified easily.

System Overview. Figure 3 shows the major components of the hardware and its associated software. Each hand is fitted with a 5DT data glove, an x-IMU, an LED module and a lavalliere microphone. A third voice microphone is also included. The auxiliary port on each x-IMU is used to control a set of RGB LEDs to act as a primary source of feedback [18].

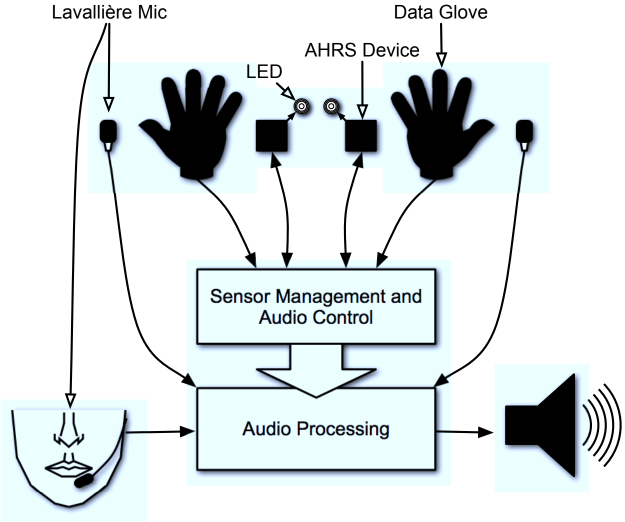

Figure 3: Gestural device components

As shown, the software consists of two distinct parts: a sensor/audio control application and an audio processing application. Communication between the applications is made via Open Sound Control (OSC) [20]. The sensor/audio control application is a C++ application written using the libraries Juce [15] and oscpack [4]. This application implements the control mechanisms set out above and transmits commands controlling the state of the audio processing application. The environment for audio processing was developed in Max/MSP to enable a range of audio recording, looping, synthesis and modification functions to be prototyped rapidly.

Mapping. In addressing questions relating to appropriate mappings [7], one approach is to analyse the natural movements emerging when subjects are asked to gesticulate while listening to music/sound [9]. Other practitioners place emphasis on including the performer in the design process [5]. The development of the control mechanism to audio parameter mapping was an iterative process directed primarily by the performer with input from co-authors and colleagues. With the requirement for a wide range of audio control options, it was clear that the audio parameters would have to be organised into modes and accessed via a state based system.

Central to the state control of the application was the discrete posture identification control mechanism. As it was easy to configure the neural network to accommodate a range of distinguishable postures, the performer was free to develop her own posture set. For all modes of control, only four postures were required for the performace, as shown in Figure 4.

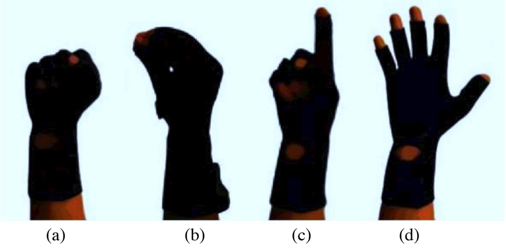

Figure 4: Chosen posture set

With the discrete posture control mechanism established, mode selection was performed using a selective orientation control gesture where the ϕ Euler angle was segmented into five regions, one for each mode: *voice*, *wrist*, *effects*, *synthesis* and *drum*. Mode selection was performed only when the left hand was open and the right hand formed a fist. This gesture was chosen because it is simple to perform and unlikely to occur incidentally. Each mode was ascribed a colour which was displayed on the left and right hand LED modules to provide feedback.

In *voice* mode, the vocal microphone signal could be recorded or overdubbed into a two channel looper. This mode was controlled using a simple grasping gesture where record was enabled on the identification of an open hand posture (Figure 4d) and disabled at all other times, an idea described previously in [12]. In *wrist* mode, the audio input received from the left and right wrist microphones can be recorded into a separate stereo looper. This enables the performer to play and record acoustic instruments where the record state was toggled when a rotational spike was detected around the axis of the right wrist.

In *effects* mode, continuous control gestures with the right hand applied effects to the output mix of the looper and left hand gestures applied effects to the live vocal input. Reverberation and panning are controlled by the Euler angles θ and ψ. Furthermore, filtering was applied using the mean average of the finger flexion sensors for each hand. The wrist flick, as described in the *wrist* mode section, toggled the recording of automation for each of the effects.

The *synthesiser* mode used a combined control mechanism in which segmented orientation of the θ Euler angle selected the current note and the posture identification of an open hand was used to trigger note playback on a software synthesiser. Similarly, in *drum* mode, sounds were triggered with the identification of peaks in the raw inertial sensor data, with the selection of the drum sound controlled by inspecting the rotation matrix. In both modes recording was toggled with the formation of a fist with the right hand.

5. APPLICATION EVALUATION

The development of the system was complete and stable and behaved as expected on the day of the performance where a positive reaction was received. However, the developmental process highlighted several notable points for consideration and identified areas for future development.

While the toolbox of control mechanisms was sufficient to implement the majority of gestures/mappings requested by the performer, it was not possible to accommodate them all. For example, the sensor apparatus was unable to track the positioning of the hands with respect to the body. By placing additional orientation sensors at multiple points on the upper body, the relative 3D positions of the arms may be tracked, an approach described in [11].

During the iterative development of the gesture to audio mappings, several areas were found to be problematic. For example, the performer wanted to engage record mode while playing the piano or Array mbira. After several attempts with other control mechanisms, rotational peak detection was found to be the most appropriate and with practice could be executed efficiently. However, on several occasions this gesture would cause the gyroscope sensor to saturate (2000°/s), which, over a sustained period, would cause the AHRS readings to drift.

Initial plans sought to map only those audio parameters that the performer used within her previous performances. However, with the capacity for gestural control, simple audio controls provided greater appeal than they would otherwise. Panning, for example, when controlled directly by pointing, was qualitatively regarded to be an effective and engaging mapping.

6. CONCLUSIONS

This paper set out a toolbox of control mechanisms for gestural music control in scenarios where finger flexion and hand orientation data is available. A selection of continuous, discrete and combined control mechanisms have been organised and presented which were developed in anticipation of a live performance to enable the rapid development of a gestural mapping system. The implementation details for the control mechanisms have been provided followed by an example application in the form of a live musical performance at the TEDGlobal2011 conference in Edinburgh. With easy access to a diversity of control mechanisms usable and robust gesture to sound mappings were quickly developed in collaboration with the performer.

7. ACKNOWLEDGMENTS

The authors would like to thank Kelly Snook, Professor Tony Pipe, Professor Chris Melhuish and all colleagues at the Bristol Robotics Laboratory.

8. REFERENCES

[1] Tedglobal, conferences.ted.com/TEDGlobal2011/.

[2] Fifth dimension technologies, www.5dt.com, 2012.

[3] x-io technologies, www.x-io.co.uk, 2012.

[4] R. Bencina. oscpack, www.rossbencina.com/code/oscpack, 2012.

[5] M. A. Bokowiec. V'oct (ritual): An interactive vocal work for bodycoder system and 8 channel spatialization. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2011.

[6] S. Chantasuban and S. Thiemjarus. Ubiband: A framework for music composition with BSNs. In *Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, 2009.

[7] S. Fels, A. Gadd, and A. Mulder. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound*, 7(2):109–126, 2002.

[8] H. Ip, K. Law, and B. Kwong. Cyber composer: Hand gesture-driven intelligent music composition and generation. In *Proceedings of the 11th International Multimedia Modelling Conference*, 2005.

[9] M. Leman, F. DESMET, F. Styns, L. Van Noorden, and D. Moelants. Embodied listening performances reveal relationships between movements of player and listeners. In *Proceedings of ENACTIVE*, 2007.

[10] P.-J. Maes, M. Leman, K. Kochman, M. Lesaffre, and M. Demey. The "one-person choir": A multidisciplinary approach to the development of an embodied human-computer interface. *Computer Music Journal*, 35(2):22–35, June 2011.

[11] P.-J. Maes, M. Leman, M. Lesaffre, M. Demey, and D. Moelants. From expressive gesture to sound. *Journal on Multimodal User Interfaces*, 3(1), 2010.

[12] T. Mitchell and I. Heap. Soundgrasp: A gestural interface for the performance of live music. In *International Conference on New Interfaces for Musical Expression*, 2011.

[13] G. Saggio, F. Giannini, M. Todisco, and G. Costantini. A data glove based sensor interface to expressively control musical processes. In *4th IEEE International Workshop on Advances in Sensors and Interfaces*, 2011.

[14] M. Sama, V. Pacella, E. Farella, L. Benini, and B. Riccó. 3dID: a low-power, low-cost hand motion capture device. In *Proceedings of the conference on Design, automation and test in Europe: Designers' forum*, 2006.

[15] J. Storer. Juce, www.rawmaterialsoftware.com, 2011.

[16] D. Sturman and D. Zeltzer. A survey of glove-based input. *Computer Graphics and Applications, IEEE*, 14(1):30 –39, jan 1994.

[17] T. Todoroff. Wireless digital/analog sensors for music and dance performances. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2011.

[18] M. Wanderley and P. Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632 – 644, April 2004.

[19] R. Y. Wang and J. Popović. Real-time hand-tracking with a color glove. *ACM Transaction on Graphics (SIGGRAPH 2009)*, 28(3), 2009.

[20] M. Wright. Open sound control: an enabling technology for musical networking. *Organised Sound*, 10(3), 2005.