Sonification of Gestures Using Specknets

Vangelis Lympouridis, Martin Parker[*], Alex Young and DK Arvind[†]
[*] School of Arts, Culture and Environment, University of Edinburgh, Scotland, UK.
vangelis@lympouridis.gr | mparker@staffmail.ed.ac.uk
[†] Research Consortium in Speckled Computing, School of Informatics, University of Edinburgh, Scotland, UK.
a.d.young-2@sms.ed.ac.uk | dka@inf.ed.ac.uk

***Abstract* — This paper introduces a novel approach to gesture recognition for interactive virtual instruments. The method is based on the tracking of body postures and movement which is achieved by a wireless network of Orient-2 specks strapped to parts of the body. This approach is in contrast to camera-based methods which require a degree of infrastructure support.**

This paper describes the rationale underlying the method of sonification from gestures, addressing issues such as disembodiment and virtuality. A working system is described together with the method for interpreting the gestures as sounds in the MaxMSP tool.

I. INTRODUCTION

There is a strong relationship between physical movement of the player and the generation of music. According to Robert Hatten: *"Musical gesture is biologically and culturally grounded in communicative human movement. Gesture draws upon the close interaction of a range of human perceptual and motor systems to synthesize the energetic shaping of motion through time into significant events with unique expressive force." (Hatten Robert, 2003)* [1].

The sound generated by human gestures is an interplay between the physical body, the data body and sound. Data body was a term introduced by Sybille Kramer to address the virtual body as a physical body in a data form that can be manipulated[3].

Jin Hyun Kim and Uwe Seifert have suggested the following role of the physical body in sound generation: *"Some strategies of integrating the physical body in the context of interactive algorithmic sound synthesis give rise to an extension and immaterialization of the physical body that seems to deviate from 'embodiment'."* [2].

In the proposed approach the movement manifests itself through sonic or possibly visual feedback. The neutral ground where the exchange takes place between physical and virtual worlds defines the area that has to be explored and better understood. Such interactive systems introduce a new reflexive relationship of the body with itself that is revealed through the virtual body: the body becomes both a subject and an object by affecting and being affected by its virtual sound environment [4].

II. METHODS FOR GESTURE RECOGNITION

There are two well established methods for gesture recognition. The first one is based on the use of cameras to track movement in a multidimensional space. Examples of this method range from David Rokeby's Soft VNS [5] to multi-camera motion capture systems. It can be argued that this approach does not directly refer to the physiology of movement, but to a projection of the movement onto a 2-D plane based on the camera angle. Several cameras can be used simultaneously to extract information in the third dimension, and libraries have been developed for analysis and synthesis (e.g. the MEGA project [6]), and the underlying method is therefore primarily a vision-based approach rather than a motion-based one.

The second approach defines a space with sensors that detects movement within it. This has been achieved successfully with infrared sensors and ultrasound beams, but have limitations similar to the camera. Movement is located within the fixed space, the data is low resolution in detail and the analysis is persistently at one removed from the gesture, i.e. the gesture itself is not analyzed but its impact in a space is.

It has been possible to attach sensors to the body, which are typically wired. Our method for tracking movement is based entirely on a wireless network of sensors attached discreetly to the limbs of the performer which detects the actual movements of the performer.

III. DESCRIPTION OF THE ORIENT-2 BASED SYSTEM

The new approach tracks in real-time the orientation and movement of parts of the human body with a network of wireless Inertial Navigation Units (INU) called Orient-2 specks [7] developed by the Research Consortium in Speckled Computing at Edinburgh. Individual devices track their own orientation relative to a global Earth-fixed reference co-ordinate frame. Orientation updates are gathered to a central point using a Time-Division, Multiple-Access wireless network. Complete frames of orientation data (comprising data from all devices) are applied to a body model to produce spatial data. In this manner six Degree of Freedom (6DOF) spatial and rotational data corresponding to body posture are produced.

The Orient-2 device is equipped with a 16-bit microprocessor, 250kbps radio, 3-axis accelerometer, two 2-axis magnetometers and three rate gyroscopes. The device measures 36x28x11mm, which includes a single-cell 120mAh lithium-polymer battery (Fig.1).

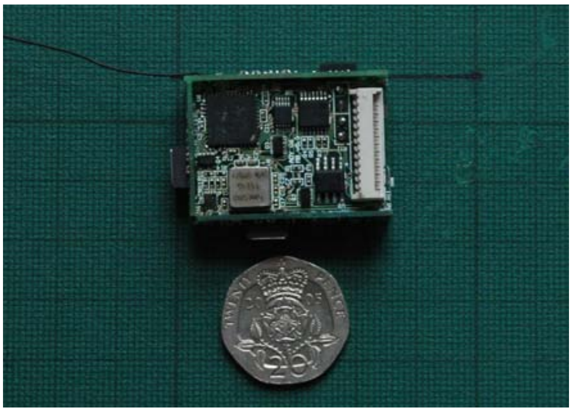

Fig.1 The Orient-2 speck prototype

Nine sensor inputs are sampled at a rate of 256 times per second, corresponding to accelerometer, magnetometer and gyroscope readings in three axes. Orientation estimation is performed on the device using a complementary filter algorithm combining an inertial estimate, generated by integrating the outputs of the rate gyroscopes, with an absolute estimate, formed from the magnetic and static acceleration vectors. The complementary blending of the two estimates results in zero-delay in the output estimate, while maintaining the high frequency response of the gyroscope estimate and the low frequency accuracy of the accelerometer and magnetometer estimate.

The output of the filter is in the form of a quaternion [8], which are 4-dimensional complex numbers combining a real part and an imaginary vector. They are used in 3-dimensional computer graphics thanks to their compact representation compared to equivalent rotation matrices and suffer less from rounding errors.

Performing orientation estimation on the device provides substantial advantages compared to off-loading raw data directly to a PC. These include:

1: Reduction in data transfer - lower bandwidth requirements allow for more devices per channel and lower power consumption per device.

2: Reduction in estimate error - as data is integrated on the device; no error is introduced due to radio packet losses.

A. Network

Each single receiver can be used to interface to multiple devices operating on a shared radio frequency channel. This is achieved by a Time-Division, Multiple-Access (TDMA) Medium Access Control (MAC) layer. Each device is allocated a slot within a TDMA frame, and is able to transmit free from interference from other devices. At the start of each frame a synchronisation packet is sent allowing all devices to resynchronise their clocks.

In each frame, each device transmits its most recent orientation estimate as of the time of the last synchronization packet. This introduces a minimum lag equal to the TDMA frame period. A typical example, tracking the arms and shoulders, would be 7 slots of 1ms, plus a 1ms synchronisation slot yielding a 8ms time lag. In this manner a consistent model of the body posture at a particular instant of time can be constructed.

B. Body Model

The Orient-2 specks provide information about the orientation and acceleration, but they do not directly provide spatial data. Translation can be estimated by integrating the measured acceleration vector in the Earth-fixed frame. However, this will suffer from drift errors that will quickly accumulate.

An alternative route to produce spatial data is to attach the speck output to a rigid-body model of the performer. The performer is modeled as a set of joints, modeled as 3-D points, connected by rigid bodies, modeled as 3-D vectors. By attaching a speck to each limb segment of the performer, the orientation of a joint and its associated bodies can be calculated. This in turns allows the calculation of the joint positions through simple vector arithmetic.

A complete model of the performer can be realised as a tree structure of connected joint objects. Each joint object has a position, an offset from its parent and an orientation. To produce each frame of motion data, the tree is traversed starting from a root joint, typically the hips or chest. As each joint is processed, it updates its position by rotating its offset vector into the parent co-ordinate frame and adding the resultant vector to its parent's position.

C. Working with Quaternions and 3D rotations in MaxMsp

Parameters which characterise the physical properties of movement can be extracted by applying mathematical functions on the incoming quaternion from the Orient-2 specks.

For example, by combining the two quaternions from the specks on the upper- and lower-arm and then extracting the Euler angles, the angle created by the limbs can be calculated. A possible use could be in the creation of a virtual trombone.

Performing measurements of orientations of the human body using quaternions depends on the default orientation of the devices providing the rotation information. This means that a rotation of a body in the *z-axis* will lead to a new arrangement of the joints in space. This could be useful for applications related to the theatre which rely on the position of the body in space. When creating an accurate virtual instrument, there has to be a method for providing the relative position for all the body parts, fixed in relation to the environment. One can imagine as if a camera is fixed to the body and rotating with it.

This can be achieved by attaching a speck on the chest or at the back of the body. The conjugate quaternion, q^*, is extracted from this speck which is combined with all the other quartenions from our specks. In this way all relative positions of the rigid body remain analogous to the rotation of the body in space [8].

It is also useful to obtain accurate measurements of the velocity of each part of the body. This is achieved by refering the quaternion to a spatial vector, and then subtracting the last two values (i.e., the previous two positions). As the devices are polling at a high rate, this measurement is accurate even for very fast movements. Another value that can be measured is the acceleration of the specks in three axes which yields the actual acceleration regardless of the orientation of the movement.

A python library has been created to calculate these functions in the MaxMSP environment [9]

IV. TESTING A SIMPLE VIRTUAL INSTRUMENT

The modeling of the movement of the upper body with rotation of vectors in 3-D space is achieved by collecting the quaternion data from each one of the Orient-2 devices. This data is used for virtual reconstruction of the human body in space with single 3-D vectors and the physical movement is modeled in MaxMsp where coordinates of the moving limbs are calculated (see Figure 2).

The user can then specify the dimensions and resolution of a virtual 3d grid. The coordinates of all the junctions of the grid are then stored into a Matrix using the external FTM libraries (IRCAM) for MaxMsp.

Classification using the Mahalanobis algorithm is then performed for the floating points at the end of each hand. In this way the user can inform the system of an interaction with the virtual areas that were modeled and have been arranged in space.

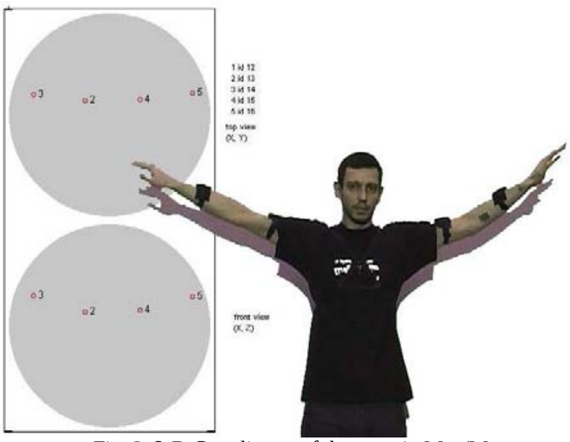

Fig. 2. 3-D Coordinates of the arms in Max/Msp

A. *Methods for Music Synthesis*

The position of hands in different planes, (x,z) (x,y) and (y,z), are mapped to two Frequency Modulation (FM) synthesizers. Physical relationships to understanding the sonic event is constructed using elementary rules. For example, the carrier frequency is related to the position of the hand in the (x,z) plane, which provides lower notes as the user moves from the right to the left and from the top to the bottom (for the left hand) giving the lowest note at the bottom left and the highest note at the top right limits (video documentation)[10].

There is an element of training (mapping) the gestures to the sound modules, which together with data filtering, results in a smooth dialogue. Simple reverberation is then applied to the generated frequencies in order to smooth the transition of data into the virtual space. The reverb has been crudely mapped to rate of movement of the gesture - slower movement extends reverb time, while faster movements create more rhythmic events.

Another synthesis method that has been tested is Scanned synthesis [11]. This offers the conceptual framework to read sounds in a dynamic wavetable based on slow movements in mechanical systems [12]. Movements with rates below 15Hz. relate to expressive gestures within this low rate, which are expressed as vibrating higher rate frequencies.

Physical models of strings are being developed to produce better sound quality and clarity.

V. CONCLUSION AND FUTURE DEVELOPMENT

The system has been tested on basic music synthesis models but the full potential of the system for generating highly responsive visuals and sound is yet to be explored.

The processing power is very limited as the system uses most of its resources to model the virtual body. This limitation is overcome by employing low level programming in Python so that more complex and advanced synthesis techniques and concepts could be implemented and tested.

Ideas for future work include themes such as kineasonics that arise from the study of rhythm and the brain. [12].

Finally, although a full-body tracking system involving sixteen Orient-2 specks has been tested successfully, processing power for sonification and stability issues have still to be resolved.

ACKNOWLEDGMENT

This work was supported by the Scottish Funding Council under the Strategic Research Development Grant # R37329 from the Scottish Funding Council, and UK Engineering and Physical Research Council under the Basic Technology Research Programme, Grant # C523881.

REFERENCES

[1] Hatten,R. “*Musical Gesture: Theory and Interpretation*”, Indiana University Press, 2003.

[2] Kramer S.,“*Maschinen, Medien, Performances. Theater ander Schnittstelle zudigitalen Welten*”, 2001, p. 471.

[3] Jin HyunKim & Uwe Seifert,“ *Embodiment: The Body in Algorithmic Sound Generation*”, Contemporary Music Review, Vol.25, 2006.

[4] Jin HyunKim & Uwe Seifert,“ *Embodiment: The Body in Algorithmic Sound Generation*”, Contemporary Music Review,Vol.25,2006

[5] Cooper D. “*Very Nervous System*”, Wired Magazine issue 3.03 , Rokeby D. “*VNS*” http://homepage.mac.com/davidrokeby/vns.html

[6] Lindstrom E., Camurri A., Friberg A., G.Volpe, and Marie-Louise Rinman “*Affect, Attitude and Evaluation of Multisensory Performances*” Journal of New Music Research 2005, Vol. 34, No. 1, pp. 69– 86

[7] Young, A., Ling, M.J., Arvind, D.K. “*Orient-2: A wireless, real-time posture tracking system using local orientation information*”, Proc. 4th Workshop on Embedded Networked Sensors (EmNets 2007), Cork, Ireland, June 2007.

[8] Kuipers J. “*Quaternions and Rotation Sequences*” Princeton University Press, 2002

[9] py/pyext - python script objects for PD and MaxMSP Copyright (C) 2002-2005 Thomas Grill http://grrrr.org/ext/py/

[10] http://xenakis.caad.ed.ac.uk/mscpages/s0566966/thesis/videos/lympouridis_hi.zip

[11] http://www.lma.cnrs-mrs.fr/~IM/Projets/en_scansynth_main.htm

[12] Jean-Michel Couturier. " *A Scanned Synthesis virtual instrument"* In Proceedings of the2nd International Conference on New Interfaces for Musical Expression (NIME02), 2002

[12] Thaut, M. "*Rhythm, Music and the Brain*" Routledge, London 2005