A MODEL-BASED SONIFICATION SYSTEM FOR DIRECTIONAL MOVEMENT BEHAVIOR

Pieter-Jan Maes

IPEM
Dept. of Musicology
Ghent University, Belgium
pieterjan.maes@UGent.be

Marc Leman

IPEM
Dept. of Musicology
Ghent University, Belgium
marc.leman@UGent.be

Micheline Lesaffre

IPEM
Dept. of Musicology
Ghent University, Belgium
micheline.lesaffre@UGent.be

ABSTRACT

Computational algorithms are presented that create a virtual model of a person's kinesphere (i.e. a concept of Laban denoting the space immediately surrounding a person's body and reachable by the upper limbs). This model is approached as a virtual sound object/instrument (VSO) that could be "played" by moving the upper limbs in particular directions. As such, it provides an alternative for visual qualitative movement analysis tools, like bar plots.

This model-based sonification system emphasizes the role of interaction in sonification. Moreover, this study claims that the integration of intentionality and expressivity in auditory biofeedback interaction systems is necessary in order to make the sonification process more precise and transparent. A method is proposed – based on the embodied music cognition theory – that is able to do this without disclaiming the scientific, systematic principles underlying the process of sonification.

1. INTRODUCTION

The intent of this study is to develop a real-time computational method for the sonification of the way a person is moving the upper limbs in the space immediately surrounding the body (i.e. kinesphere). The movement feature that is of particular interest is the direction in which the different parts of the upper body are moving in reference to a person's torso. The method emphasizes the role of interaction in sonification. To describe the type of interaction that is facilitated by the system, we refer to the different categories of interactive sonification outlined by Hermann [6]. The type of interaction that comes closest to the one characterizing the presented method, is denoted with the term *auditory biofeedback*. In this type of interaction, the user is actively involved in generating and controlling the input data for the sonification system. The data specifying the movements of the upper limbs is delivered in real-time to the sonification system by an inertial sensor system attached to the upper body of the user. The actual algorithm performing the sonification consists of different sub-algorithms: (1) a method for calculating the position of the upper limbs in reference to the chest, (2) a smoothing Savitzky-Golay FIR filter, (3) an algorithm for calculating the direction of movement of a particular point of the upper limbs, (4) a method to virtually model and segment the user's kinesphere, and (5) the creation of a model-based sonification system which uses the virtual kinesphere as a virtual sound object (VSO) controllable by directional movement behavior.

2. TECHNICAL DESIGN

An inertial sensor system is used to sense the movement behaviour of the upper body in real-time. We made the choice to use the custom-made HOP inertial sensor system produced by the Centre for MicroSystems Technology (http://www.cmst.be/) at Ghent University [7]. Five HOP sensor nodes are attached to the different rigid bodies constituting the upper body; one on the torso, two on the upper arms, and two on the forearms. As such, the orientation of each rigid body in reference to an earth-fixed reference coordinate system is obtained. These signals are then inputted in the model presented by Maes [10] to calculate the position of the elbows and wrists in reference to a right-handed coordinate system with a relative origin located at the middle of the torso.

2.1. Savitzky-Golay filter

A first problem that we encounter when using movement data originating from an inertial sensing system is the presence of random high-frequent noise in the signal. Because of the fact that the direction of movement will be calculated from sample to sample, the high-frequent noise will result in an unstable and fluctuating output deforming the actual direction of motion. To avoid this problem, a Savitzky-Golay FIR smoothing filter was developed in Java and further implemented as a Max/MSP *mxj*-object. It facilitates a real-time smoothing device that removes high-frequent noise in the signal specifying the position of a point of the upper body from which we want to estimate the direction of movement.

The central occupation of the Savitzky-Golay filter is the computation of a polynomial fit to the data inside a specified frame window around each incoming data point. This fitted signal is expressed as a polynomial function (see Equation 1) of a specific order and from which the polynomial coefficients are computed by the Least Square Error (LSE) estimation method.

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_2 x^2 + a_1 x + a_0 \quad (1)$$

To optimize the results, the LSE estimation can be weighted with a rectangular, triangular, hamming or Blackman weighting vector.

The development of the Max/MSP object is conceived in a way it enables the user to manually configure the polynomial order, the frame size and the type of weighting vector specifying the filter. This is of particular interest because of the fact that (1) different inertial sensor systems could differ slightly in the noise they produce and (2) the type of movement that is sensed.

The Savitzky-Golay FIR smoothing filter has some interesting advantages over other types of filters, like the linear moving average and the IIR filters. The Savitzky-Golay filter preserves – in contrast to moving average filters – much more the spatial characteristics of the original data, like the widths and heights of peaks. Compared to an IIR filter, a FIR filter is much more stable, which is essential in real-time environments.
However, a trade-off for smoothing the original signal with a Savitzky-Golay filter is the occurrence of a delay in the smoothed signal. This delay results from the fact that the polynomial fitting is computed on the basis of values that come after the point of interest. Expressed in terms of milliseconds, the amount of delay is equal to:

$$\text{delay}(ms) = \frac{(f-1)/2}{Fs} * 1000 \qquad (2)$$

So, when working at a sample rate (*Fs*) of 100 ms with the default value 5 for the frame size (*f*), this results in a delay of 20 ms which is acceptable for real-time performance.

2.2. Calculation of orientation

The algorithm that is presented in this section calculates the direction of movement executed by the two wrists and elbows in reference to the body's centre of gravity (i.e. the middle of the torso). The smoothed positional *(x,y,z)* coordinates of the two wrists and elbows – outputted at a rate of 100 Hz – are taken as input of this algorithm. The direction of movement is represented at each instance by a vector drawn between the 3D position of each incoming sample and the successive sample. According to the model of Maes [10], the direction vector is defined in a right-handed coordinate system of which the origin is located at the position of the chest (see Figure 1). This is particularly convenient in the light of Laban's opinion that all directional energy irradiates from the chest and must as such be determined in relation to this centre of gravity. By calculating the 4-quadrant inverse tangent (i.e. *atan2* method in Java's *Math* class), each direction vector is expressed in terms of its spherical coordinates (see Equation 3).

$$\begin{aligned} S &= \sqrt{x^2 + z^2} \\ \theta &= \text{atan}\,2(y,S) \\ \phi &= \text{atan}\,2(z,x) \end{aligned} \qquad (3)$$

In the specification of the spherical coordinates, we follow the conventions outlined by Dray [3]. The angle in the vertical, XY plane (i.e. elevation) is specified by the theta (Θ) value expressed in radians, while the angle in the horizontal, XZ plane (i.e. azimuth) is defined by the phi (Φ) value expressed in radians. The azimuth expresses the angle in reference to the X-axis. The direction pointed by the Y-axis has an azimuth value of $\pi/2$ radians. The negative X-axis direction has maximum of π. The negative Y-axis has an azimuth value of $-\pi/2$ radians. The elevation expresses the difference in angle of a vector with the reference XY-plane. The Z-axis direction accounts for the maximal elevation value of $\pi/2$ radians. The opposite direction accounts for $-\pi/2$ radians (see Figure 2).

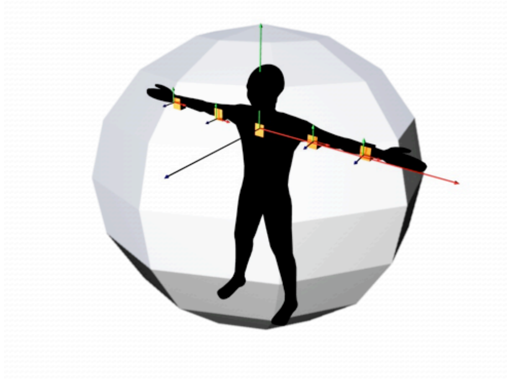

Figure 1. *Representation of the coordinate system defining the (1) position, (2) direction of motion, and (3) virtual kinesphere.*

2.3. Virtual kinesphere model

Laban's notion of kinesphere is used to indicate the imaginary sphere-like space immediately surrounding the human body and reachable by the limbs [5]. This section proposes a method to virtually model this kinesphere and subdivide it into different *directional segments*. The virtual kinesphere is represented in the same coordinate system that was used to define the orientation vector (see Section 2.2.). The method is developed in Java and implemented as a Max/MSP *mxj*-object. The amount of segments (i.e. resolution) could be determined by the user by way of an argument. Each segment is labelled with a number and defined in terms of a unique pair of spherical coordinates specifying the maximum and minimum azimuth/colatitude values.
The virtual sphere can now be approached as a virtual sound object (VSO) by attaching sounds to the different segments. Each sound can then be triggered and controlled by directing the movements of a specified part of the upper body to the corresponding segment. Before we go deeper into how the VSO is configured, we present how it is visualized.

2.3.1 2D and 3D visualization

For the 2D visualization (see Figure 2), a four-plane matrix is created with a resolution of *n*-by-*n* cells. Each cell corresponds to a segment of the VSO. The cell of the matrix that corresponds to the label attached to the segment towards a movement occurs, is coloured with a user-specified ARGB colour. The other segments, where no movement is directed towards, are coloured black. The user can put a command inside each cell of the matrix indicating to which sound process or sound sample it is mapped.

For the 3D visualization (see Figure 2), we use the OpenGL (http://www.opengl.org) implementation in Jitter. OpenGL is a widely used 2D and 3D graphics application programming interface (API). With the *jit.gl.gridshape* object, a 3D sphere object is created with a resolution of *n*-by-*n* segments. This sphere is the actual virtual representation of the human's kinesphere. Then, the 2D matrix specified in the previous paragraph is used to colour the particular segment in which the direction of movement occurs. This is done with the *jit.gl.texture* object. Again, comments could be added in order to specify which sample or sound process is coupled to a specific segment. The 3D sphere could be rotated in a way it coincides with the perspective of the user creating a virtual model that helps the user to explore his own, real kinesphere.

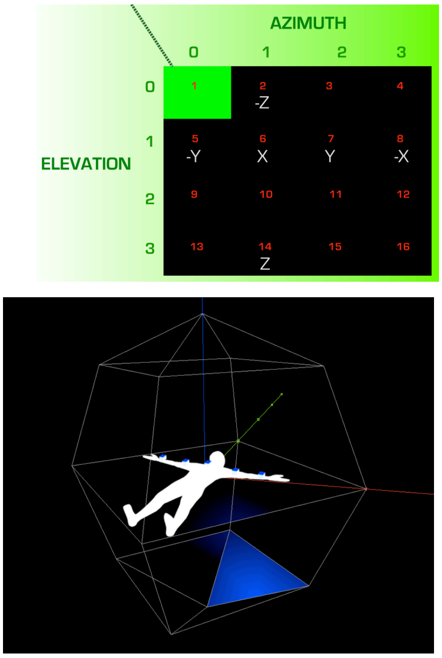

Figure 2. *2D (top) and 3D (bottom) visualization of the VSO.*

2.4. Configuration of the VSO

The method used by the VSO to turn incoming movement data into sound is standardized: by directing a particular body part towards a particular segment of the surrounding kinesphere, it is possible to trigger the sound synthesis process attached to the corresponding segment of the VSO. Nonetheless this standardized method, there are some dynamic features implemented in the VSO. First, the number of segments of the VSO could be changed. Second, different parameter mappings are possible specifying what kind of sonic process is assigned to a particular segment of the VSO. It can be sound synthesis parameters, sound control parameters and/or sound sampling parameters. The system that is presented in this paper takes advantage of this dynamic approach. As we will explain, this will enable the integration of the aspect of intentionality in the process of sonification without disclaiming the scientific and systematic aspect of sonification. But before we come to that part, we present a version of the VSO based on additive synthesis techniques that could be considered as a sonic alternative for the qualitative, visual data observation.

2.4.1. The additive synthesis model

The model presented in this section is used to sonify the complexity of directional movement behaviour performed by the upper body. The sonification is based on traditional additive synthesis techniques facilitating the creation of complex sounds and timbres according to the addition of sinusoidal waveforms. Each segment of the VSO is assigned to a different, pure sinusoidal tone (i.e. frequency) that could be triggered in the way specified in the previous paragraph.

Two configurations are proposed in this study. A first one presents an offline process for the sonification of the directional movement behaviour of only one point of the upper body (e.g. the right wrist). For a recorded movement trajectory of *n* samples specifying the position of the wrist, *n-1* direction vectors can be calculated (see Section 2.2). Then it is calculated how many times the *n-1* direction vectors intersect each of the segments of the VSO. A number is assigned to each segment representing the number of times it is being crossed during the performed movement trajectory. All numbers were then normalized between 0 and 1. The sonification exists in the activation of all sinusoidal waveforms attached to the VSO segments with an amplitude that corresponds with the normalized number representing how many times each segment was crossed. Now, if the performed movement behaviour was homogeneous, in the sense that it was dominated by the same repeated directional pathways over and over again, the corresponding sonification is also homogeneous, in the sense that the sound is dominated by a few number of frequencies. This could be compared with a high, narrow peak in a plot visualizing the statistical distribution of the frequencies in the spectrum (see Figure 3).

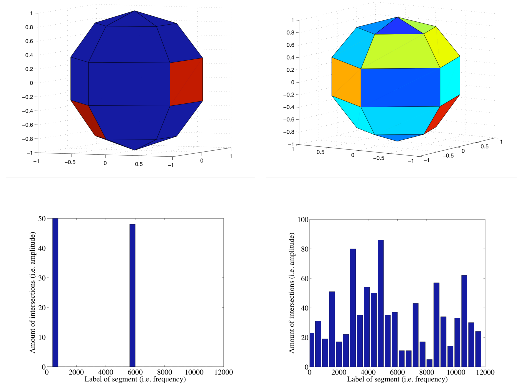

Figure 3. *The offline sonification method of a simple (left) and complex (right) directional gesture.*

Moreover, the chosen resolution can be compared to the bin size characterizing a data histogram. The more VSO segments (and attached frequencies), the more fine-grained the movement behaviour can be sonified (compare with Figure 4).

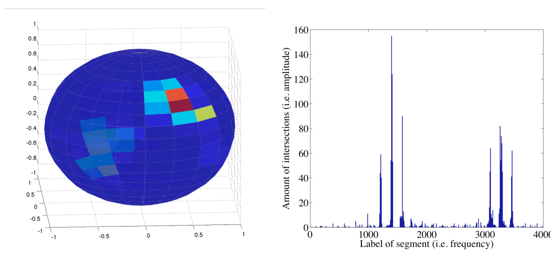

Figure 4. *Visualization of how an increasing resolution of the VSO creates a more detailed analysis of the movement behaviour.*

The same process could now be applied in an online manner taking into account movement behaviour of the full upper body. If we take the movement behaviour of the two wrists and elbows into account, each of the four points activates at each instance one sinusoidal waveform. If the four points move in accordance with each other across the same directional pathways, the simplicity of movement behaviour will be reflected in the simplicity of the sonification.

2.4.2. The sampler model

The sampler model provides means for the user to define (1) the resolution of the VSO, and (2) the sounds or sound processes attached to each segment.
Now, each segment of the VSO could be interpreted as being a pad of a traditional sampling device used for triggering samples (e.g. Akai MPC1000, Roland, SP-404, etc.). A user can activate samples or sound processes attached to specific segments by moving pre-defined points of the upper limbs (e.g. wrists or elbows) in the spatial direction that corresponds to the specific segment of the VSO. But instead of limiting the activation of sounds to touching/hitting pads with the fingers, the embodied sampler allows a more expressive interface between human and computer. It is possible to control sound synthesis and control processes by spontaneous movement of the full upper body. More important, it becomes possible to match the intentions linked to bodily directional behaviour to the intention expressed by sound synthesis processes.

This model is dynamic in the way the application provides the structural framework that can be filled in at wish. It provides a platform for the user to establish sonifications based on the active, explorative engagement of the user. It stimulates exploration of sound and sound qualities. It sharpens the awareness of how the psycho-sensory experience of a sound must be linked to the psycho-sensory awareness of the sound producing gesture in order to allow the exploration and communication of musical expressiveness. The action and the sonification of that action executed by the embodied sampler contribute to the same kind of intentional idea creating the illusion of biomechanical based control and causality. In doing so, the embodied sampler provides an interface for a dynamic interplay between corporeal, spatial, auditory and expressive components.

3. DISCUSSION

This study pointed out that, when dealing with the interpretation and comprehension of movement behaviour, we have to take into account that this can occur on different levels. First, we have the pure physical properties of a movement that can be measured, quantified and quiet easily transformed (i.e. reflected) into physical sound relations. However, this transformation is done on a pure cognitive level and therefore easily liable to randomness and arbitrariness. Moreover, it is forgotten that there is "something behind" the data specifying the movement behaviour. This "something behind" involves the intentionality of a movement. An extensive body of research [4; 9; 2; 8; 1] shows how directionality in a movement, and the relations among the different parts of the upper body are linked to expressivity and intentionality. So what is often forgotten is that the precise and transparent interpretation of a movement is first and for all a matter of the understanding of the intention behind the movement. Nonetheless the subjectivity of that, it is proved [9] that the relationship between intentionality and the formal characteristics of movement and sound could be expressed in a systematic – and therefore, repeatable and general – way. So, when dealing with auditory biofeedback interaction loops, where the user is seen as an active contributor to the generated input data, we can integrate the aspect of intentionality without departing the systematic, scientific methodology. Moreover, it helps the transparency and preciseness of the interpretation of how specific interactions cause the sound to change.

4. CONCLUSIONS

We presented a system based on additive synthesis techniques that could be considered as a sonic alternative for the qualitative, visual data observation. Moreover, we presented an embodied alternative for the classical sampler device that integrates the expressive qualities of the human body in the process of music production. It provided a more intuitive and spontaneous sampling device in comparison with traditional sampling devices where sounds are triggered by finger tapping.

The structural algorithms that make up the model-based sonification system are developed each as standalone Max/MSP *mxj*-objects and can as such be implemented in other HCI-design projects: (1) real-time Savitzky-Golay FIR filter, (2) algorithm to extract the direction of movement from 3D position data, and (3) algorithm to virtually model a user's kinesphere.

5. ACKNOWLEDGMENTS

This work is carried out in context of the EmcoMetecca project that is supported by the Flemish Government. A special thanks goes out to Bart Moens for implementing Java code, Pieter Coussement for the help with 2D and 3D visualizations and Ravan van Noorden for her help with the figures.

6. REFERENCES

[1] Bartenieff, I. 1993. *Body movement: Coping with the environment*. New York: Gordon and Breach Science Publishers.

[2] Broeckx, J. 1981. *Muziek, ratio en affect: Over de wisselwerking van rationeel denken en affectief beleven bij voortbrengst en ontvangst van muziek*. Antwerpen: Metropolis.

[3] Dray, T, and CA Manogue. 2003. Conventions for spherical coordinates. *this Journal* 34:168-169.

[4] Godøy, RI, and M Leman. 2010. *Musical Gestures: Sound, Movement, and Meaning* New York: Routledge.

[5] Groff, E. 1995. Laban Movement Analysis: Charting the Ineffable Domain of Human Movement. *The Journal of Physical Education, Recreation & Dance* 66 (2).

[6] Hermann, T. 2008. Taxonomy and definitions for sonification and auditory display. Proceedings of ICAD 2008, Paris, France.

[7] Kuyken, B, W Verstichel, F Bossuyt, J Vanfleteren, M Demey, and M Leman. 2008. The HOP sensor: Wireless Motion Sensor. Proceedings of NIME 2008, Genova, Italy.

[8] Laban, R, and L. Ullmann. 1966. *Choreutics*. Estover, Plymouth, UK: Macdonald and Evans.

[9] Leman, M. 2007. *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.

[10] Maes, P-J, M Leman, M Lesaffre, M Demey, and D Moelants. 2010. From expressive gesture to sound: The development of an embodied mapping trajectory inside a musical interface. *Journal on Multimodal User Interfaces* 3 (1-2):67-78.