SoundThimble: A High Resolution Gesture Sonification Framework

Ben Trovato
Institute for Clarity in Documentation
1932 Wallamaloo Lane
Wallamaloo, New Zealand
trovato@corporation.com

G.K.M. Tobin
Institute for Clarity in Documentation
P.O. Box 1212
Dublin, Ohio 43017-6221
webmaster@marysville-ohio.com

Grigore Burloiu
CINETic
UNATC "I.L. Caragiale"
Bucharest, Romania
gburloiu@gmail.com

ABSTRACT

The abstract should preferably be between 100 and 200 words.

Author Keywords

sonification, motion tracking, gesture spotting, interactive installation

ACM Classification

H.5.5 [Information Interfaces and Presentation] Sound and Music Computing, H.5.2 [Information Interfaces and Presentation] User Interfaces—Auditory (non-speech) feedback

1. INTRODUCTION

- motivation
  - challenges
  - **the Vicon system**

With advanced technology and intelligent controls, Vicon Vantage represents a powerful tool to capture motion.Vicon Vantage system used in this project contains 8's 5 megapixel camera which transmit movement in realtime to Nexus or Blade. Having accelerometers and temperature sensors, each camera detects any decalibration of the system. [1].

2. STATE OF THE ART

- Vicon & related projects
  - interactive / movement sonification examples[2].
  - *Sound in space* represents another innovative element in this project because introduces the ideea of *controling sound* by movement. This means that the sound is an entity, gets a materialization and it becomes switchable [1]. It is not about the **localization** of sound in space (detecting the direction of sound source), it is about its **position** in space, the coordinates of the sound object in space, like an actual *object* in a room. *Sound objects* concept represents an innovative tool for multimedia arts such as sketches, imaginary games and realtime interactions.

[1]See `https://www.vicon.com`.

NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark.

Interaction between sound control and human gesture has constantly increased over the last years [3]. Probabilistic models for analysing motion and sound relationships became a necessity and a forthcoming tool [4].

3. PROJECT DESCRIPTION

3.1 Concept

- Performance aesthetic

- Gestures, virtual objects, dynamic mapping

Manny interactions and programable elements which are included in the concept have an exactly function and usefulness. These elements are created in MAX with the help of Vicon SDK and OSC C++ library. The concept suppose an interactive action of a performer or a simple user with an virtual object which has associated gestures defined by person. Virtual object interaction acts in sound design utility like a master controller. Searching for a certain object comes with an audio feedback which makes the search easier. Object's asscociated gestures are compatible with sound design patch and contribute at audio performance. A dynamic mapping of marker's coordinates is necessary to transfer data between Nexus and MAX.

- Visual environment

3.2 Implementation

- Character design (Nexus)

Implementation of the concept presented above requires Nexus, Vicon SDK and MAX sotware. Every character involved in the scene is defined by a limited number of markers. In this case, two markers are positioned on the head, one marker is positioned on elbow and the others 2 on the hand (thumb and index finger). Every marker has associated a name in Nexus and between them 6 segments are drawn. It is very important in realtime capture motion, that the marker to have assigned correct coordinates.

- Vicon extensions (SDK plugin)

Vicon's SDK is a versatile and simple tool for users to gain easy access to Vicon DataFlow created in Nexus, Blade or Tracker applications.The Vicon DataStream Software Development Kit (SDK) provides intuitive programable access to data with custom functions created in C++. With the help of some functions, Vicon's SDK forwardes the Vicon DataStream to other constructive softwares and plug-ins to create custom applications [2]. In combination with Open Sound Control protocol, Vicon's SDK forwards data to any

[2]See `https://www.vicon.com/products/software/datastream-sdk/`.

software compatible with this communication protocol (eg. MAX). OSC is a protocol for communication among computers, sound synthesizers and other multimedia devices[3]. Hence, any marker can be routed in MAX using its parameters and coordinates. Also, the Vicon DataStream Software Development Kit (SDK) admits inside changes such as labeling markers, timecode generation and framerate.

3.2.1 Max modules

- Objects generation & performance mechanics

Manipulating objects algorithm consists of some big steps: object generation, finding the object, picking up the object, throwing the object on the floor.

Object generation is realized by random generators with the help of *drunk* object, but with certain limits. These limitations are influenced by the dimensions of the room in which the Vicon system is installed. Finding the object supposes continuous mathematic operations between the coordinates of the object and coordinates of the left hand's marker. This process comes with an audio feedback. When these coordinates are close enough one to another, the object is retreived and manipulated by performer (eg. define gesture). After all these processes, a simple comparison between the coordinates of the floor and the value of the z axes of the marker is done in order to put down the object. According to this, a performer can handle as many objects as he wants.

- Gesture recognition

Mubu containers provided by Ircam laboratories in MAX software represent a handy tool to record and analyze gesture, captured with Vicon system [5]. Our gesture recognition algorithm is based on Hierarchical Hidden Markov Models (HHMM) implemented in *mubu.hhmm* object of MAX/MSP. HHMMs are a generalization of HMM where each state is considered to be a self-contained probabilistic model [6]. The system is trained by captured data which is essentially a gesture. This process requires a predefined indicator in order to delimitate gestures from all data flow. The algorithm analyzes all input data and generates a probability of similarity between data and saved gestures. In order to control every generated object, there are associated 2 or 3 gestures saved by the performer, but there is a limited time for the gestures to be executed. Predefined gestures offer the possibility to delete the gesture just saved and also indicate the moment the gesture is recorded.

- Sound design
- Visualisation (jitter)

4. CASE STUDIES

4.1 Interactive Installation

4.2 Performance

- Solo / duet / tutti ...

5. CONCLUSIONS AND FUTURE WORK

- Areas of improvement
  - Eye tracking?

6. ACKNOWLEDGMENTS

This section is optional; it is a location for you to acknowledge grants, funding, editing assistance and what have you.

[3]See `http://opensoundcontrol.org`.

7. REFERENCES

[1] BrianKane. *Sound Unseen - Acousmatic Sound in Theory and Practice*. Oxford University Press, New York.

[2] T. Hermann, A. Hunt, and J. G. Neuhoff. *The sonification handbook*. Logos Verlag Berlin, 2011.

[3] K. N. Jorge Solis. *Musical Robots and Interactive Multimodal Systems*. Springer-Verlag Berlin Heidelberg, Berlin, 2011.

[4] R. B. F. B. Jules Francoise, Norbert Schnell. Probabilistic models for designing motion and sound relationships. *International Conference on New Interfaces for Musical Expression*, pages 287–292, June 2014.

[5] D. S. G. P. R. B. Norbert Schnell, Axel Robel. Mubu and friends assembling tools. *International Computer Music Association*, pages 423–426, August 2009.

[6] N. T. Shai Fine, Yoram Singer. The hierarchical hidden markov model: Analysis and aplications. *Machine Learning*.