3dinmotion - a Mocap Based Interface for Real Time Visualisation and Sonification of Multi-User Interactions

Alain Renaud
MintLab
Geneva, Switzerland
alain.renaud@mintlab.ch

Caecilia Charbonnier
Artanim
Geneva, Switzerland
caecilia.charbonnier@artanim.ch

Sylvain Chagué
Artanim
Geneva, Switzerland
sylvain.chague@artanim.ch

ABSTRACT

This paper provides an overview of a proposed demonstration of 3DinMotion, a system using real time motion capture of one or several subjects, which can be used in interactive audiovisual pieces and network performances. The skeleton of a subject is analyzed in real time and displayed as an abstract avatar as well as sonified based on mappings and rules to make the interplay experience lively and rewarding. A series of musical pieces have been composed for the interface following cueing strategies. In addition a second display, "the prompter" guides the users through the piece. 3DinMotion has been developed from scratch and natively, leading to a system with a very low latency, making it suitable for real time music interactions. In addition, 3DinMotion is fully compatible with the OpenSoundControl (OSC) protocol, allowing expansion to commonly used musical and sound design applications.

Keywords

Human computer interaction, mapping design, body as a musical instrument, motion capture and music, audiovisual display, sonification.

1. Introduction

This project uses motion capture (mocap) as a means for generating visual data and music. 3DinMotion is a flexible framework combining several types of technologies with the ultimate goal to make interplays as lively and interactive as possible. The mocap part uses state of the art algorithms to accurately grab information emanating from motion analysis. The system has the advantage of being compatible with several motion capture devices ranging from simple Microsoft Kinect [1] to cutting edge professional technologies such as Vicon systems [2]. The visual display part is able to accurately display "avatars" on a screen. The avatars can be modulated and transformed in real time on a 3D pane. The sonification part relies on information sent by motion analysis. All the x,y,z positions are automatically formatted as OSC messages [3]. An application, built with Max/Msp [4] receives the OSC messages and sonifies the result of the motion analysis in a quad soundscape. The audio application also relies on a pre-defined set of cues to guide the users though the interplays [5].

The results of the cues are displayed on a prompter guiding the users through the results emanating from their various body movements. The rationale behind this project is to develop a framework which is robust yet scalable enough to fit a wide variety of applications. The aim is also to offer a system with low latency, so that users do not notice time lags, which can potentially kill the immersive aspect of interaction. The system has been designed to allow the development of interactive audiovisual pieces and since it is network enabled, it can also be used for network music performances.

2. Motion Capture

The system is compatible with a wide range of mocap devices depending on the level of accuracy, capture volume and number of users required. In this framework, the following motion capture systems have been integrated: Vicon system (best accuracy, multi-users), Xsens MVN suit [6] (unlimited capture volume, single user) and Microsoft Kinect (simplest to use, no suit, multi-users). When using the first two motion capture devices, data is transferred to the framework using network streaming and handled by a dedicated C++ class. For the Kinect, the Microsoft SDK [1] was directly integrated into the application. The information retrieved from the various devices is then processed to obtain positions of each joint of the body. Additional C++ code was produced to correctly handle entry/exit in/from the capture volume, multiples users, as well as missing tracking data, which ensures application's robustness and playability. Eventually, the joint's positions (x,y,z) of each user are formatted for each frame (frequency: 30-120 Hz depending on the mocap system used) as OSC messages and sent to the sonification application.

3. Visual display

The visualization is integrated into the application processing the motion capture data. The generation of 3D content is performed thanks to the open-source C++ library Cinder [7] which provides a powerful and intuitive toolbox for programming graphics. All users of the system are displayed as virtual avatars (Figure 1). Their joint positions are represented as small spheres or markers. Each marker defines a particles system and its position is the emitter. The particles have a limited lifespan and change color with motion velocity. Their other properties are customizable (eg: number of particles, size). Another optional feature is the possibility to draw or sculpt in 3D. This is achieved by tracking the trajectories of the hands' positions. These trajectories are displayed as 3D tubes enabling to draw in the 3D space "traces" of ephemeral duration. The 3D environment includes a reflective floor plane lit by a spot light. Moreover, it is possible to easily modify the visualization by adjusting the shaders' parameters. An avatar representation was chosen as the initial visual display. The visual representation will be more customizable in future versions.

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

Figure 1: Real time 3D visualization of 3DinMotion

4. Sonification

The sonification relies on a set of layers, which can be superimposed depending on the desired effect. It is based on gesture capture dimensions introduced by Bevilacqua et al., which include body posture, space and time [10]. The expected sound design needs to be musical whilst keeping coherence and clarity emanating from the movements and display of the users interacting with the system. There are four types of layers, which can be combined in any order or number:

Underlay: A layer, which runs in the background, such as a drone or a rhythm. User input is limited to slowing down or up, or applying an effect such as adding reverberation or changing the center frequency of a low pass filter. This layer is controlled by one message at a time (eg: /right_hand/x).

Melodic: A layer, which can be controlled accurately through, for example, a hand elevation, controlling a melody with distinct pitches or the frequency of a continuous signal. An auxiliary input layer can be applied, for example with the right hand's position in the space to add effects such as pitch bend or vibrato on the note being played. This layer is controlled by one message at a time (eg: /right_hand/x).

Spot effect: This layer relies on the spot effects most commonly found in sound design theory [9]. The idea here is to allow users to trigger sonic events, which provide a break in the musical sequence being performed. The spot effect can also have attached rules, such as triggering a new section in the musical sequence or giving up a specific role to another user interacting with the system. This layer is controlled by one message at a time (eg: /right_hand/x).

Subtractive: This layer relies on a combination of messages to achieve a desired result and is controlled by two or more messages at a time (eg: if /right_hand/x == /left_hand/y then trigger a new spot effect).

The design of each layer is based on a sonification design method developed by Walker and Nees (2011) that considers the following questions [8]:

- What do we want the user to accomplish?
- What part of the data is relevant to achieve a task?
- How much information is needed by the user to achieve the task?
- How does the data need to be manipulated to make sense musically?

5. Cueing

For each piece, a cueing strategy is chosen based on the following types of cues [5]:

Temporal: A type of cue that is sent out as information related to timing. An example of such a cue would be a counter indicating the time left in a cue, or a warning signal.
Behavioral: A type of cue that is sent with a certain scenario attached to it. This can, for example, include the triggering of a waveform, or the suggestion that a given node needs to play certain notes only above the note C4.
Notational: A type of cue that is able to display content that can be identified by the performers as being helpful in the good running of the performance. This can include the display of a cue number, a countdown or dynamic shapes that can be activated by various factors in the performance. The cues are forming a basis for the performance. The users of the system through their actions within layers are modifying the cues and/or triggering new ones.

6. Prompter

A prompter, separate from the display of the "avatars", provide crucial information to the users, including the type and description of the layer and how to "play" it, which cue the piece is in and suggestions for actions to take full advantage of the functionalities offered. The prompter receives OSC messages from the sonification application.

7. Conclusions

The aim of this demonstration proposal is to offer an overview of our 3DinMotion framework and how it can be considered as a multi-user interface for musical expression. The system provides a good basis for developing complex sonic interactions. Its low latency and flexibility for customization will allow for the development of other applications in the future. Constant feedback is being gathered when presented to music festivals and other events. This gives a good basis for further developing the system both in terms of robustness and musicality.

8. References

[1] Jana, Abhijit. "*Kinect for Windows SDK Programming Guide*". Birmingham: Packt Publishing, Limited, 2012.

[2] Vicon motion capture system, http://www.vicon.com/.

[3] Schmeder, A., Freed, A., and Wessel, D. "*Best Practices for Open Sound Control.*" In Linux Audio Conference. Vol. 10, 2010.

[4] Lyon, E. *"Designing Audio Objects for Max/Msp and Pd.",* Middleton: A-R Editions, 2012.

[5] Renaud, A., 2011. "*Cueing and composing for long distance network music collaborations*." In AES 44th Conference on Audio Networking, 18-20 November 2011, University of California, San Diego, USA,. AES.

[6] Rosenberg, D., Luinge, H., and Slycke, P. "*Xsens MVN: Full 6dof human motion tracking using miniature inertial sensors*". Xsens Technologies, 2009.

[7] Rijnieks, K. "*Cinder – Begin Creative Coding*", Packt Publishing, 2013.

[8] Thomas, H., Hunt, A., and Neuhoff, J. "*The Sonification Handbook."* Berlin: Logos Verlag, 2011.

[9] Kaye, D., and LeBrecht, J. "*Sound and Music for the Theatre: The Art & Technique of Design"*. Oxford: Focal Press, 2009.

[10] Bevilacqua, F., Schnell, N., Fdili Alaoui, S., Klein, G., and Noeth, S. "Gesture Capture: Paradigms in Interactive Music/dance Systems." *Emerging Bodies: The Performance of Worldmaking in Dance and Choreography*, 2011, 183.