Gestural control of concatenative sound synthesis

A. Gabriel Vigliensoni

Introduction

Nowadays, it is possible to use one's own sound library as a sonic palette for making music. By segmenting a collection of sounds into small *units*, extracting their acoustic features, and arranging them into a descriptor space, a *unit selection* algorithm can find the closest unit to a *target* sound, and concatenate it to the previous one. This kind of synthesis is called *data-driven concatenative sound synthesis* [1].

As an extension, in *user-driven concatenative sound synthesis* [2] we can freely interact with the *units* by concatenating one audio segment after another without using an algorithm for a target sound. This kind of interaction has been implemented—with musically interesting results—mainly through the navigation of a two-dimensional descriptor space in a computer interface, typically by using a mouse and keyboard, or a graphic tablet. I expect that giving access at the same time to more low- and high-level features will improve the playability of such a system.

The main goal of my research is to provide musicians a system to compose and perform with a *sound corpus* by exploring a three-dimensional space by means of *non-contact gestures*. A "touch-less" interface like this provides a challenge, but also opens channels for expression and performance due to its natural physical degrees of freedom, and the immersiveness of an untethered performance.

Previous Work

A taxonomy for *alternate controllers*—those that are not modeled or inspired by existing acoustic instruments—have been proposed by Mulder in [3], and expanded by Piringer in [4]. In this classification, the non-contact gesture controllers are considered as *expanded-range* or *immersive*, depending if the performer can or cannot "escape" the interface. Using these concepts, Paradiso [5] and Miranda and Wanderley [6] have reviewed and summarized a number of interfaces for music creation, performance, and control, using non-contact gesture sensing.

For measuring a user's free gesture through her body position and motion over time, Burdea and Coiffet come up in [7] with a review of *three-dimensional position trackers*. They address the technology, performance parameters, advantages, and caveats of mechanical, magnetic, ultrasonic, optical, and hybrid inertial systems. These input devices have been used extensively over the last decades for the acquisition of musical gestures, as can be seen in [8, 9, 10], among many others.

Concatenative sound synthesis origins and systems are reviewed and studied by Schwarz in [11] and by Casey in [12]. Both state that this technique emerged closely related to speech synthesis, and it is located at the intersection of many fields of research for music and technology: MIR, DSP, real-time interaction, sound synthesis, and database technology. The former also establishes that we are in front of a "relevant technology advance (...): Large databases of sound, with a pertinent description of their contents, ready for content-based retrieval, (...) and musical sound synthesis".

During the last years, a growing number of pieces and techniques for real-time interaction using these techniques have been developed. *Plumage* by Jacquemin et al. [13] uses the metaphor of a playback head passing through a space full of feathers, controlled by a performer using two mice and a keyboard. *Grainstick* by composer Pierre Jodlowsky and Leslie et al. [14] uses IR tracking and game controllers to give users the illusion of playing a rainstick full of small grains of sound.

Proposed Research / Methodology

My research will be structured in four major parts: (i) a review of the sensing techniques available for spatial data acquisition, (ii) a comparison of professional and consumer-oriented tracking systems, (iii) an overview of concatenative sound synthesis software suites, and (iv) a prototype design considering the previous insights.

Part one will be an extensive description of sensing techniques for three-dimensional position data acquisition. Musical interfaces, instruments, and controllers using these approaches will be surveyed and described. The main objectives of this chapter will be to identify the technologies already used for spatial acquisition, and how they have been implemented in interface design.

In the second part, I will test and compare three professional tracking systems: the Polhemus Liberty 8[1], the Vicon 460[2], and the Xsens[3]; and three consumer-oriented controllers with tracking capabilities: the Microsoft's PS Kinect[4], the Nintendo's Wii remote controller[5], and the MadCatz' Gametrak[6]. Each device will be set up in the same environment, and will control the same variables of a synthesis system. By this means, critical characteristics for music performance such as range, sensing space, occlusion, portability, calibration, and set-up time will be studied.

Part three will explore the following software suites for concatenative sound synthesis: Casey's *Soundspotter* [12], Schwarz's *CataRT* [15], and Brent's *TimbreID* [16]. Their characteristics in terms of segmentation types, sound descriptors, database handling, concatenation types, mapping flexibility, expandability, real-time ability, and dimensionality reduction will be compared.

The fourth stage of my research will involve the development of a prototype to track the position of a performer's hands in three dimensions, the mapping of the acquired gestures to the synthesis engine, and the selection of features for each axis of the descriptor space to create a meaningful interface for musical performance. Thus, the performer will have the impression of being immersed in the sound corpus, and the audio output of each one of the played segments will be spatialized in order to provide properly-located aural feedback in the space.

Contributions

As part of courses taken in the M.A. Music Technology program, I have developed interfaces for gesture acquisition in one and two-dimensional spaces [17, 18]. This thesis will extrapolate these previous works to three dimensions.

Substantial outcomes of my research will be threefold: (i) To provide a comparison of professional and consumer-oriented position trackers. This objective will update previous literature [7], reviews[7], and will consolidate information presented only in the form of catalogs or unstructured internet information. (ii) To update previous work by Piringer [4] on compiling historical interactive musical instruments, especially the free gesture ones. Both updated summaries will be part of the ISIDM database[8]. (iii) To create an interface for exploring, selecting and performing with sonic objects. Such an interface could also be used for more general uses, as in the interactive exploration of a catalog of sound effects in the context of audiovisual post-production, or the navigation of a user's song collection beyond the text-based music browser standard.

[1] Polhemus Liberty 8 Electromagnetic Motion Tracking System, `http://www.polhemus.com/?page=Motion_Liberty`, accessed October 5, 2010

[2] Vicon Motion Capture Systems, `http://www.vicon.com/`, accessed October 5, 2010

[3] Xsens 3D Motion Tracking, `http://www.xsens.com/`, accessed October 5, 2010

[4] Microsoft Kinect for Xbox 360, `http://www.xbox.com/en-CA/kinect/`, accessed November 11, 2010

[5] Nintendo Wii controllers, `http://www.nintendo.com/wii/console/controllers/`, accessed October 5, 2010

[6] Myers, E. E. 2002. A transducer for detecting the position of a mobile unit. UK Patent Application. GB2373039A

[7] CNMAT. Position Sensing Technology and Product Summary. `http://cnmat.berkeley.edu/Position-Sensing`, accessed October 5, 2010

[8] ICMA Workgroup on Interactive Systems and Instrument Design in Music, `http://sensorwiki.org/doku.php/isidm/introduction?s[]=isidm`, accessed November 9, 2010

References

[1] Schwarz, D. 2004. *Data-driven concatenative sound synthesis*. Ph.D. thesis, Universite Paris 6 - Pierre et Marie Curie, Paris, France.

[2] Aucouturier, J. and Pachet, F. 2005. Jamming with plunderphonics: Interactive concatenative synthesis of music. *Journal of New Music Research* 35 (1): 35–50.

[3] Mulder, A. 1998. *Design of virtual three-dimensional instruments for sound control*. Ph.D. thesis, Simon Fraser University, Vancouver, Canada.

[4] Piringer, J. 2001. *Elektronische musik und interaktivität: Prinzipien, konzepte, anwendungen*. Master's thesis, Institut für Gestaltungs- und Wirkungsforschung der Technischen Universität Wien.

[5] Paradiso, J. and Gershenfeld, N. 1997. Musical applications of electric field sensing. *Computer Music Journal* 21 (2): 69–89.

[6] Miranda, E. and Wanderley, M. M. 2006. *New digital musical instruments: control and interaction beyond the keyboard*. Computer Music and Digital Audio Series, A–R Editions Inc., Madison, WI, USA.

[7] Burdea, G. and Coiffet, P. 2003. *Virtual reality technology*. John Wiley and Sons Inc., Hoboken, NJ, USA.

[8] Gehlhaar, R. 1991. Sound = Space: an interactive musical environment. *Contemporary Music Review* 6 (1): 59–72.

[9] Flety, E. 2000. 3D gesture acquisition using ultrasonic sensors. In *Trends in Gestural Control of Music*, Wanderley, M. M. and Battier, M., eds. Paris, France, 193–208.

[10] Camurri, A. and Trocca, R. 2000. Movement and gesture in intelligent interactive music systems. In *Trends in Gestural Control of Music*, Wanderley, M. M. and Battier, M., eds. IRCAM-Centre Pompidou, Paris, France, 95–109.

[11] Schwarz, D. 2006. Concatenative sound synthesis: The early years. *Journal of New Music Research* 35 (1): 3–22.

[12] Casey, M. 2009. Soundspotting: a new kind of process? In *The Oxford Handbook of Computer Music*, Dean, R., ed. Oxford University Press, New York, NY, USA, 421–53.

[13] Jacquemin, C., Ajaj, R., Cahen, R., Olivier, Y., and Schwarz, D. 2007. Plumage: design d'une interface 3D pour le parcours d'échantillons sonores granularisés. In *Proc. of the 19th International Conference of the Association Francophone d'Interaction Homme-Machine*. Paris, France, 71–74.

[14] Leslie, G., Zamborlin, B., Jodlowski, P., and Schnell, N. 2010. Grainstick: A collaborative, interactive sound installation. In *Proc. of the International Computer Music Conference*. New York, NY, USA, 123–126.

[15] Schwarz, D., Beller, G., Verbrugghe, B., and Britton, S. 2006. Real-time corpus-based concatenative synthesis with catart. In *Proc. of the 9th International Conference on Digital Audio Effects*. Montreal, QC, Canada, 279–282.

[16] Brent, W. 2010. A timbre analysis and classification toolkit for pure data. In *Proc. of the International Computer Music Conference*. New York, NY, USA, 224–9.

[17] Vigliensoni, G. and Wanderley, M. 2010. Soundcatcher: explorations in audio-looping and time-freezing using an open-air gestural controller. In *Proc. of the International Computer Music Conference*. New York, NY, USA, 100–3.

[18] Vigliensoni, G. 2010. The enlightened hands: navigating through a bi-dimensional feature space using wide and open-air hand gestures. MUMT620 graduate seminar final report, Schulich School of Music, McGill University.