Gesture–based Control of Physical Modeling Sound Synthesis: a Mapping-by-Demonstration Approach

Jules Françoise
STMS Lab
IRCAM–CNRS–UPMC
1, Place Igor Stravinsky 75004
Paris, France
jules.francoise@ircam.fr

Norbert Schnell
STMS Lab
IRCAM–CNRS–UPMC
1, Place Igor Stravinsky 75004
Paris, France
norbert.schnell@ircam.fr

Frédéric Bevilacqua
STMS Lab
IRCAM–CNRS–UPMC
1, Place Igor Stravinsky 75004
Paris, France
frederic.bevilacqua@ircam.fr

ABSTRACT

We address the issue of mapping between gesture and sound for gesture-based control of physical modeling sound synthesis. We propose an approach called *mapping by demonstration*, allowing users to design the mapping by performing gestures while listening to sound examples. The system is based on a multimodal model able to learn the relationships between gestures and sounds.

Categories and Subject Descriptors

H.5.5 [**Information Interfaces And Presentation**]: Sound and Music Computing; J.5 [**Arts and Humanities**]: Music

Keywords

music, gesture, sound synthesis, physical modeling, HMM, multimodal

1. INTRODUCTION

Gestural interaction with audio and/or visual media has become ubiquitous. Many applications, including music performance, gaming, sonic interaction design, or rehabilitation, involve mapping from physical gestures to sound, requiring solutions for quick prototyping of gesture–based sound control strategies. The relationship between gesture and sound, often called *mapping*, has been recognized one of the crucial aspects of such interactive systems, as its design influence the interaction possibilities. In this paper, we address the issue of mapping for gesture–based control of physical modeling sound synthesis. This is a companion paper of a short-paper presented at ACM Multimedia 2013 [3] that will demonstrate concrete examples of a general a multimodal probabilistic model for gesture–based control of sound synthesis.

Physical modeling sound synthesis aims at simulating the acoustic behavior of physical objects. Gestural control of such physical models remains difficult, since the captured gestural parameters are generally different from the physical input parameters of the sound synthesis algorithm. For example, sensing gestures using accelerometers might pose difficulties for controlling the physical model of a bowed string where force, velocity, and pressure are the expected input. The design of the mapping between gesture and sound is therefore complex, and would hardly be realized by direct relationships between input and output parameters.

To tackle such an issue, we propose an approach we call *mapping by demonstration*, based on machine, allowing users to craft control strategies by demonstrating gestures associated with sound examples. Therefore, the system supports a design of the mapping driven by listening and interaction, as in many cases the training examples are defined by gestures performed while listening to sound examples, as proposed by Caramiaux [1] or Fiebrink [2].

Our approach places the user at the center of an interaction loop integrating training and performance. During *training*, sounds can be designed using a graphical editor. By performing gestures while listening to the sounds, the user feeds the system with examples of the mapping he intents to create, therefore translating his intentions through the direct demonstration of specific control strategies. During *performance*, the learned mapping can be used for sound control, allowing the user to explore the control strategies defined by the training examples.

2. APPLICATION OVERVIEW

2.1 Gesture capture

We use the *Modular Musical Objects* (MO) for gesture capture [5]. These wireless devices include an accelerometer and a gyroscope, and can be integrated to various objects, or extended with additional sensors, for example piezo-electric sensors.

Figure 1: MO: Modular Musical Objects

MM'13, October 21–25, 2013, Barcelona, Spain.

2.2 Modal synthesis

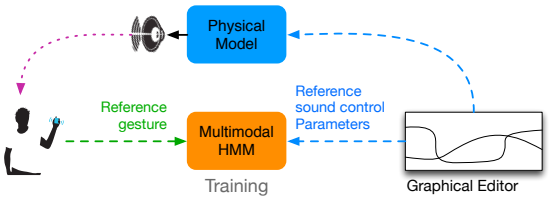
Our system uses *Modalys*, a software dedicated to modal synthesis, i.e. that simulates the acoustic response of vibrating structures under an external excitation. It allows to build virtual instruments by combining *modal elements* – e.g. plates, strings, membranes – with various types of connections and exciters – e.g. bows, hammers, etc. Each model is governed by a set of physical parameters – e.g. speed, position and pressure of a bow. Specific sounds and playing modes can be created by designing time profiles combining these parameters.

2.3 gesture–sound mapping

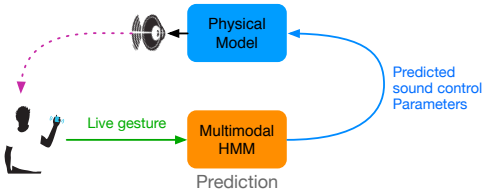
Our goal is to learn the mapping between gestures, captured with accelerometers, and specific time profiles of the control parameters of the physical models. We adopt a multimodal perspective on gesture–sound mapping based on a probabilistic multimodal model. This approach is inspired by recent work in other fields of multimedia, such as speech–driven character animation [4]. The system is based on a single multimodal Hidden Markov Model (HMM) representing both gesture and sound parameter morphologies. The model is trained by one or multiple gesture performances associated to sound templates. It captures the temporal structure of gesture and sound as well as the variations which occur between multiple performances. For performance, the model is used to predict in real-time the sound control parameters associated with a new gesture. Additional details about the model and its implementation can be found in [3].

2.4 Workflow

The workflow of the application is an interaction loop integrating a *training* phase and a *performance* phase. It is illustrated in figure 2, and a screenshot of the software is depicted in figure 3. In the *training* phase, the user can



(a) *Training*: sounds are designed using a graphical editor, and reference gestures can be recorded while listening.



(b) *Performance*: the model is used to predict the sound parameters associated with a live gesture.

Figure 2: Application workflow.

draw time profiles of control parameters of the physical models to design particular sounds. Each of these *segments* can be visualized, modified, and played using a graphical editor

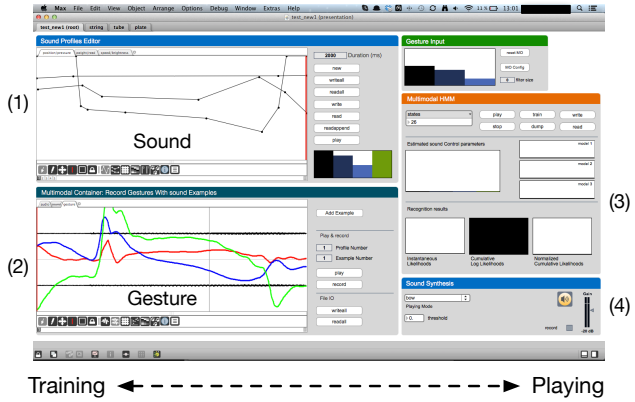


Figure 3: Screenshot of the system. **(1)** Graphical Editor. **(2)** Multimodal data container. **(3)** Multimodal HMM: control panel and results visualization. **(4)** Sound synthesis.

(top left of figure 3). Then, the user can perform one or several demonstrations of the gesture he intents to associate with the sound example (figure 2a). Gesture and sound are recorded to a multimodal data container for storage, visualization and editing (bottom left of figure 3). Optionally, segments can be manually altered using the user interface. The multimodal HMM representing gesture–sound sequences can then be trained using several examples. During the *performance* phase, the user can gesturally control the sound synthesis. The system allows for the exploration of all the parameter variations that are defined by the training examples. Sound parameters are predicted in real-time to provide the user with instantaneous audio feedback (figure 2b). If needed, the user can switch back to *training* and adjust the training set or model parameters.

3. ACKNOWLEDGMENTS

We acknowledge support from the French National Research Agency (ANR project Legos 11 BS02 012).

4. REFERENCES

[1] B. Caramiaux. *Etudes sur la relation geste-son en performance musicale*. Phd dissertation, Université Pierre et Marie Curie, 2012.

[2] R. Fiebrink, P. R. Cook, and D. Trueman. Play-along mapping of musical controllers. In *In Proceedings of the International Computer Music Conference*, 2009.

[3] J. Françoise, N. Schnell, and F. Bevilacqua. A Multimodal Probabilistic Model for Gesture-based Control of Sound Synthesis. In *Proceedings of the 21st ACM international conference on Multimedia (MM'13)*, Barcelona, Spain, 2013.

[4] G. Hofer. *Speech-driven animation using multi-modal hidden Markov models*. Phd dissertation, University of Edimburgh, 2009.

[5] N. Rasamimanana, F. Bevilacqua, N. Schnell, E. Fléty, and B. Zamborlin. Modular Musical Objects Towards Embodied Control Of Digital Music Real Time Musical Interactions. *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 9–12, 2011.