Blending the Physical and the Virtual in Music Technology: From Interface Design to Multi-modal Signal Processing

George Tzanetakis
University of Victoria
Victoria, BC, Canada
gtzan@cs.uvic.ca

Sidney Fels
University of British Columbia
Vancouver, BC, Canada
ssfels@ece.ubc.ca

Michael Lyons
Ritsumeikan University
Kyoto, Japan
michael.lyons@gmail.com

ABSTRACT

Recent years have seen a significant increase of interest in rich multi-modal user interfaces going beyond conventional mouse/keyboard/screen interaction. The new interface technologies are broadly impacting music technology and culture. New musical interfaces use a variety of sensing (and actuating) modalities to receive and present information to users, and often require techniques from signal processing and machine learning in order to extract and fuse high level information from noisy, high dimensional signals over time. Hence they pose many interesting signal processing challenges while offering fascinating possibilities for new research. At the same time the richness of possibilities for new forms of musical interaction requires a new approach to the design of musical technologies and has implications for performance aesthetics and music pedagogy. This tutorial begins with a general and gentle introduction to the theory and practice of the design of new technologies for musical creation and performance. It continues with an overview of signal processing and machine learning methods which are needed for more advanced work in new musical interface design.

Categories and Subject Descriptors

H.5.5 [**Sound and Music Computing**]: Methodologies and techniques, Modeling, Signal analysis, synthesis, and processing systems

Keywords

New interfaces for musical expression, musical instrument design, human computer interaction, signal processing, machine learning, multimodal interaction, digital musical instruments

MM'13, October 21–25, 2013, Barcelona, Spain.
ACM 978-1-4503-2404-5/13/10.
http://dx.doi.org/10.1145/2502081.2502238.

1. INTRODUCTION

Traditional musical instruments offer some of the most fascinating technological artifacts ever created by human beings. The complexity and richness of control afforded by acoustic musical instruments, such as violin or piano, to professional musicians is impressive. Research in new interfaces for musical expression [2, 3, 1] has explored how such complex and delicate control can be combined with the ability to interact in real-time performance situations with computers. In some cases, research on musical interface technology has anticipated aspects of the development of rich multimodal interfaces in the field of augmented reality and has provided excellent examples of sensorially rich and temporally-detailed human-machine interaction. With this in mind, musical performance interfaces present intriguing working examples for gaining new insight into the field of multi-modal interaction.

From the viewpoint of music technology, advances in digital technologies have led to a situation where computers play a role in most music production and performance. Digital technologies offer unprecedented opportunities for the creation and manipulation of sound, however the flexibility of these new technologies implies a confusing array of choices for musical composers and performers. Some artists have faced this challenge by using computers directly to create music and leading to an explosion of new musical forms. However, most would agree that the computer is not a musical instrument, in the same sense as traditional instruments, and it is natural to ask "how to play the computer" using interface technology appropriate for human brains and bodies.

We are at the beginning of a new era in human computer interaction in which users will come to view the way we interacted with computers using keyboards and mice, as we view today punched cards and teletypes. The large diversity of affordable hardware in terms of sensors and actuators today, provides a fertile environment for creating rich multimodal interfaces. Frequently, their design and development requires interdisciplinary expertise combining concepts from digital signal processing (DSP) and human computer interaction (HCI).

There is an increasing interest in rich user interfaces that go beyond the traditional mouse/keyboard/screen interaction. In the past few years, their development has accelerated due to the wide availability of commodity sensors and actuators. For example, the Microsoft Kinect provides, at low cost, a structured light infrared depth camera, a regular color camera, and a microphone array. Moreover, smart phones contain a variety of additional sensors such as accelerometers that provide unique opportunities for control. Unlike traditional controllers such as a mouse that provide direct and simple sensor readings, these new rich interfaces provide high dimensional, noisy and complex sensor readings. Therefore sophisticated digital signal processing and machine learning techniques are required in order to develop effective human computer interactions using such interfaces [4]. At the same time they offer fascinating possibilities of blending physical and virtual worlds such as non-invasive full body control and augmented reality. Different algorithms and customizations are required for each specific application and possibly even for each user so there is a large design space for novel research and contributions.

The goal of this tutorial is to provide an overview of the field of new musical interface design and to give participants a brief but broad introduction to the major techniques needed to conduct research in the field: ranging from design and aesthetics of music technology to the digital signal processing and machine learning techniques required for advanced work in the field. Case studies will be taken from some of the recent work on new musical interfaces.

2. COURSE OUTLINE

2.1 Introduction and Tutorial Overview

The introductory module will give an outline and schedule of the tutorial and motivate the major questions to be addressed using a selection of case studies. Why does music technology represent an excellent application domain for the study of real-time multi-modal interaction? How can digital technologies be designed effectively for the requirements of real-time expressive performance? How have advances in sensor and actuator technologies changed the field of music technology research? What do advances in digital signal processing and machine learning techniques bring to the field?

2.2 A Practical Guide to Creating Multi-modal Musical Interfaces

This module offers a clear six step guide to prototyping a new musical interface. An overview of sensors and actuators as well as sound synthesis techniques is given. The issue of mapping between sensors, sound synthesis, and actuators is introduced. Finally the role of demonstration and performance on the research and development cycle will be described.

2.3 Design, Theory, and Aesthetics of Musical Interaction

This module makes contact with approaches drawn from the field of human-computer interaction that serve to guide the design of multi-modal real-time musical interfaces. A model of a multi-modal musical interface is described and some theoretical issues, such as feedback, mapping, transparency, control intimacy, and flow are introduced. Some aesthetic considerations relating to the design of real-time multi-modal interfaces are described. Finally some useful design heuristics (aka rules of thumb) applicable to the domain are introduced.

2.4 Case Studies I

The material from the first part of the course will be reviewed in the context of case studies from the published literature. These will be illustrated with video demonstrations.

2.5 Signal Conditioning & Feature Extraction

This module covers the basics of signal conditioning with topics such as denoising, dimensionality reduction, filtering, resampling, and dealing with missing samples with concrete examples using multi-modal sensor data. Topics to be introduced include: various types of time-frequency analysis (short time fourier transform (STFT), wavelets), basic image and video analysis, and modeling dynamics. The algorithms will be illustrated with specific examples using audio, video, and sensor data in the context of new interfaces for musical expression.

2.6 Dealing with Uncertainty and Time

In this module, techniques from machine learning and data mining such as supervised and unsupervised learning will be described. The algorithmic constraints of real time and causal implementations will also be covered. Techniques for time modeling such as Kalman Filters, Hidden Markov Models (HMM), and Dynamic Time Wrapping (DTW) will be described through specific examples from multi-modal interaction. Sensor fusion will also be discussed.

2.7 Case Studies II

Case studies to be presented include: a surrogate sensor methodology for training non-invasive signal acquisition using direct sensors through classification and regression, teaching a virtual violinist to bow using machine learning to automatically determine sound quality, self-tuning and self calibrating music robotic instruments, extending the performance possibilities of a vibraphone using the Microsoft Kinect controller, and the Soundplane, multi-touch, pressure sensitive new music interface that provides highly expressive control possibilities.

2.8 Conclusion

The conclusion will summarize the major points covered in the tutorial, provide suggestions for further study, as well as specific links to the published literature, conferences, and other resources in the related research community.

3. REFERENCES

[1] F. Bevilacqua, S. Fels, A. R. Jensenius, M. J. Lyons, N. Schnell, and A. Tanaka. Sig nime: music, technology, and human-computer interaction. In *Extended Abstracts CHI'13*, pages 2529–2532. ACM, 2013.

[2] S. Fels and M. Lyons. Creating new interfaces for musical expression: introduction to NIME. In *SIGGRAPH 2009 Courses*. ACM Press, 2009.

[3] I. Poupyrev, M. Lyons, S. Fels, and T. Blaine. New interfaces for musical expression. In *Extended Abstracts CHI'01*, pages 491–492. ACM Press, 2001.

[4] A. Tindale, A. Kapur, and G. Tzanetakis. Training surrogate sensors in musical gesture acquisition systems. *Multimedia, IEEE Transactions on*, 13(1):50–59, 2011.