http://cnmat.berkeley.edu/publications/enactive-approach-computer-music-performance

An Enactive Approach to Computer Music Performance

David Wessel
Center for New Music and Audio Technologies (CNMAT)
Department of Music
University of California Berkeley
wessel@cnmat.berkeley.edu

Résumé

Experimental computer music performance practices designed from an *enactive* view of musical perception, cognition, and motor control are described. The *enactive* view emphasizes the role of *sensory-motor* engagement in musical experience. The enabling elements required for the approach include, rich and precise gestural interfaces, connectivity devices, real-time gesture analysis and mapping software, richly controlled sound synthesis and processing, and the composition of musical worlds in the form of generative algorithms worthy of extensive exploration. These practices in human-instrument symbiosis require a commitment on the part of musicians to develop both refined motor skills and engagement in the development and refinement of real-time software.

1 Introduction

The *enactive* view of perception and cognition (Varela *et al*, 1991) as well as the vast majority of work on the link between perception and action (Noe, 2005) is dominated by concerns with visual experience. The approach has yet to be widely applied to auditory experience. Traditional psychoacoustics has ignored the role of *sensoy-motor* engagement in hearing. Research in auditory and musical cognition has a lot of catching up to do to reach the level of understanding we have accumulated concerning the perception-action link in vision.

One could imagine an auditory version of the classic perception action link experiment by Held and Hein (1963). Their apparatus is shown in Fig 1 and their basic paradigm has one of the animals transacting with the environment while its yoked control is passively driven about. Perceptual tests showed that the kittens who rode in the gondola had considerable perceptual impairment compared to those who where actively transacting with the environment. Though full blown music perception experiments using Held and Hein's paradigm have yet to be carried out, pilot experiments in our lab suggest that passive listeners do not develop perceptual skills to the same extent as those actively manipulating musical material. Though laboratory experiments remain to be done, this essential link between perceptual development and actively performing music seems all too obvious to teachers of *solfege* and instrumental practice.

Indeed, both traditional instrumental and vocal musical practices require rich *sensory-motor* engagement. On the other hand, electro-acoustic music has for the most part been a studio art and modern computer-based musical instrumentation remains far from involving the body. As computers begin to populate the musical stage they are most often found before performers who manipulate the keyboard and mouse in manner all too reminiscent of office work. This is certainly not a situation that invites the development of musical virtuosity nor enthusiasm on the part of the audience.

A general framework for reasoning about computer-based musical instruments is shown in Fig. 2. The left part of the diagram shows key cognitive mechanisms in the human performer and the region from the sensors to the loudspeaker de-

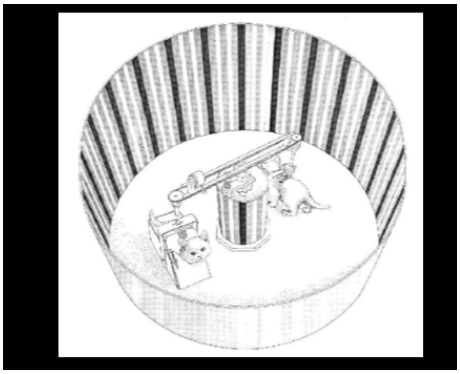

Fig. 1 – The experimental apparatus for the classic experiment by Held and Hein (1963). The kitten on the left in its gondola is driven about in the environment by the kitten on the right

fines the instrument. The motor program is developed and refined by bodily transactions with the instrument and perceptual evaluation of the resulting sound. The exploratory control mechanism combines intentional messages to the motor program with not so intentional messages - *babbling* - and specifies the bodily transaction with the instrument. The term *babbling* was chosen because of the critical role it appears to play in the development of speech in infants. Babbling involves sequences of commands sent to the motor system in order to learn the relationship between these commands and the sonic result produced by the vocal apparatus. Babbling is distinguished from intentional commands in that auditory feedback provides information about the relationship between a gesture and a resulting sound whereas auditory feedback to intentional commands provides information about the match between the desired output sound and the actual output sound. In this albeit theoretical framework, the exploratory control mechanism provides a variational method for refining, generalizing, and otherwise evolving the motor program. We will return in subsequent sections to the import role of *sensory-motor* exploration of a musical world specified in part by such computer-based instruments.

Sensors provide physical measurements of gestures and processing of these measurements is required before useful parameters can be extracted. Gesture processing involves both classification, i.e., gesture recognition, and the extraction of continuous control parameters.

The controller is where the mapping between parameters extracted from the gestures are mapped to the parameters that control the generation of the musical result. In most applications the dimensionality of the gesture data is smaller than the dimensionality of the generative algorithm's control parameters. The controller then handles the mapping between low and high dimensional data. To get a feel for this problem of low to high dimensional mapping consider a generative algorithm that does additive synthesis with a large number of oscillators, say a 1000. If both the amplitudes and frequencies are to be specified we have 2000 parameters. No matter how rich the gesture sensing, it it hard to imagine independent gestural control of that many parameters.

2 Rich and precise gestural interfaces

Ease of use is an important design criterion for user interfaces. Clearly, the instrument should be easy to play. Unfortunately, the ease of use goal, when achieved, often makes the resulting instrument seem like a toy, interesting for a while, but not something with which one would want to invest a lifetime's worth of musical practice. So additional design criteria must be included. The instrument should inspire the development of virtuosity. It should have a vast potential for musical expressiveness. It is here where aesthetic criteria play a large role. The instrument must be composed.

Musical control intimacy and virtuosity require both spatial and temporal precision in the sensing of gestures. Control intimacy refers to a tight connection between a body movement and change in an auditory feature. Wessel and Wright (2002) argue that while low latency responsiveness is essential for control intimacy, low variation in latency is even more critical. Consider the manner in which temporally tight grace note gestures control timbre as in the percussionist's flams. In these situations the performers control precision is in the millisecond range as is our perceptual sensitivity to tem-

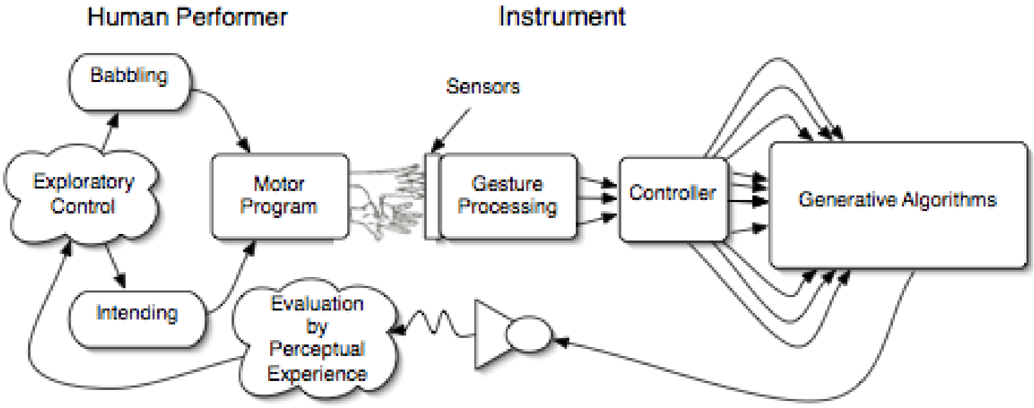

FIG. 2 – A framework for reasoning about computer-based musical instrumentation

poral fine structure. The jitter in MIDI and other low rate gesture sensing devices is inadequate for this level of control.

Gestures select and control. They are symbolic as well as analog in character. They have names as well as shapes. Gestural interfaces and the analysis software associated with them must both classify and derive control parameters. Triggering events has been popular but triggering and subsequently shaping events is less well developed in computer music practice and needs more attention. The much maligned modal interface provides a simple mechanism for selection and control. Better are interfaces that locate different controllers in a spatial array. In my own work I have found it best to combine a modal approach with a multi-controller array. I change from situation to situation via present style selection and the multiple gestures in my interface change their meaning.

Control of polyphony with multiple continuous controllers is important in my own musical work. The piano keyboard is polyphonic but little in the way of continuous control is afforded. The *Thunder* controller from Don Buchla, Fig 3, provides for continuous control of several different regions simultaneously. A new version of a *Thunder*-like controller is in development at CNMAT and makes use of a tiled array of 24 independent rectangular track pad style controllers. Each of them is sensitive to x, y, and pressure. They are separated by small ridges providing tactile refererences. The array is sampled at a much higher rate than *Thunder*, upwards to 8kHz. The fingers can control a number of the sensors simultaneously. In fact without stretching the hands all ten digits can be employed.

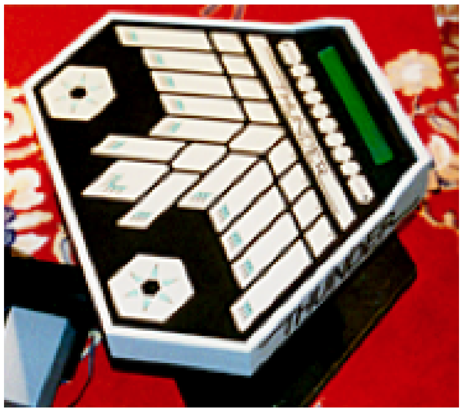

FIG. 3 – The Thunder controller by Don Buchla. Each strip is sensitive to both finger location and pressure in a continuous manner. It provides for polyphonic control of continuous variables

3 Connectivity devices

At CNMAT we have designed and constructed a series of connectivity devices that provide for several channels of audio input and output, a large number of analog signals from sensors, as well as MIDI input and output (Avizienis, *et al* 2000). CNMAT's connectivity processors multiplex the audio, sensor data, and MIDI into a single stream of Ethernet packets assuring a tight temporal coupling between gestures and audio that is, in effect, sample accurate. Furthermore, gestures from the sensors are treated as continuous functions of time and sampled at rates appropriate to their application. Interestingly, Andy Schloss (personal communication) has used a CNMAT connectivity processor in conjunction with the Radio Baton developed by Max Mathews and Tom Oberheim (Mathews 1991). To generate the 30kHz radio signals for the batons Schloss used a 96kHz sampling rate and as a result the gesture sensing and analysis was carried out entirely in Max/MSP at the 96k rate. It is not that such a high rate was required for the gestures themselves but that the signal processing on the raw Radio Baton output was carried out directly in the computer without the circuitry Mathews and Oberheim had built for the coding of sensor data into MIDI, yielding a greater degree of control intimacy with the bonus of additional signal processing flexibility for gesture analysis provided by Max/MSP.

Fig. 4 – The gesture input and digital audio input panel of CNMAT's connectivity device. Two Ethernet ports, MIDI I/O, and multipin inputs for sensors are provided

The multiplexing of gesture and audio data into the same stream goes a long way to providing control intimacy. The operating system is not required to make context switches to receive data from different input sources that would produce temporal jitter. Consequently a tight link between the audio and gesture is assured.

We have used our devise in a variety of contexts including the augmentation of traditional instruments such as the cello and guitar as well as with specialized multitouch and related input structures. The connectivity processor uses both the Apple's Core Audio and Core MIDI with the current version of the Macintosh operating system, OSX 10.4.

4 Gesture analysis and mapping software

As we noted earlier that gesture analysis requires both classification and the extraction of continuous control parameters. There is a growing interest in multi-touch controllers. Let's consider for a moment a multi-touch sensor system with high *taxel* density. *Taxels* are the tactile analog of pixels in video systems. Futhermore, let's sample the array with a rate of 8kHz or so and apply computer vision techniques to the recognition of gestures and the extraction of parameters for continuous control from the pressure dynamics of the hand when it is in contact with the surface. Dynamic "hand force images" constitute the gesture volcabulary of the system. At CNMAT we have taken some small steps in this direction using the *taxel* array developed by the Tactex firm in Canada (Wessel & Wright 2002). The results are promising and provide for a remarkably rich mapping to control parameters.

Unlike many of the lap-top performers I have chosen to avoid automatic generation of sound without my intervention and control. I subscribe to a no action no sound principle. Furthermore, dynamics are controlled by the pressure applied to the controller surfaces. Specifying dynamic control and the silencing of my instrument is the first mapping issue I resolve.

5 Richly controlled sound synthesis and processing

Where does timbral richness come from ? While concentrating on the spectral content, much of the perceptual research on timbre has all but ignored the temporally evolving features of sound and that includes much my own work on the subject. Timbral research should pay more attention to modulations in both amplitude and frequency. Such sub-audio modulations are useful for the characterization of amplitude envelopes, grain, roughness, and frequency trajectories (Atlas & Shamma 2003). Even spectrally impoverished signals sound rich if their amplitude envelopes and frequency trajectories are complex and varied.

Our gestures, if treated as continuous signals, can provide direct control of these perceptually potent amplitude and frequency envelopes. Try this simple experiment. Place a Force Sensing Resistor (FSR) on your finger tip. Sample the signal from the FSR at the audio rate and use it to scale the amplitude of a sine wave. Tap with different gestures and on surfaces with different textures. The timbral variation possible with this primitive but direct connection to the sound is remarkable.

While the simple instrument just proposed leaves much to be desired it suggests that extracting dynamic shapes from gestures and mapping them properly will have powerful effects in the shaping of musical expression. It also provides the musician with a tight feedback loop for exploratory control.

6 The dual role of feedback

As shown in Fig 2, auditory feedback is proposed to play an important role in both musical exploration an control. It plays an essential role in assessing the match between intentions and the resulting sound. What then is the role of feedback with respect to *babbling.* It turns out that many control problems can be difficult to learn with simple adaptive feedback control. There is often a many-to-one relationship between a set of control parameters and the resulting sound. Furthermore the space of parameter settings that achieve the appropriate result can be non-convex. Non-convexity means that averages of the parameter settings that achieve an appropriate result may themselves not provide for an appropriate result. In this case the development of a *forward model* (Jordan & Rumelhart 1992) can aid in learning a controller. A forward model characterizes the relationship between the commands sent by the controller and the resulting sound output. *Babbling* is a non-goal-directed variation of the control parameters and is a key to the exploration of an instrument's potential for musical expression.

References

Atlas, L. & S. Shamma (2003) Joint acoustic and modulation frequency. *Eurasip J. on App. Signal Processing* No. 2003.

Avizienis, R., A. Freed, T. Suzuki, and D. Wessel (2000), Scalable Connectivity Processor for Computer Music Performance Systems. *Proceedings of the International Computer Music Conference 2000*, Berlin, Germany, International Computer Music Association, San Francisco.

Held, R. & A. Hein. (1963) Movement-Produced Stimulation in the Development of Visually Guided Behavior. *J Comp Physiol Psychol.* 1963 Oct ; 56 :872-6.

Jordan, M. & Rumelhart (1992) Forward Models : Supervised Learning with a Distal Teacher. *Cognitive Science,* Vol. 16, No. 3, 307-354

Mathews, M. (1991) The Radio Baton and Conductor Program. *Computer Music Journal,* Vol 15, No. 4, 1991.

Noe, A. (2004) *Perception in Action.* MIT Press, Cambridge Mass.

Wessel, D. & M. Wright. (2002) Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal,* 26 :3, pp. 11-22, Fall 2002

Wessel, D. L., (1979) Timbre space as a musical control structure. Computer Music Journal, Vol. 3, No. 2, 1979, 167-171

Varela, F. J., E. Thompson , E. Rosch (1991) *The Embodied Mind : Cognitive Science and Human Experience,* MIT Press, Cambridge Mass.