

TRAKHUE - INTUITIVE GESTURAL CONTROL OF LIVE ELECTRONICS

Nils Peters[‡], Megan Evans[‡], Eliot Britton[‡]

Schulich School of Music, McGill University, Montreal

[‡]Music Technology Area - Real Time Multimodal Lab

[‡]Department of Performance - French Horn

[‡]Music Composition Area - Digital Composition Studios

CIRMMT - Centre for Interdisciplinary Research in Music Media and Technology

ABSTRACT

This paper discusses a new control method for live electronics which bypasses traditional haptic models through a virtual "landscape" of control parameters activated by a video capture system. The performer navigates the virtual landscape by his/her physical motion, moving through the various sets of parameters. The paper also discusses the interdisciplinary approach used to realize this project, involving software development, performance application, and the realization of a compositional idea.

1. INTRODUCTION

Every method of interaction between performer and electronic music system has its strengths and weaknesses. In our research we sought a system that left the performer unencumbered by physical mechanisms but with good control of the performance environment. Such a system must allow the performer latitude in interpreting a musical composition. Also natural physical gestures must be possible so the performer can quickly master the relationship between a gesture and its musical effect.

As described in [5] we sought to use concepts of immersive virtual landscapes to create a sound environment where the organization of sound parameters is simple and intuitive. Taking this concept, a virtual representation of the performer is placed inside this sound environment and through a motion tracking method, he/she navigates this landscape.

TrakHue was developed to control audio signal processes in Max/MSP by tracking a performer's head motions. Unlike systems that require performers to use an independent cueing method, TrakHue is a way to incorporate physical gestures which are more natural for the performer. The TrakHue concept is based on the Max/MSP software environment and a small video camera, which is the only visible hardware component on stage (see Figure 1). To explore the musical possibilities of this application, the pieces *Hilbert Space no.1 and no.2* were composed for a french horn player who controls the blend and dynamic of multiple prerecorded audio samples within TrakHue.

2. IMPLEMENTATION

2.1. Jamoma

The software was implemented using Jamoma [6], a framework for developing high-level modules in the Max/MSP/Jitter environment, encouraging a structured approach, interchangeability, flexibility and controllability. It offers a standardized way of dealing with presets and simplifies complex handling of parameters: The entire communication of messages to and from a module employs the Open Sound Control protocol [9], whereby every module is named and reachable through the first OSC-nametag. The message

```
/player.1/audio/gain/midi 80
```

would set the gain of a module named */player.1* to the midi-value 80. Jamoma includes a system for remote communication which also supports wildcards. This enables the assigning of messages to multiple receivers. The following example mutes all audio modules by sending:

```
/*audio/mute 1
```

Due to the high-level modularity concept and a variety of ready-to-use Jamoma modules, complex audio/video patches can be quickly created ¹ and adapted. The Jamoma-module *jmod.cuelist.mxt* enables the user to store, load, set and edit snapshots of the parameter state within the Jamoma-environment. This allows the user to organize, structure and develop new musical ideas.

2.2. Processing

2.2.1. Motion tracking

Motion tracking is the most critical component of the TrakHue implementation because its reliability and accuracy determine the usability and versatility of the entire system. The core of this tracking system is the *tap.jit.colortrack* object, which is a part of Tap.Tools [2]. This object includes an algorithm that can track up to four different colors in a video stream simultaneously. A conventional webcam, is used to capture the performer's head

¹ For this project three new Jamoma modules were developed which are available as a part of the Jamoma UserLib.

movement, at a speed of 30 frames per second, through the use of a hairband (see Figure 1). The performer must wear a hairband of a solid color, not duplicated within the camera's field of view, in order that the system may easily identify and thereby track the head motion. A preview window inside the patch shows the incoming video stream of the performer with the hairband. By clicking on the hairband, the target color for the tracking algorithm is chosen. The colortracker outputs the center coordinates (X/Y) of the recognized area. In order to increase the robustness of the colortracker process, tolerance levels can be adjusted in terms of hue, saturation, and luminance. Furthermore to avoid unwanted noise and jitter in the outputted center coordinates a lowpass and/or a median filter can be applied as necessary. Before the video signal is processed the image can be cropped and adjusted for coloration so as to give the optimal view of the performer.

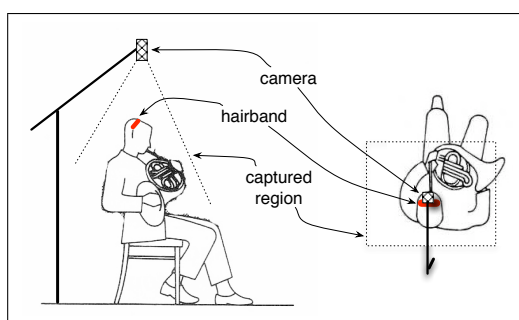


Figure 1. Camera setup for TrakHue

2.2.2. Colormap

The colormap object employs a geometric approach to the arrangement of sound processing parameters in a virtual environment, as described in [5]. The colormap is a bird's-eye projection of a set of five gaussian kernels, freely organizable in the virtual space. Each kernel is adjustable in position, size, and weight, and represents a stored cue of parameter settings controlling musical processes for example audio effects, synthesizer, or sampleplayer. With the head gestures the performer essentially controls their location within the map. Depending on their current location within the map, the colormap calculates an appropriate new parameter state through interpolation and weighting between the overlapping kernels and their associated cues.

2.2.3. Hiking on the colormap

Two options of mapping a gesture into a location of the virtual map are implemented. For the first, the physical motion of the head is directly mapped onto a location in the virtual map. The other option is a "joystick-like" control where the location within the virtual map moves in the same direction of the head as long as the performer's head is off-center of the camera image. The speed of this

motion is a function of the distance of the head to the central position, whereby a position further off-center yields a faster motion. Both mapping methods have different strengths which can be applied to different kinds of musical ideas.

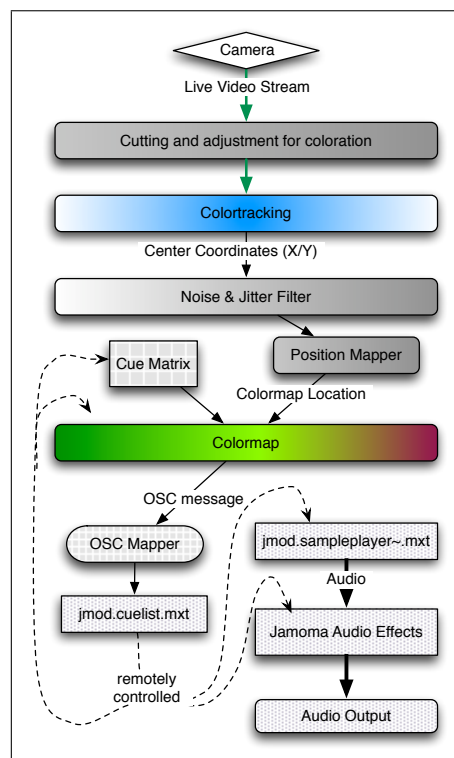


Figure 2. processing overview

3. A PERFORMER'S PERSPECTIVE

3.1. Control over timing, dynamics and mixing

Performance testing of TrakHue proves that this system allows the performer a surprising amount of musical control over the live electronic part. When navigating the virtual map the performer has precise control of the velocity at which he or she enters and exits different areas of the virtual landscape. As the colormap is a bird's eye view of Gaussian kernels, the visual representation of the kernel is a color blob on the map which fades from very saturated to invisibility. The ability to control the speed that the kernels are entered and exited allows the performer to create an electronic part which crescendos and decrescendos musically. Dynamic control, interpretation and manipulation of the electronic part allows the performer freedom to interpret and vary the prerecorded samples or processing that they are triggering. With TrakHue the performer can move seamlessly between the different areas of the virtual map, and blend the different musical elements of live electronics. This virtual world is very convincing and easily comprehensible by the performer, resulting in high performance satisfaction.

3.2. Musical Gesture vs. Physical Gesture

Ancillary physical gestures (movements which are not essential to the operation of the instrument) are a naturally occurring response to expressive musical gestures. Gesture, whether physical or musical, can be defined as "the carrier set of temporal/spatial features responsible for conveying expressiveness." [1]. A variety of studies of the ancillary gestures of musicians have shown that these movements seem to be an integral part of musical performance. Studies have also shown that physical gestures tend to be consistent throughout a musician's performance, and also across many performances [8]. Gesture is not simply a byproduct of performance but also helps convey expressive intent to an audience. A silent video of a performance contains as much expressive information as an audio recording of the same performance [7]. With TrakHue, these physical gestures can be used as a way for the performer to communicate with the live electronics as well as the audience.

3.3. Intuitive Control

TrakHue gives the user the feeling that the physical space around them has actually been transformed to contain sound. These sounds can be accessed by moving the torso within a comfortable range and using naturally expressive gestures. As McNutt in [4] points out, any type of electronic prosthesis is distracting and cumbersome for the performer. McNutt also says that "the more unfamiliar the technology is to the musician, the more distracting, even for a brilliant and skilled performer." In motion tracking, aside from a colored hat or headband there is no additional equipment attached to the performer and the required gestures are based on the natural gestures of a performer and not difficult to master.

4. COMPOSITIONAL ISSUES

Hilbert Space no.1 places the performer in a sound environment where the performer can move towards four sound sources that are placed at the cardinal points. The map is composed in such a way that the performer has precise control of the levels and blending of four sampleplayers. A variety of tension and release events can be achieved by pairing different instrumental sounds with contrasting or similar sounds. The four electronically generated sounds in *Hilbert Space no.1* have very different textures and attack densities. They are all tied together with a harmonic structure that revolves around the pitch class E. The french horn line acts as a unifying force, bridging the textural gap between the sound sources. *Hilbert Space no.2* has a colormap that allows the performer linear movement through a sound space. In this case, joystick gesture mapping is appropriate because it allows maximum control of the colormap with minimum physical gesture. Instead of blending contrasting sounds slowly as in *Hilbert Space no.1* this colormap allows only abrupt transitions between sounds. In *Hilbert Space no.2*

the sound files are rhythmically related. The sound files are the same length so that when they loop, their patterns remain synchronized. As the performer moves through the color map, sounds are added and taken away. As a compositional method, this allows the composer to create sound files in layers, becoming more or less complex. Also, each line on the color grid can act as a trigger for events in the live processing. *Hilbert Space no.2* has a completely different set of compositional restrictions and aesthetic consequences than *Hilbert Space no.1*

4.1. Composing with the colormap interface

Care is taken to balance functionality and detailed control. This balance is important when creating a work environment conducive to creativity. As part of the precompositional process the motions of the performer can be simulated by moving the cursor within the color map. Composers are able to test ideas and gestures before the performer is required. This simulation feature enables the composer to save rehearsal time to work out more important musical ideas and performance details.

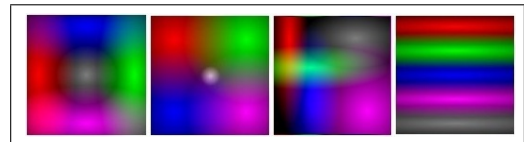


Figure 3. examples for different colormap-arrangements

4.2. Notation

Controlling electronics using new interfaces poses notational challenges. It is difficult to standardize notational systems because new interfaces often require innovative representational notation. For performing with TrakHue two notational systems were developed that are different in function and purpose.

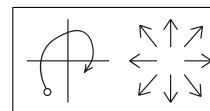


Figure 4. graphical notation for performing TrakHue

Grid line: This notation solution represents the performer's gesture through time. This type of notation is most effective in pieces where precise paths through the colormap are required. As a visually descriptive notation, it is straightforward but requires the gestures be practiced and memorized (see Figure 4 left).

Arrow: This representative style of notation is intuitive and easy to follow in real time, however, it lacks the precision of the grid line notation. The arrows give freedom of interpretation to the performer that is useful in many musical situations. This notation style allows the

performer to more easily match the required gesture to their natural gestures. (see Figure 4 right).

4.3. Compositional Gesture

When composing instrumental works for motion tracking systems, attention must be paid to the pairing of physical, musical and electronic gestures. The composer has to juggle these three layers at all times, while maintaining a unified aesthetic goal. There are two ways of controlling the relationships between these three levels. The first is to create a colormap and then compose a piece based on its characteristic features. In this working method, gesture is notated according to the colormap, not necessarily the performer's motions. This working method tends to create contrasting gestures between the performer and the electronics. The notated physical gestures tend to be less natural due to an indirect relationship to the performer's musical gesture. The other method is to compose instrumental gestures and study the natural gestures of the performer. As the natural gestures are cataloged, a colormap can be created where natural gestures will give the desired aesthetic result. This working method tends to produce a score that feels more natural for the performer.

5. RESULTS

5.1. Flexibility

The scalability and modular design of TrakHue system makes implementing new features fast and easy. As the demands of the performer and composer increase, the structure of the program can adapt to their changing needs. For example, during rehearsals of *Hilbert Space no.1*, we found that the transitions between sound sources needed to be slower and less defined. Aesthetic considerations easily translated into a technical solution: reverb and dynamic effects were quickly included and controlled through the TrakHue.

5.2. Production Costs and Longevity

The associated costs and complicated setup of a performance with live electronics is often a key limiting factor. A musician interested in performing these works can be intimidated by an equipment list that includes rare, expensive and specialized technical apparatus. The use of readily available, low cost equipment increases the portability of a piece. A performance using TrakHue requires a modest setup considering the processes involved.

6. DISCUSSION AND PERSPECTIVE

TrakHue presents an intuitive approach to the gestural control of live electronics. Collaboration between software developer, composer and performer throughout the project resulted in a system that is simple and inexpensive to set up, facilitates compositional projects and improves the performers control of the electronic part.

After TrakHue's successful premiere in a concert setting we identified several areas for improvement.

- Compensating radial distortion by the camera's lens
- Adding inverse perspective warp to derive a perspective-free image when the camera is not directly above the performer.
- Testing an infrared camera to compensate difficult lighting conditions
- More advanced gesture recognition
- Integration of the Gesture Description Interchange Format (GDIF) [3]
- Experimenting in performance with a variety of different musicians both seated and standing

7. ACKNOWLEDGMENT

This work has been funded by the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT). Thanks to Sean Ferguson and Fabrice Marandola for comments and suggestions.

References

- [1] A. Camurri. *Expressive Gesture*. PhD thesis, University of Genoa, 2002.
- [2] Electrotap. Tap.Tools Max 2.3. *Web site: <http://www.electrotap.com>*, retrieved in April 2007.
- [3] A. Jensenius, T. Kvitte, and R. Godøy. Towards a gesture description interchange format. In *Proceedings of the 2006 conference on New interfaces for musical expression*, pages 176–179. Paris, France, 2006.
- [4] E. McNutt. Performing electroacoustic music: a wider view of interactivity. *Organised Sound*, 8(03):297–304, 2004.
- [5] A. Momeni and D. Wessel. Characterizing and controlling musical material intuitively with geometric models. In *Proceedings of the 2003 conference on New interfaces for musical expression*, pages 54–62, Montreal, Canada, 2003.
- [6] T. Place and T. Lossius. Jamoma: A modular standard for structuring patches in Max. In *Proceedings of the International Computer Music Conference 2006*, pages 143 – 146, New Orleans, US, 2006.
- [7] B. Vines, M. Wanderley, C. Krumhansl, R. Nuzzo, and D. Levitin. Performance gestures of musicians: What structural and emotional information do they convey. *Gesture-based communication in human-computer interaction*, pages 468–478, 2003.
- [8] M. Wanderley and P. Depalle. Gesturally-Controlled Digital Audio Effects. In *Proceedings of the Cost G-6 Conference on digital Audio Effects (DAFX-01)*, Limerick, Ireland, 2001.
- [9] M. Wright and A. Freed. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. *Proceedings of the 1997 International Computer Music Conference*, pages 101–104, 1997.