

**Mark Pedersen**

Chailight Productions

PO Box 26

Moreland, 3058

Australia

markp@chailight.com

# User Configurable Intention Maps for Gestural Control and Live Spatialisation in Computer Music performance

## Abstract

*We describe an implementation of a gestural interface for manipulation of computer mediated musical material based on a user-configurable intention map, in order to support more direct expression of the performer's expressive intention than is typically achievable with direct manipulation of parameters.*

*The particular application of intentional maps and gestural control to live spatialisation in surround sound performance is explored.*

## Introduction

Interfaces for computer music performance typically suffer from being directly oriented toward the control of sound production (e.g. directly controlling pitch, duration, tempo). In addition, more subtle controls relating to timbre of synthesised sounds are frequently oriented toward a single parameter among many, making physical control of complex timbres difficult. The overall result is that computer music performance interfaces lack expressivity, inasmuch as they hinder the performer's effort to express complex performance intentions (Rovan 1997). Furthermore, the typical lack of obvious connection between a performer's action and the music being performed is a well-documented issue with respect to audience engagement during computer music performance (Schloss 2003).

## Related Work

Canazza et. al. (2003) describe an abstract control space for the expression of intentions in music performance which incorporates the results of studies on the emotional response of listeners to differing performances of a piece of music. The listeners'

responses are correlated with changes in the performance characteristics in order to identify the broad dimensions of a control space oriented around the expression of emotional intentions and a mapping to performance parameters.

The development of an interface which incorporates knowledge of a mapping from emotional intention to performance parameters holds promise for overcoming the limitations typically associated with computer music interfaces. The challenge presented by Canazza's proposed abstract control space is how to implement a physical interface which retains transparency with respect to intention and performance in such a way that it is engaging for both the performer and the audience.

Designing such a physical interface has drawn upon prior work in gestural control and parameter mapping and interpolation. Both topics are too broad to fully review here, however in the area of parameter mapping particularly relevant work has been done by Hunt & Wanderly (2002) and Verfaillie et. al. (2006). Bencina's (2005) work on parameter interpolation is the inspiration for the system used here. In the specific area of gestural control for spatialisation, work by Schacher (2007) was a significant starting point.

## User Configurable Intention Maps

We propose a gestural interface for manipulation of computer mediated musical material, utilising the concept of a user-configurable intention map, in order to support more direct expression of the performer's communicative intention than is typically achievable with computer music interfaces oriented around direct control of sound production parameters. We augment the control space proposed by Canazza et. al. (2003) with a gesture recognition system that allows for nuanced expression of emotional intention by physical means.

## Description

The architecture of the system consists of a gestural recognition layer, an intention control layer, and an audio control layer, with mappings between each layer. The user is able to configure the gestures to be recognised, the mapping from gestures to intentions, and the mapping from intention to performance parameters. The user can interact with the intention control layer directly (i.e. the mapping to the gesture space is one-to-one) or via complex gestures. Gesture recognition can be used to create an expressive 'shorthand' for complex control positions within the intention space.

The system can be used to control performance at varying levels of detail, for example the tempo of sequence playback as well as the articulation of notes and the timbre of synthesised material. Multiple sound objects (or parts in a polyphonic performance) can be controlled individually. An implementation of the system in Max/MSP has been developed, providing a variety of physical input methods including video recognition, Wacom tablet, and wireless motion sensors. The particular interface explored in this paper is based on the monome 40h (Crabtree and Cain 2007): an 8x8 button grid device augmented with a two dimensional accelerometer and an infra-red distance sensor. The infra-red distance sensor is mounted underneath the device, measuring distance to the surface over which the device is being held.

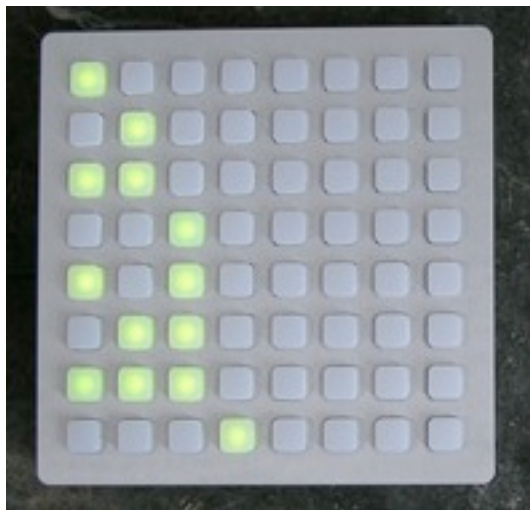


Figure 1. The monome 40h provides an 8x8 button grid with independent visual feedback via LEDs within the buttons.

## Implementation

The implementation in Max/MSP incorporates a number of pre-existing modules.

The gesture recognition layer makes use of Frederic Bevilacqua et al.'s Gesture Follower, which uses a Hidden Markov Model to recognise patterns in multi-dimensional data based on a supplied set of examples (Bevilacqua 2007).

The gesture recognition layer outputs a label for each recognized pattern, corresponding to a position in the intention map. The intention map is based up Oliver Larkin's INT.LIB (Larkin 2007): an interpolated preset interpolation control for Max/MSP. Positions in the intention map are arranged according to the two dimensional distribution of performance intentions proposed by Canazza (2003), namely: Bright, Dark, Hard, Light, Soft, Natural, Heavy. Movement between these positions in the intentional map results in smooth interpolation of control parameters associated with each position, as shown in Figure 2.

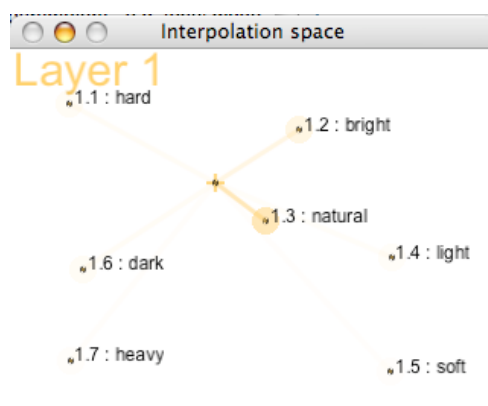


Figure 2. Example layout of interpolation space following Canazza 2003.

User configuration involves two phases:

- training the gesture recognition layer to recognise input patterns corresponding to one of the available intention labels
- configuring/adjusting performance parameters according to the target performance device.

The gesture recognition system takes 3 inputs from the monome 40h: the X and Y axis of the accelerometer input, and distance from infra-red sensor, providing the user with an approximation of 3-dimensional motion sensing. Gesture recognition is not continuous: the user triggers the start and end of an intended gesture by selecting a dedicated button on the monome 40h.

In addition to the mapping of symbolic gestures to intentions, the system also affords direct manipulation of position in the parameter space. Direct manipulation maps the 2 dimensional accelerometer input from the monome 40h into motion in the 2 dimensional intention mapping space, al-

lowing the user to tilt the device in the desired direction. Direct manipulation can be handled either in absolute positioning or as interpolation from the current position, allowing the user to 'nudge' the interpolation cursor in the direction of tilt. Switching between gesture recognition and direct manipulation modes is selected using dedicated control buttons on the monome 40h.

## Gesture Selection

The author's experience arising from both development of and performance with the intention mapping and gesture recognition system suggests that care must be taken to select gestures which are sufficiently orthogonal so as to give robust results from the recognition engine. Simple flicks to the left and the right, forward and back gave strong results. Likewise, vertical movement is easily recognised. Combinations of the above resulted in some ambiguous results in terms of recognition. For instance, rapid side-to-side motion was not always distinguished from rapid front back motion, presumably due to the difficulty of isolating rapid movements to be only along one axis.

This issue highlights another challenge with the gesture recognition, in that the user must train themselves to repeat gestures reliably enough when training the system that the gestures are recognisable. Limited variation in the gestures helps the robustness of the recognition, as minor variations are mapped to the same intention, however gross variations will reduce the effectiveness of the recognition algorithm.

A further limitation is that gestures are mapped to intention categories without any scalar factor indicating the "strength" of the gesture. In comparison, typical human interaction, for instance with an orchestra conductor, would afford recognition of symbolic gestures as well as the degree or strength of the gesture through factors such as speed or expansiveness of the gesture. Alternative inputs, such as pressure sensitivity, which remain orthogonal from the recognition system, are being considered as means of overcoming this problem.

## Mapping to Parameter Space

In the context of computer music performance, the available parameters for intention mapping are vast. However it is sensible to restrict the possibilities to those which are typical for musical performance.

Working within the framework of using traditional instruments such as a piano to render sequenced note information, Canazza et. al. focus on mappings

from intention states to energetic and kinetic parameters of performance such as tempo and dynamics.

In terms of traditional instrument performance, Paine (2007) groups performance parameters broadly under the headings of Dynamics, Pitch, Vibrato, Articulation, Attack and Release as the mediating parameters between the physical instrument and the desired tone colour. For specific instruments, additional parameters (such as bow pressure, position and speed for violin or cello) also contribute to the desired tone colour.

We adopt a similar approach, replacing vibrato with a more general and direct set of parameters relating to timbral brightness and noise. Given the possibility of using the intention mapping technique to manipulate sequenced material, it is also possible to influence the tonal mode of the playback, in much the same way that a composer may choose to work in a specific key, or modulate to a new key in order to express an emotional state or intention.

Mapping of these abstract qualities needs to be made specific for any particular synthesis engine. For example, brightness may be expressed in terms of filter cutoff frequency and resonance, EQ or even reverb amount. The system allows for parameters to be communicated via MIDI to external synthesis engines or communicated directly to embedded effects that have been developed within the system itself.

## Application to Live Spatialisation

Spatialisation of sound for surround-sound performance can be handled as simply another parameter in the intention space, however a number of challenges emerge:

- spatialisation performances typically manipulate multiple sound objects, while the existing intention mapping architecture is designed for the manipulation of only a single sound object
- merely associating a single spatial position with an expressive intention is overly restrictive
- directly manipulating a parameter which effects the rendering of a sound object on a 2 dimensional plane is potentially confusing for a performer who is navigating an abstract 2 dimensional plane.

In terms of the first issue, control of multiple sound objects is afforded by the use of layers in

Larkin's (2007) interpolation space. Each layer in the interpolation space is configured with an identical intention map layout, with each layer connecting to the control parameters for the associated sound object. Thus separate sound objects, with separate synthesis engines, can be controlled from a common set of intentional gestures. Selection of the active sound object is achieved through dedicated buttons on the monome. Simultaneous control is limited currently to affecting all layers simultaneously.

A significant disadvantage of the layering approach is the issue of additional cognitive load generated by using a single interface device to control multiple virtual objects. Context shifting between the objects tends to break the continuity of parameter settings and the control surface, requiring the user to keep track of which object they are controlling and beware of making unintended changes to the active object.

This issue is somewhat alleviated by having a common intention map across all objects, and common gestures for each intention. Interpolation from the current position in the intention map to the desired goal helps to reduce the incidence of discontinuities arising from context switching, and for this reason gestural control or 'nudging' is preferred over direct manipulation of the intention mapping space when working with multiple objects and layers.

In addressing the issue of decoupling spatial position and intention, a number of approaches have been taken. While it may suit the compositional/performance requirements of some pieces to associate a particular performance intention or tonal quality (dark, bright, heavy, etc) with a particular position in the sound field, being limited to *always* positioning sound objects in the "bright corner" or the "dark corner" is not desirable. Hence a more sophisticated approach to mapping intention to spatialisation parameters is required.

In the first instance, we associated complex trajectories with each intention. Thus allowing, for instance, a shift into the "light" intention space to results in the sound object bouncing swiftly around the sound field, whereas a move into the "heavy" intention space slows the movement down to settle in the centre. Care must be taken to manage switching between trajectories to avoid unintended discrete jumps from one spatial position to another.

The current implementation associates fixed trajectories with each intentional position and applies changes in trajectory once the sound object passes through a common point. Trajectories can be recorded by the user directly and stored for later use. This allows the user to develop a library of spatialisation trajectories that can be linked with

performance intentions when configuring the system. Other scalar factors such as translation in the XY plane, and scaling of trajectories can likewise be applied.

Trajectory coordinates are not the only parameter available for control. Speed, and direction along the trajectory (forward/reverse), are controllable and, as scalar values, are easily interpolated using the current architecture.

Apart from automated control of spatial position via fixed trajectories, direct manipulation of the spatial position of a sound using the monome 40h is also possible. In this instance, a separate spatialisation layer is added to the parameter interpolation control, in which the preset reference points are positioned to mirror the speaker layout of the performance space.

Movement in this layer directly maps on to movement in the sound field. Again, the use of additional layers of control adds additional cognitive load, subjectively speaking, however the context switching is not as problematic as when switching between multiple objects with an overlaid control space. While possible, maintaining control of a separate intention map and spatial position map for multiple objects has proven challenging in performance, particularly as context switching between layers becomes more problematic.

## Visual Feedback

The advantage of the monome 40h as a control device is that it affords a degree of visual feedback within the device itself, which is not available in other devices such as Wacom tablets or handheld accelerometers such as the Wiimote, although these devices may provide more detailed feedback via computer displays.

Selection of sound objects is achieved by selecting a button along the bottom row of the monome. Buttons representing available objects (up to eight) are lit, and the selected object flashes.

Likewise, toggling between the timbral control layer and the spatial control layer can be done by selecting a button on the top right of the monome, with a flashing status indicating spatial control.

Position within the control space (for the purpose of direct manipulation) can be indicated by a "cursor" on the monome, which moves according to the tilt of the device. However, this form of feedback was found to be inadequate for performance purposes compared to the display of the interpolation space on computer screen, which provided details of the

various significant positions within the space. In addition, the spatial control position could also be monitored concurrently on the computer screen, whereas the monome 40h display required context switching.

## Future Work

A number of areas of future work remain. Foremost is a more formal evaluation of the intention mapping system with a range of performers to assess usability in a performance context, particularly looking at issues of cognitive load in surround surround performance, and robustness of gesture recognition.

Switching between fixed trajectories is somewhat of a limitation in terms of choice of trajectories and achievement of a similar level of interpolation in the spatial parameter space as in the timbral parameter space. Interpolation between trajectories requires calculation of intermediate trajectories which can be applied progressively without noticeably interrupting the current motion of the object. A solution to this problem, based on approaches to non-intersecting graph morphing, is currently being investigated.

Refinement of the interface, including the addition of pressure inputs or other means of indicating the intended strength of a gesture is highly desirable.

Given the limitation on the number of directly controllable sound objects, hierarchical grouping and control along the lines suggested by Schacher (2007) has been identified as a potentially fruitful direction for further research into the management of multiple objects within the intention mapping architecture. Linking gestures to complex transitions within the parameter space, either for individual sound objects or multiple groups would also extend the expressive potential of the system.

## Conclusion

The system developed addresses issues with expressivity and transparency in computer music performance interfaces by providing the performer with a spatial representation of expressive options that can be navigated either directly or by means of configurable gestures, allowing complex control of performance parameters in a manner more closely linked to the performer's own expressive intention.

In particular, the application of intentional maps to control of live spatialisation has been explored and a number of issues identified and workable solutions developed.

## References

- Bevilacqua, F., Guédy, F., Schnell, N., Fléty, E., Leroy, N., 2007, "Wireless sensor interface and gesture-follower for music pedagogy", *Proceedings of the International Conference of New Interfaces for Musical Expression (NIME 07)*, New York, NY, USA, pp 124-12
- Canazza, S., et al., 2003, "An Abstract Control Space for Communication of Sensory Expressive Intentions in Music Performance," *J. New Music Research*, vol. 32, no. 3, 2003, pp. 281-294.
- Crabtree, B. and Cain, K. 2007, monome, Philadelphia, USA, <http://www.monome.org>
- Hunt, A. and Wanderley, M. 2002, "Mapping Performance Parameters to Synthesis Engines". *Organised Sound*, 7(2):97-108, 2002.
- Larkin, O., 2007, "Int.Lib - A Graphical Preset Interpolator For Max MSP", *Proc. International Computer Music Conference 2007*, Copenhagen
- Paine, G., 2007, "Interfacing For Dynamic Morphology In Computer Music Performance", *International Conference on Music Communication Science*, Sydney, Australia
- Schacher, J., 2007, "Gesture Control of Sounds in 3D Space" *Proceedings of the 2007 New Interfaces for Musical Expression Conference (NIME 07)*, New York, NY, USA.
- Schloss, W. A., 2003, "Using Contemporary Technology in Live Performance: The Dilemma of the Performer", *Journal of New Music Research*, Volume 32, Issue 3 September 2003 , pages 239 - 242
- Rovan, J., Wanderly, M., Dubnov, S., Depalle, P., 1997, "Instrumental gestural mapping strategies as expressivity determinants in computer music performance" *Proceedings of the Kansei - The Technology of Emotion Workshop*, Genova - Italy, Oct. 1997.
- Verfaille, V., Wanderley, M. and Depalle, P., 2006 "Mapping Strategies for Gestural and Adaptive Control of Digital Audio Effects", *Journal of New Music Research*, 2006, Vol. 35, No. 1, pp. 71 - 93