

Towards a Model for Instrumental Mapping in Expert Musical Interaction

Andy Hunt, Marcelo M. Wanderley* and Ross Kirk

York Music Technology Group
University of York
Heslington – UK
{adh, ross}@ohm.york.ac.uk

*Analysis-Synthesis Team
Ircam – Centre Pompidou
Paris – France
mwanderley@acm.org

ABSTRACT

This paper reviews models of the ways in which performer instrumental actions can be linked to sound synthesis parameters. We analyse available literature on both acoustical instrument simulation and mapping of input devices to sound synthesis in general human-computer interaction. We further demonstrate why a more complex mapping strategy is required to maximise human performance possibilities in expert manipulation situations by showing clear measurements of user performance improvement over time. We finally discuss a general model for instrumental mapping, by separating the mapping layer into two independent parts. This model allows the expressive use of different input devices within the same architecture, or conversely, the use of different synthesis algorithms, by only changing one part of the mapping layer.

1. INTRODUCTION - INTERACTIVE SYSTEMS AND EXPERT MANIPULATION

In acoustic musical instruments the sound generation device is inseparable from the human control device, and this yields complex control relationships between human performers and their instruments. However, in the case of electronic musical instruments (where the interaction - or *input* - device is independent of the sound synthesis device) there is no implicit mapping of one to the other [Win95]. Too often the instrument design will default to a single control device corresponding to a single musical (synthesis) parameter – a 'one-to-one' mapping.

It is also known that acoustic musical instruments have evolved over many hundreds of years into systems with greatly superior interaction capabilities than today's state-of-the-art real-time processors. Thus, the study of mapping strategies seems to be of paramount importance, since in many instruments the input variables are inter-related in both complex and non-linear relationships. Simple mapping strategies acting in isolation do not usually allow an expert performer to simultaneously control multiple parameters in an expressive way [RWDD97].

In different real-life scenarios, similar situations occur (e.g. when driving a car). The problem in these situations is that there is not necessarily an obvious model of the mapping strategies that relate the input devices being manipulated by the user/performer to the system being controlled. Recent studies have shown [HK99] that even in those situations that are not directly modelled on existing systems (i.e. acoustic instruments), complex mapping strategies do perform better than one-to-one relationships. The question to be investigated in this case is the way in which these relationships should be set up.

1.1 What do we mean by *Mapping*?

In this article, the word 'mapping' refers to the liaison or correspondence between *control parameters* (derived from *performer actions*¹) and *sound synthesis parameters*.² This concept is illustrated in fig. 1 that represents a general computer-based musical instrument; what might be called a 'composite electronic' musical instrument.

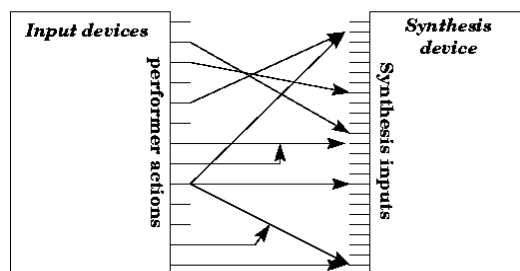


Figure 1: Mapping of performer actions to synthesis parameters.

1.2 A Discussion of the Role of Mapping

A first interesting point to consider is the role of mapping in interactive systems. Two main points of view exist:

- *Mapping is a specific feature of a composition;*
- *Mapping is an integral part of the instrument.*

We subscribe to the second point of view, that mapping is part of the instrument, and therefore influences the way a performer makes use of it in different contexts.

1.3 Types of Mapping

One can devise two main directions from the analysis of the existing literature on mapping:

- The use of generative mechanisms (e.g. neural networks) to perform mapping [LW92] [FeI94] [MZ97] [Mod00].
- The use of explicit mapping strategies [BPMB90] [RWDD97] [MFM97] [WSR98] [HK99], among others.

The main difference consists in the chosen approach to mapping: either the use of a method that provides a mapping strategy by means of internal adaptations of the system through training, or the proposition of mapping strategies which explicitly define the relationships. Each approach has its own advantages and drawbacks. In this article we will focus on the second approach.

2. EXPLICIT MAPPING STRATEGIES

The available literature generally considers mapping of performer actions to sound synthesis parameters as a *few-to-many* relationship, mostly in the case of synthesis by 'signal' models, such as source-filter or additive synthesis.

¹ Sometimes called *performer gestures*. See [CW00] for a review.

² Note that we do not include in the concept of mapping the actions related to data preparation, such as segmentation, scaling, limiting, etc.

But considering any two sets of parameters, three basic strategies relating those parameters can be devised: *one-to-one*, *one-to-many* or *many-to-one*. Obviously, a combination of the above basic strategies is also possible, termed *many-to-many*. This has been presented in the musical literature in different ways.

Ryan [Rya91] has categorised different strategies by proposing Euclidean analogies (points, lines and curves)³. Therefore, mapping one event to a set of musical parameters would be between a point and a curve. Conversely, various performer actions relating to one musical parameter would be considered as a curve to a point. Other possible relationships could then be between lines, curves, and so on.

Rovan et al. [RWDD97] have identified the three basic categories, using the words *convergent* (many-to-one) and *divergent* (one-to-many).

Garnett and Goudeseune [GG99] have also considered the general case with three strategies: *direct mapping* from individual controls to individual synthesis parameters, *one control driving several parameters* and *one parameter being driven by several controls*.

3. REVIEW OF PREVIOUS WORKS RELATED TO MAPPING

One interesting work on mapping was published by Bowler et al. [BPMB90]. They presented a general way to map N articulation parameters to M synthesiser control parameters through interpolation of the M control parameters placed at coarsely quantised positions in the N dimensional performance space. This strategy reduced the total number of multiplications from $M \cdot 2^N$ to $M \cdot N$, allowing real-time control of additive synthesis in a transputer network, at times when memory and processor speed were much different from our current technological standards.

Choi et al. [CBG95] [Cho00] proposed a mapping of a point in a *control space* to a point in a *phase space*⁴ by identifying the co-ordinates of the cell it is in, and relating them to the corresponding cell and co-ordinates in the phase space. A refinement of this method is reported in [GG99].

3.1 Mapping Strategies for the Simulation of Traditional Instrument Performance

Favilla [Fav96] proposed a mapping of empty-handed gestures to pitch-bend ornaments (called *gamakas*) in the context of Indian music performance. It is implemented through the use of non-linear maps consisting of sloping and steep curves. These maps allow the use of different ranges of gesture amplitudes captured by electromagnetic proximity controllers.

Rovan et al. [RWDD97] have shown that in an expert interaction context the input device (connected to a sound synthesis generating system) must match the original instrument's characteristics. The specific example cited is of a clarinet's embouchure, which needs to be set to an appropriate level before the 'air-flow' parameter has any effect – a *biasing* effect.

This approach relates to the use of *timbre spaces* [Wes79], meaning that the mapping in this case was implemented between performer actions and the *axes of the timbre space*, not individual additive synthesis parameters. The authors also show that the resulting instrument's expressivity is much dependent on the specific mapping strategies employed. Using the same set-up, differences in expressivity were reported depending on the mapping strategy chosen.

An interesting point suggested in this paper is the adaptation of the mapping strategy to the user level: beginners may profit from simpler or more direct mappings, whilst skilled musicians take advantage of complex mappings.

One direct conclusion from this study [WD99] is that one can consider mapping as providing the level of interaction between the user and the machine. A continuum between two mapping strategies may allow a control level from a *macroscopic* (e.g. phrasing, rhythm) to a *microscopic* (e.g. precise timbre control) level, such as suggested in [Sch90].

3.2 Mapping Strategies for General Interactive Instruments

For the case of general interactive instruments, where new instruments may not have any acoustic counterpart, formalisms must be proposed.

Mulder et al. [MFM97] suggested the use of geometrical shapes as an attempt to reduce the cognitive load of manipulating several simultaneous parameters. The user then virtually manipulates a geometrical shape, whose parameters are mapped to sound synthesis parameters.

Recently it has been shown that the use of complex mapping strategies does improve performance of complex tasks. This work is summarised in the following section.

3.3 The Effect of Mapping on Instrument Effectiveness

A series of tests was carried out at the University of York UK in order to study the effectiveness of different interfaces when used for a real-time musical control task. The data that was gathered was used to compare how a group of human test subjects performed on different interfaces over a period of time. The following three interfaces were chosen for the study:

- A set of on-screen sliders controlled by a mouse.
- A set of physical sliders moved by the user's fingers.
- A multiparametric interface which uses a mouse and one hand and two sliders in the other.

The first interface (mouse) represents the standard way of operating computer interfaces by clicking and dragging with a mouse pointer. The second (sliders) gives the user the opportunity to operate all parameters simultaneously but with one slider control per parameter, rather like a small mixing-desk. The third (multiparametric) presents the user with an interface more reminiscent to a conventional musical instrument, with many cross-mappings of parameters.

Several users performed identical sets of tests using each of the above interfaces over a period of time. These tests involved listening to increasingly difficult audio signals (consisting of variations in pitch, volume, timbre and stereo panning) and attempting to recreate them on the interfaces. Details are given in [HK99].

The first two interfaces ('mouse' and 'sliders') have direct *one-to-one* mappings between each control input and one of the four controllable audio parameters. The 'multiparametric' interface is different. It uses the same hardware as the other interfaces (the mouse and a set of physical sliders), but it uses them in two very different ways.

a) Firstly the system expects the user to expend some physical energy to continuously activate the system. Sound is only made when the mouse is *moved*. The sound's volume is proportional to the *speed* of mouse movement. This ensures that the user's physical energy is needed for any sound to be made.

b) Secondly, there is only one direct one-to-one correspondence (mapping) between a physical control and an internal sound parameter (for panning). All other mappings are complex (many-to-many).

³An interesting overview of this article and a discussion on mapping is presented in chap 2, section 4 of [Mar00].

⁴The n-dimensional Euclidean space where points correspond to states of a parameterised system.

The volume, pitch, timbre and panning are controlled by combinations of the mouse position and the position of two sliders, as shown here:

- Volume = speed of mouse + mouse button pressed + average position of two sliders.
- Pitch = vertical position of the mouse + speed of movement of slider no. 2.
- Timbre = Horizontal position of the mouse + difference in the two slider positions.
- Panning = Position of slider no. 1.

Qualitative analysis was carried out by interviewing every subject after each set of tests on each interface. They were asked how they felt about their performance and the interface. Towards the end of the session they were asked to sum up how they had done overall and to compare the different interfaces.

The interviews indicated that the majority of users enjoyed playing the multiparametric interface and thought (quite enthusiastically) that it had the best long-term potential. The following three comments summarised from the users indicate that there was something about the multiparametric interface that allowed spatial thinking that was entertaining and engaging.

- The multiparametric interface allowed people to think gesturally, or to mentally rehearse sounds as shapes.
- The majority of users felt that the multiparametric interface had the most long-term potential. Several people commented that they would quite like to continue to use it outside the context of the tests!
- Several users reported that the multiparametric interface was fun.

In contrast the sliders interface often elicited the opposite response and the majority of people found the sliders interface confusing, frustrating or at odds with their way of thinking. This was often focused on the requirement to mentally break down the sound into separate parameters. Maybe the test subjects were experiencing cognitive overload with this task on this particular interface.

Since both the sliders and multiparametric interfaces allowed the user to have continuous control over all four sound parameters, we can conclude that the above differences can be accounted for by the parameter mapping alone. In other words:

Mapping strategies which are not one-to-one can be more engaging to users than one-to-one mappings.

These qualitative comments were supported and extended by the quantitative results. Every test result was stored on the computer and later given a score by both a computer algorithm and a human marker, giving over 4000 tests over a period of several weeks [Hun99][HK00].

It was clear that the multiparametric interface allowed users to perform in a different manner to the other two interfaces:

- For the simplest audio test the scores were lower than those for the mouse or sliders, but they improved over time.
- The scores got *better* for *more complex tests* and were much higher than the other two interfaces, for all but the simplest tests.
- There is a good improvement over time across all test complexities.

In other words the multiparametric interface, which differed from the others only in the *mapping* employed, showed dramatically improved results over time. Figure 2 shows a three dimensional graph which portrays the average results on the multiparametric interface for all the test subjects. The dimension to the right represents the *complexity* of the audio tests, whilst that to the left shows *session number* (i.e. subjects gaining familiarity with the system). The vertical

axis shows the average score for all test subjects. The graph shows two distinct tilts. The upward tilt towards the left shows that the interface allows people to improve their scores over time (a feature not shown by the other two interfaces for all but the simplest tests). The upward tilt to the right shows the remarkable feature that better scores were achieved for *more complex tests*.

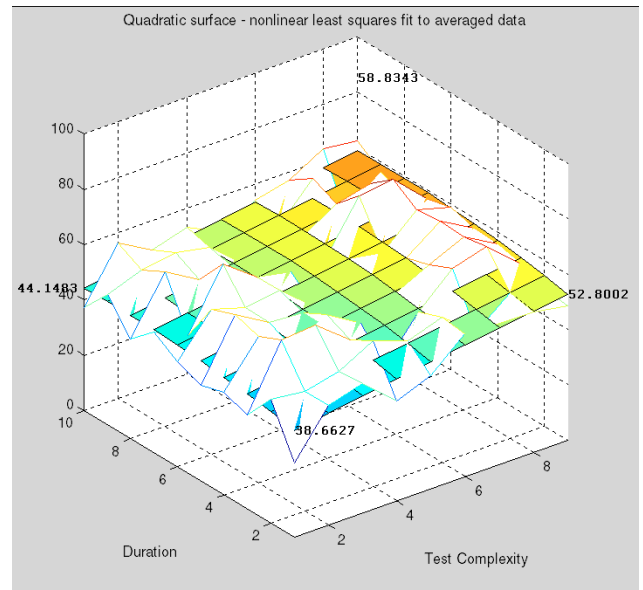


Figure 2: Subjects' performance evolution over time

4. TOWARDS A GENERAL MODEL OF MAPPING FOR EXPERT INTERACTION

It may be interesting to provide the performer with control of higher-level parameters than frequencies, amplitudes and phases of sinusoidal partials in additive synthesis, or carrier to modulation ratios in frequency modulation. One approach comes straight from research on timbre spaces [Wes79][VB94] or speech synthesis. For instance, frequencies of the first two formants of French vowels can be distributed in a plane, the vowel formant frequencies amounting to a *vocalic triangle*. Navigation in this plane then allows the interpolation between the different vowels.

Wanderley et al. [WSR98] have proposed a real-time synthesis system called ESCHER, where they divided the mapping layer in *two independent layers*⁵. This was done by the definition of an intermediate abstract parameter layer, based on perceptual characteristics of sounds or arbitrarily chosen by the composer or performer. The interest in this approach is that the first layer depends only on the choice of the input device for a given set of abstract parameters, whilst the mapping from the *abstract* parameter set to the actual synthesis variable is a function of the synthesis algorithm used [WD99].

In the same direction, Garnett and Goudeseune [GG99] report a method for *automatically generating* perceptual parameters without manually crafting the timbre space, what they called the *timbre rover*.

This general model (*with the inclusion of complex mappings, and derivatives of the user's input in the first mapping layer*) should allow composers and instrument designers to more carefully 'fit' the output of their novel interfaces to more well-developed 'synthesis engines'. A widespread acceptance of this model may result in synthesis engines being developed

⁵ Similar ideas have been proposed in [UK98], where the mapping layer is also divided into an input mapping and an output mapping layer. This can also be considered as the way the system proposed in [MFM97] works.

with built-in complex mappings, so that their inputs are more readily accessible in terms of perceptual parameters.

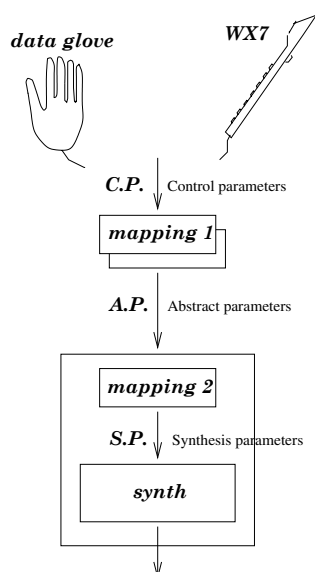


Figure 3: Two-level mapping using an intermediate user-defined abstract parameter layer [WSR98].

5. CONCLUSIONS

In this paper we have reviewed the available literature on the mapping of performer actions to sound synthesis parameters. We have discussed the role of mapping in both the simulation of existing instruments and in general human-computer interaction.

We contend that the topic of parameter mapping deserves more study in its own right, as it dramatically affects the perceived operation of electronic musical instruments. Complex mappings cannot be learned instantaneously, but then again, we have never expected this from acoustic instruments. Complex mappings also appear to allow users to develop strategies for controlling complex parameter spaces.

In summary, we recommend that complex mappings (using cross-coupling of input parameters to synthesis parameters, and derivatives of input parameters related to the performer's energy) be utilised widely in the next generation of electronic performance instruments.

ACKNOWLEDGEMENTS

Thanks to John Szymanski for rendering the graph in fig. 2.

The second author would like to acknowledge contributions, discussions and suggestions by Ph. Depalle, Butch Rován, S. Dubnov, N. Schnell and X. Rodet.

The second author is supported by a grant from CNPq/Brazil.

REFERENCES

- [BPMB90] Bowler, I., Purvis, A., Manning, P., and Bailey, N.. 1990. "On Mapping N Articulation onto M Synthesiser-Control Parameters." In *Proc. ICMC'90*, pp. 181-184.
- [CW00] Cadoz, C., and Wanderley, M. 2000. "Gesture-Music." In M. Wanderley and M. Battier, eds. *Trends in Gestural Control of Music*. Ircam – Centre Pompidou.
- [CBG95] Choi, I., Bargar, R., and Goudeseune, C.. 1995. "A Manifold Interface for a High Dimensional Control Space." In *Proc. ICMC'95*, pp. 385-392.
- [Cho00] Choi, I. 2000. "A Manifold Interface for Kinesthetic Notation for High-Dimensional Systems" In M. Wanderley and M. Battier, eds. *Trends in Gestural Control of Music*. Ircam – Centre Pompidou.
- [Fav96] Favilla, S. 1996. "Non-Linear Controller Mapping for Gestural Control of the Gamaka." In *Proc. ICMC'96*, pp. 89-92.
- [Fel94] Fels, S. 1994. *Glove-Talk II: Mapping Hand Gestures to Speech Using Neural Networks - An Approach to Building Adaptive Interfaces*. PhD Thesis, University of Toronto, Canada.
- [GG99] Garnett, G., and Goudeseune, C. 1999. "Performance Factors in Control of High-Dimensional Spaces." In *Proc. ICMC'99*, pp. 268 – 271.
- [HK99] Hunt, A., and Kirk, R. 1999. "Radical User Interfaces for Real-time Control." In *Proceedings of the Euromicro Conference*. Milan.
- [HK00] Hunt, A., and Kirk, R. 2000. "Mapping Strategies for Musical Performance." In M. Wanderley and M. Battier, eds. *Trends in Gestural Control of Music*. Ircam – Centre Pompidou.
- [Hun99] Hunt, A. 1999. *Radical User Interfaces for Real-time Musical Control*. DPhil thesis, University of York UK.
- [LW92] Lee, M., and Wessel, D. 1992. "Connectionist Models for Real-Time Control of Synthesis and Compositional Algorithms." In *Proc. ICMC'92*, pp. 277-280.
- [Mar00] Marrin-Nakra, T. 2000. *Inside the Conductor's Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture*. PhD Thesis. MIT Media Lab.
- [MZ97] Modler, P., and Zannos, I. 1997. "Emotional Aspects of Gesture Recognition by a Neural Network, using Dedicated Input Devices." In *Proc. of the Kansei Workshop, Genova*, pp. 79- 86.
- [Mod00] Modler, P. 2000. "Neural Networks for Mapping Gestures to Sound Synthesis." In M. Wanderley and M. Battier, eds. *Trends in Gestural Control of Music*. Ircam – Centre Pompidou.
- [MFM97] Mulder, A., Fels, S., and Mase, K. 1997. "Empty-Handed Gesture Analysis in Max/FTS." In *Proc. of the Kansei Workshop, Genova*, pp. 87-91.
- [RWDD97] Rován, J., Wanderley, M., Dubnov, S., and Depalle, P. 1997. "Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance." In *Proc. of the Kansei Workshop, Genova*, pp. 68-73.
- [Rya91] Ryan, J. 1991. "Some Remarks on Musical Instrument Design at STEIM." *Contemporary Music Review* 6(1):3-17.
- [Sch90] Schloss, W. A. 1990. "Recent Advances in the Coupling of the Language MAX with the Mathews/Boie Radio Drum." In *Proc. ICMC'90*, pp. 398-400.
- [UK98] Ungvary, T., and Kieslinger, M. 1998. "Creative and Interpretative Processmilieu for Live-Computermusic with the Sentograph." In Kopiez und Auhagen, eds. *Controlling Creative Processes in Music*. Schriften Musikpsychologie und Musikästhetik 12. Peter Lang Verlag, pp. 173-227.
- [VB94] Vertegaal, R., and Bonis, E.. 1994. "ISEE: An Intuitive Sound Editing Environment." *Computer Music Journal* 18(2): 21-29.
- [WSR98] Wanderley, M., Schnell, N., and Rován, J.B. 1998. "Escher - Modeling and Performing Composed Instruments in Real-Time." In *Proc. IEEE SMC'98*, pp. 1080-1084.
- [WD99] Wanderley, M., and Depalle, P.. 1999. "Le Contrôle gestuel de la synthèse sonore." In H. Vinet and F. Delalande, eds. *Interfaces homme-machine et création musicale*. Paris: Hermes Science Publishing, pp. 145-163.
- [Wes79] Wessel, D. 1979. "Timbre Space as a Musical Control Structure." *Computer Music J.* 3(2): 45-52.
- [Win95] Winkler, T. 1995. "Making Motion Musical: Gestural Mapping Strategies for Interactive Computer Music." In *Proc. ICMC'95*, pp. 261-264.