Action and Perception in Interactive Sound Installations: An Ecological Approach

Jan C. Schacher
Zurich University of the Arts,
Institute for Computer Music and Sound Technology
Baslerstrasse 30
CH-8048 Zürich
jan.schacher@zhdk.ch

Abstract

In this paper mappings and adaptation in the context of interactive sound installations are discussed. Starting from an ecological perspective on non-expert audience interaction a brief overview and discussion of mapping strategies with a special focus on adaptive systems using machine learning algorithms is given. An audio-visual interactive installation is analyzed and its implementation used to illustrate the issues of audience engagement and to discuss the efficiency of adaptive mappings.

Keywords: Interaction, adaptive mapping, machine learning, audience engagement

1. Introduction

"In the beginning was the deed" Wittgenstein quoting Goethe's faust, in *Culture and Value* [1]

"To perceive is to have a body and to have a body is to inhabit a world." Taylor Carman in *Merleau Ponty* [2]

One of the challenges of performing or composing music with computers in realtime is to establish a meaningful and cognitively sustainable relationship between a directed physical gesture and the virtual digital instrument which produces the desired sound. This relationship traverses a number of transformation layers which are governed by their own laws and the whose nature gradually change in a way that potentially generates a strong separation of the initial action from the synthesis process. The sound produced finally travels back to the interacting person and closes the action-perception loop which is critical for a well controlled and executed performance.

In the extensive literature about mapping in the field of electronic music and NIME the research has been focused on describing and formalizing the mechanisms that govern mapping from a structural point of view. The prevailing interaction paradigm in this research is one of the instrumentalist performing with an electronic instrument in the same role and with the same intentionality as with a traditional instrument. The interaction process is segmented into separate problem domains which are linked by abstraction layers. These are needed in order to translate the meaning or the energy of the original gesture into the appropriate form or domain for the intended technical realization. Of course this functional method is imposed in large parts by the structure of the tools used to create the desired results.

The ecological view on mapping takes into account a wider scope of the original action, including aspects which are non-technical but rather psychological and perceptual and are more closely related to a given socio-cultural context and the perceptual or cognitive aspects of expressing musical intentions through digital means. It considers the mapping as a mediator of essential performance aspects.[3] Beyond the technical link between the gesture acquisition device and the synthesis algorithm lies the core issue of perceiving the interaction and the re/en/acting (to) the result in an way that is influenced by a richer environment than just the instrument in play.

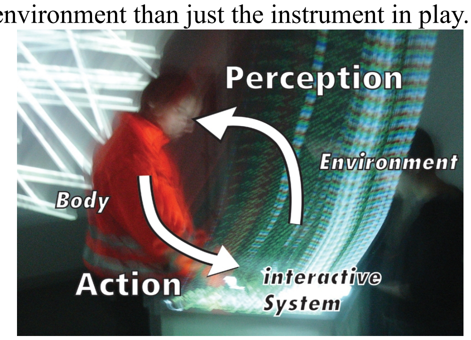

Figure 1. The Action-Perception Loop

Embodied interaction looks at this issue from a ontological point of view. There can be no action without a body, the body cannot interact with the world without us having a knowledge, an internal representation of the situatedness in our own environment. (Figure 1.)

2. Interactive Sound Installations

In interactive sound installations the problem of mapping and its effectiveness shifts depending on the complexity of the deployed interactive model but also on the visitor's level on a novice-to-expert scale. The interaction and mapping cannot count on an extended learning process by which the visitors could grasp the work's essential qualities, and therefore an installation will aesthetically fail to satisfy if it doesn't contain an inherent richness of relationships between its constituting elements. *“In interactive sound space installation, the mapping is the procedure that translates gestural input to audio output. The mapping largely defines the aesthetic expression of the space, and simultaneously plays a significant role in audience engagement.“*[4] Traditionally interactive sound installations reduce the complexity in the interaction compared to expert performance systems and make up for this simplicity by using more dense or varied musical structures. When taking visitor behavior into account at the conceptual stage of an installation, a more experience oriented interaction pattern can be designed, a gesture type can be specified which facilitates playfulness, rewards curiosity and ultimately builds a more intuitive bridge between gesture and sound or action and perception. The quality of a gesture and its directed or indirectedness clearly can denote the intentionality in a ecological sense.[5] By analyzing not only the gesture space but also the visitor's attitude in the entire interaction space and the behavior patterns in time spans that exceed the duration of a normal gesture, information about the intention and directedness of a gesture can be obtained.

3. The Learning Process

Interacting with an instrument or a reactive installation always entails a process of learning, exploring and adapting. Exploratory control and 'babbling' are strategies often used in the performance of electronic music.[6]

3.1 Visitor interaction and learning

In interactive installations this searching behavior is essential but only happens in an undirected and open way, since there is not enough time to accumulate more than a very superficial experience of the system. It is crucial that the audience engage immediately with an interactive work. To improve the engagement, an interactive system should be able adapt to the intentions, intensity or effectiveness of the behavior exhibited by the audience in order to increase the impact of the critically short first exploration of an installation. The visitor might also adapt to the task he or she perceives as being part of the interaction and thus alter their behavior towards the system. This learning process forms part of what could be called interaction literacy. The sum of experience in interaction with technical systems enables us to adapt to new situations by projecting a tentative gesture and optimizing the gesture based on the feedback the system provides. This constitutes the first outer circle of adaptation in the action-perception relationship. In the design of the system a set of typical audience reactions needs to be taken into account., thus informing the way the outer adaptation happens.

3.2 Machine learning

The learning process on the machine side usually lacks the open and exploratory behavior. However in the more narrowly defined scope tightly coupled adaption processes *do* constitute learning and can make the interaction evolve beyond a static mapping situation. Artificial intelligence processes such as neural networks and genetic algorithms are commonly used for these tasks. They mostly depend or prior training or the concise definition of fitness conditions. This imposes bounds to the variation space such an algorithm can explore. Yet the overall expression space is much larger than these bounds, since the inner adaption loop of the machine learning algorithm also depends on the self adapting behavior of the visitor. This extended scope can only be fully understood, if the external ecological factors form part of the model.

4. Mapping strategies

Arfib et al. (2002) address the subject of mapping between a gesture and a sound synthesis algorithm in a somewhat ecological way. The strategy of mapping between gesture data and a synthesis model is to use perceptual spaces rather than gesture or synthesis parameter spaces.[7] They differentiate between related-to-gesture perception parameters and related-to-sound perception parameter spaces, which between them form an abstract interpretation domain which is completely decoupled from the specific affordances of the gesture acquisition device and the specific parameter space of the sound synthesis algorithm. The mapping takes place in this abstract domain and any feedback to the performer is expressed through the state of the intermediate space rather than that of the synthesis system or the gesture analyzer. A variety of feedback scenarios is tested, but exhibit a relatively static structure.

Other mapping strategies extend from immediate and direct mappings to indirect mediated transmissions. Several layers of mapping mediate from a device specific representation to an abstracted domain and from there to the synthesis algorithm.[8] The structure of the mapping and its topology again remains static and is the fruit of an a priori exploration of the expression space by the composer/performer.

5. Adaptive Mapping

“Adaptive mapping is a computational method that utilizes feedback to continuously transform a mapping, enabling continuously novel output in response to the behavior of the participant.“[9] Not all of the common mapping topologies are appropriate for adaptive mapping. The straight one-to-one mappings can only be adaptive in their scaling or transfer function, something which most software environments implement as auto-scaling. One-to-many mappings tend to use meta-descriptors to bundle complex parametric spaces into control of salient features. Here adaptation can change weightings but rarely changes

the fundamental mapping structure in such a way that meta-descriptors would completely change their semantic relationship with the synthesis algorithm. The projection from the meta-space onto the parameter space remains persistent. *"[The] variation range clearly defines the boundaries of the domain for the performer to explore. Adapting mapping to the performer creates evolutions of these boundaries, for example when gesture is confined in a small part of the gesture perceptual space whereas the sound explores a greater part of the sound perceptual space."*[7] In many-to-many mappings like the ones specifically necessary for multi-modal gesture interfaces the adaptation can change not only the weighting of the connections but also the topology of the node-connection mapping itself.

5.1 Artificial Neural Networks ANN's and other machine learning algorithms

Neural networks have been used in systems such as Glove Talk II by S. Fels and G. Hinton to map gestures to speech synthesis.[10] In this case the adaptation happens on the algorithmic level, the domain of variation is described by the network. The result is flexible but constrained to a response by a set of training data or narrow sets of rules, depending on the morphology of the network. George Lewis' Voyager system builds on unpredictability and surprise as fundamental interaction paradigm.[11] The algorithm 'listens' to the performer and reacts in ways that are within its space of potentialities but are interlinked in such a way with the ongoing action that the performer can not anticipate the output. Adaptation happens on symbolic musical levels, the structural link between performer and machine interpreter is expressed as a set of rules rather than an adaptive feedback driven system.

Commonly used algorithms implemented in the fast artificial neural network library, such as multi-layer perceptrons, self organizing feature maps and radial basis functions become now available in standard music synthesis languages.[12] A newer trend in music information retrieval but also in sonic interaction design is the use of dynamic systems which anticipate the evolution of vectors of values.[13]

6. Implementation Example

Codespace is an interactive sound and image installation that was commissioned for the Today's Art festival at the Hague in the Netherlands in 2005 and was subsequently shown at the NIME '06 conference in Paris. A new edition was developed at the media art lab of the art museum in Graz, Austria in early 2009. The interaction happens through video tracking in the gallery space in the old and a multi-touch surface in the new version. In both cases the visitor is exposed to a surround audio scene, created using our own ambisonic tools in MaxMSP[14] and one or more large video projections showing abstract motion graphics. (Figure 2.)

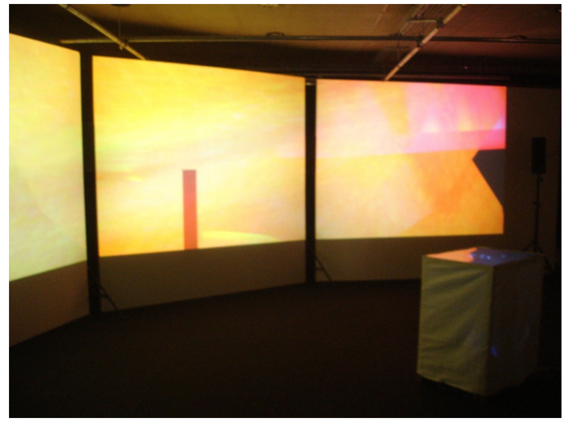

Figure 2. *Codespace* at the Graz art museum, 01/2009.

6.1 Structure

At the heart of the system two conceptual elements exist that form the basis for the algorithmic control. The first concept is that of an affect space, which represents the state of overall system in three orthogonally arranged emotion pairs. An pre-trained ANN receives a wide number of streams of data and indicates the system state as the position within the affect space. The second element is a flocking algorithm, which establishes relationships with the media output on the one hand and is the visible interaction space shown on the multi-touch surface on the other hand. The flock's behavior is directly influenced by the visitor interaction and in the simplest mapping layer immediately applies the position data of the flocking agents to the position of the sound sources in the surrounding audio scene. The output of the flocking algorithm also contains information about neighborhood and attractor relationships which is analyzed and sent back to the ANN in order to controls the affect space and influence higher level control parameters for the flock such as attraction or repulsion forces. This is a prime example of double adaptive loop since the flock is directly dependent on the user's interaction and through its own influence on the affect space adapts to the outside behavior.

The gesture analysis layer extracts salient features from the interaction and also scales and maps certain information directly to interpreter parameters. The salient features extracted are based on higher order derivatives of position, speed, density and distribution within the sensed space. These are applied to sound and image synthesis parameters through a traditional mapping layer which auto scales the values with gliding windows. In parallel the affect space state influences the node connections in the mapping. For a position in the happy, angry, nervous quadrant for example the mapping will apply motion extracted values directly to image motion and sound synthesis onset time controls whereas in a different affect state these streams might be mapped to image color or frequency control nodes. A simple statistical analysis of the weights of certain streams is performed and fed back to the gesture analysis layer.

This adaptive loop tends to reinforce often repeated interaction behavior. (Figure 3.)

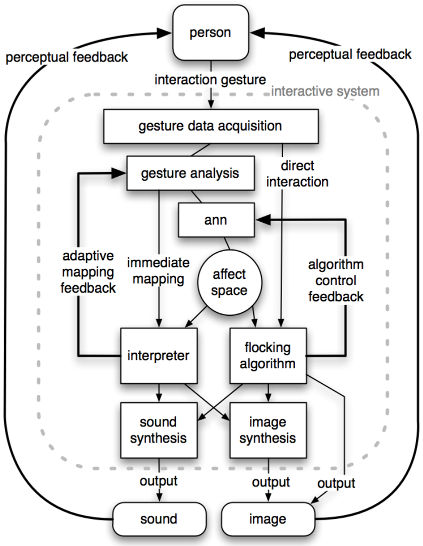

Figure 3. Structure of the interactive system in *Codespace*.

6.2 Audience Engagement

Experience has shown that audience engagement only works if the interaction present can be grasped within a few moments. In the first version of the *Codespace* installation the visitor would not have any direct visual or aural feedback of their action. Without it the recognition of interaction in an installation only happens with the simplest of mappings. As soon as there is one translation layer involved the direct action-perception mechanism fails. With a multi-touch interface the situation changes dramatically. Not only is exploration through touch a natural reflex, but the direct link of the action with a spatially coherent visual feedback becomes immediately obvious. In both versions of the installation adaptive mapping helps making the results of an action become more apparent. The focus on listening and tentative exploratory behavior is stronger on the system which doesn't provide direct visual feedback, the overall audience engagement however is stronger in the clearly perceivable interaction model.

7. Conclusion and Outlook

Mapping is normally designed as the special solution to a specific set of constraints. Once the mapping is established and tested it tends to become almost static. In the case of digital instrument design the persistence of mapping and repeatability are an imperative for success. In the context of novice-level public installation however the mappings that adapt and learn about audience behavior can help to reinforce audience engagement in particular during the initial moments of an encounter. The challenge in this methodology is to find the right paths for feedback, be it inside the interactive system or in a more ecological scope by providing clues to help the visitors perceive their role in the system. More work is needed to formulate a comprehensive methodology for this type of adaptive mappings. The available machine learning tools need to be examined and classified to facilitate their use in interactive realtime systems. More use cases need to be studied by creating new implementations of this principle. Finally the engagement and psychological impact of adaptive systems should be examined more closely in order to find those behaviors, responses and mappings that are valid for a large number of people, yet effective and pleasant to experience.

8. Acknowledgments

This paper has been written thanks to a short term scientific mission to the Center for Intelligent Machines at McGill University, Montreal for the European COST Action IC0601 Sonic Interaction Design. Special thanks go to Yon Visell for gracefully hosting me and to Nils Peters for many fruitful discussions.

References

[1] Wittgenstein, L.; *Culture and Value*, G.H. Von Wright, ed. P. Winch, trans. Chicago: University of Chicago Press, 1980

[2] Carman, T.; *Merleau Ponty*, Routledge, 270 Madison Avenue, New York 10016, 2008, ISBN 10: 0-415-33980-4

[3] Leman, M.; *Embodied Music Cognition and Mediation Technology*, p. 164-165, MIT Press, 2008,

[4] Jakovich, J., Beilharz, K.; ParticleTecture: *Interactive Granular Soundspaces for Architectural Design*, In: Proceedings of the 2007 Conference on New Interfaces for Musical Expression; New York, NY

[5] Wessel, D., Wright, M.; *Problems and Prospects for Intimate Musical Control of Computers* In: Proceedings of the 2001 Conference on New Interfaces for Musical Expression, Seattle, WA

[6] Wessel, D. *An Enactive Approach to Computer Music Performance,* In: Le Feedback dans la Creation Musical, Lyon, France, p.93-98, (2006)

[7] D. Arfib et al. *Organised Sound* **7**(2): 127–144, 2002 Cambridge University Press. Printed in the United Kingdom.

[8] Van Nort, D, Wanderley, M.; *Control Strategies for Navigation of Complex Sonic Spaces*, In: Proceedings of the 2007 conference on New Interfaces for Musical Expression; New York, NY

[9] Wessel, D.; *Instruments that learn, refined controllers, and source model loudspeakers.* Computer Music Journal 15, *4 (1991), 82-86.*

[10] Fels, S. S., Hinton, G. E. *Glove-talk: A neural network interface between a data-glove and a speech synthesizer.* IEEE Transactions on Neural Networks 4: 2–8. 1993.

[11] Lewis, G. E.; *Too Many Notes: Computers, Complexity and Culture in Voyager* In: Leonardo Music Journal 10 (2000)

[12] Fast Artificial Neural Network Library (FANN), http://leenissen.dk/fann/ URI verified January 2009

[13] Visell, Y. *Spontaneous organization, pattern models, and music.* In: Organised Sound 9, 2 (2004), 151-165.

[14] Schacher, J., Kocher, P. *Ambisonics Spatialization Tools for Max/MSP*, In: Proceedings of the 2006 International Computer Music Conference; New Orleans, LA