

Interactive Control of Musical Structures by Hand Gestures

PAUL MODLER

Staatliches Institut für Musikforschung PK
Tiergartenstraße 1, D-10785 Berlin, Germany
pmodler@compuserve.com

1 Abstract

This paper summarizes our ongoing work on the mapping of hand gestures to musical parameters in an interactive music performance and virtual reality environment.

We use data obtained from a sensor glove, a high end input device for digitizing hand and finger motions into multi-parametric data. These data are processed in order to extract meaningful data to control musical structures.

For that we use a neural network architecture to achieve meaningful control parameters.

We focus on the mapping of gestural variations onto equivalent musical parameters, motivated by a creative situation like a performance. We set up a dictionary of symbolic and parametric subgestures. Different hand gestures of this dictionary and characteristic variations will be evaluated with respect to their applicability to intuitive and musical control of musical structures.

The system is complemented with a 3D VRML environment, i.e. an animated hand model and behaving representations of musical structures. This 3D representation combines with the gesture processing module and the sound generation to powerful so called Behaving Virtual Musical Objects.

2 Components of the System

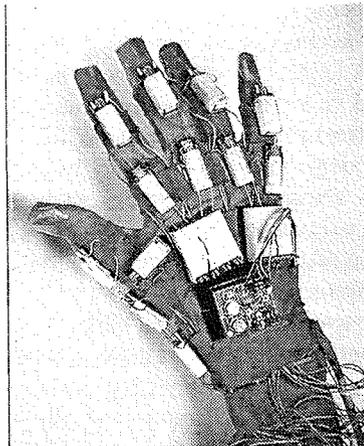
The interactive system comprises the following components which are described below in greater detail.

- a dedicated sensor glove which tracks hand and finger motions

- a design and control environment for the data glove including preprocessing features (written in JAVA)
- a data processing section based on neural networks for gesture recognition and postprocessing
- a real-time sound synthesis module for advanced synthesis algorithms
- a virtual reality framework (VRML) for the interaction of the performer with virtual objects which is included in the JAVA environment

3 Digitization of Hand Movements

The sensor glove developed by Frank Hofmann is used as an input device (Picture 1).



Picture 1 Sensor Glove Version 3 (by Frank Hofman)

By tracking 24 finger angles and 3 hand acceleration values, gestures of the hand can be processed by the connected system. As a first approach, direct mapping of single sensor values to sound parameters was used. Although good results concerning the possibilities of controlling parameters of the sound algorithm (Frequency Modulation, granular Synthesis, analog Synthesis) have been obtained, disadvantages of such a direct connection occurred as well. E.g., intuitive control of multiple parameters simultaneously turned out to be hard to realize. The data from the Sensor Glove are fed into a postprocessing unit which provides feature extraction and gesture recognition abilities, as well as agent features for intelligent control of the subsequent sound synthesis module.

4 Separation of Gestures

We assume that a gesture consists of subgestures of symbolic and parametric nature (cf. Modler & Zannos 1997). The symbolic type does not have to be time-invariant. It can as well be a time-varying gesture to which the user denoted symbolic contents. The parametric type should always be time-variant for the control of sound parameters. With this subgesture architecture gesture-sequences can be recognized using only a part of the hand as a significant symbolic sign, while other parts of the hand movement are used as a parametric sign.

Subgestures allow for both, the application of smaller neural nets as well as the possibility of using trained nets (subgestures) in various compound gestures.

We aim at setting up a set of gestures, suited for the gestural control of musical parameters. The gestures are subdivided into symbolic and parametric subgestures as described above. We show how a dedicated neural network architecture extracts time varying gestures (symbolic gestures). Besides, secondary features such as trajectories of certain sensors or overall hand velocity will be used to extract parametric values (parametric gestures).

Special attention is given to the way a certain gesture can be emphasized or altered with significance for both emotions and music. We are investigating whether the concept of symbolic and parametric gestures can adequately describe the situation of an emotionally expressive performance.

The set of gestures will be evaluated regarding their potential of providing meaningful symbolic and parametric subgestures, as well as how these subgestures can deal with gestural variations.

4.1 Dictionary of Symbolic Subgestures

A set of 15 to 25 gestures was selected and used as a prototype dictionary. For a preliminary classification the following categories were used to organize the gesture dictionary.

- A) gestures with short (not relevant) start and end transition phases and on static state (pose) (e.g. finger signs for numbers)
- B) gestures with repetitive motions (e.g. flat hand moving up and down [slower])
- C) simple (most fingers behave similar) gestures with relevant start and end state and one transition phase (e.g. opening hand from fist)
- D) complex (most fingers behave differently) gestures with relevant or not relevant start and end states and transition phase
- E) compound gestures with various states and continuous transitions.

The dictionary is based mainly on gestures of categories B) C) and A).

Since category A) contains poses, those types of instances have been selected as part of the dictionary, which the performer can use as very clear signs.

Few examples of category D) have been chosen, because of their more complex character.

4.2 Dictionary of Parametric Subgestures

Besides the symbolic gestures a set of parametric gestures has been selected in order to set up a dictionary for classification the subgestures according to the following categories:

- a) alteration of the absolute position of the hand: translation
- b) alteration of the absolute position of the hand: rotation
- c) alteration of velocity (energy)
- d) alteration of the cycle duration

In the dictionary of parametric subgestures we included instances of categories a), b) and c). Additional work has to be done regarding the extraction of repetitive cycle time and detection of resulting timing variations.

For the prototype implementation, an agent-type module which supervises the combination of symbolic and parametric subgestures has been included. This coordinating device recognizes the influence the extracted subgestures have on the output of the symbolic gestures.

5 Pattern Recognition of Gestures by Neural Networks

5.1 Neural Network Architecture

Based on a prototype implementation for the recognition of continuous performed time-variant gestures (cf. Modler & Zannos 1997) we extended the proposed architecture to deal with the demands of the selected dictionary. Additional input layers have been added for the recognition of subgroups of the gesture dictionary. The layers of the subgroups are connected by separate hidden layers.

5.2 Training Procedure

The Network is trained with a set of recorded instances of the gestures of the symbolic subgesture dictionary.

Both the 2D representation of the sensor data as well as the 3D-representation (section 6.2) offered a good feedback about recorded instances.

Each gesture class was recorded two times in 4 different velocities.

The training of the Neural Net was conducted offline. The resulting net parameters were transferred to the Sensor Glove processing section and integrated in the C/JAVA environment.

5.3 Recognition of Subgestures by Neural Networks

For evaluation, time-varying continuous connected phrases of instances of the symbolic subgesture dictionary were presented to the trained net. This was realized online, i.e. the data were passed directly from the glove to the network.

For the selected part of the gesture dictionary the proposed net architecture produced good results, i.e. a recognition rate of about 90 %.

5.4 Extraction of Parametric Subgestures, and Combination with Symbolic Subgestures

The parametric subgestures as proposed in section 5.2 were achieved by online processes. Further investigations will show whether neural networks can also provide the desired parameters.

The combination of both of them produced good results in both recognition of a subgesture and altering the overall gesture by altering the parametric subgesture (e.g. flat hand, fingers moving up and down [slower] combined with translation movements of the whole hand).

5.5 Results

The proposed combination of gesture recognition of symbolic subgestures with parametric ones brought up good results. In other words, they promise to become a powerful tool to extend the possibilities and variability of a performance conducted with the Sensor Glove.

The concept of symbolic and parametric subgestures as well as the proposed categories offer the performer a guideline to fix a parameter mapping with connected sound synthesis and visualization modules. The definition of Virtual Musical Objects (see below) is less cumbersome using this categorization.

The extension of the neural network for the processing of a larger number of features provided seems to be manageable, but an extension to a multiprocessing parallel architecture has to be considered, too.

6 3D Representation: Integration of Virtual Reality Worlds

6.1 Animation of a Hand Model by the Sensor Glove

As a feasibility study, we have created a visual representation of a hand and Virtual Musical Objects in VRML language (cf. Picture 2).

The VRML language is a standardized toolkit which provides a potential for creating three-dimensional environments: virtual worlds. VRML offers the advantage of a platform and browser independent application. Since VRML is so widely accepted, its disadvantage of reduced speed is acceptable.

The hand model is animated with the input from the Sensor Glove. This is realized by a JAVA-VRML interface.

This prototype world can be viewed with a VRML browser that has been integrated into the design and control environment and runs on a combination of JAVA and C.

The VRML - JAVA interface also offers the possibility of dynamically creating or altering existing VRML worlds, in other words, user-provided interaction models such as the ani-

mated hand model can then be introduced into unknown worlds (e.g. scenarios downloaded from somewhere else).

Complex worlds can be generated with special tools like MAX3D or VREALM, which then can be animated, investigated, altered, and viewed with the VRML browser.

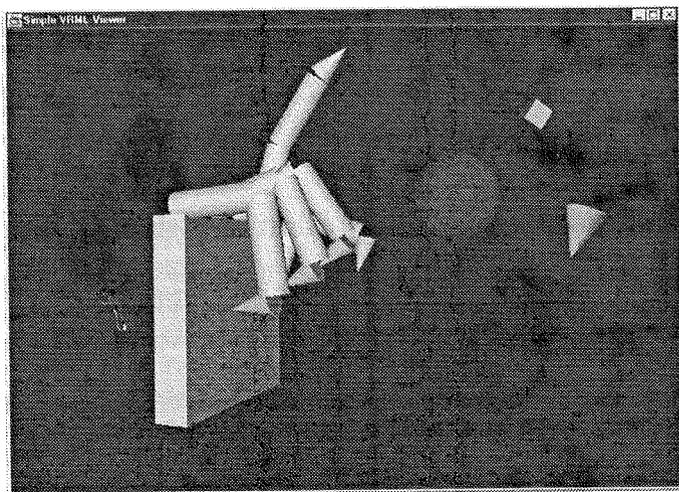
6.2 3D Representation of Behaving Virtual Musical Objects (BVMO)

In addition to the hand animation, we developed a framework for the creation of VMOs.

These objects – in connection with the Sensor Glove - constitute the gesture processing section and the sound synthesis module. An extended form (EVMI) of a Virtual Musical Instrument (VMI) has been proposed by Alex Mulder (Mulder 1994).

The VMOs are variable in color, size, form, and position in the surrounding world. Additional features can be defined and controlled, e.g. time-varying alterations of a certain aspect such as color or motion trajectory. This can be regarded as a behaving VMO (BVMO) or a resident.

The graphical representations (e.g. the hand model) are realized in VRML. For data passing the JAVA, the VRML interface can be used. Good results have been achieved for animating the hand model and altering BVMOs by user input from the Sensor Glove.



Picture 2 VRML World with Animated Hand Model and Virtual Musical Objects

7 Musical Application of the System

The synthesis module is designed in order to provide the ability of controlling different sound algorithms in real-time by continuous control parameters. The algorithms are realized in SuperCollider, a special sound synthesis language offering real-time synthesis possibili-

ties as well as LISP-style list processing and a limited SMALLTALK like class/object structure.

Different synthesis techniques (Frequency Modulation, Physical Modeling and Granular Synthesis) are compared and described with respect to their controllability through the Sensor Glove and the proposed processing system.

8 Parallel Processing Aspects

Currently the system is based on a 200MHz MAC compatible 604PPC covering the data acquisition and processing from the Sensor Glove. It also realizes the gesture preprocessing recognition and postprocessing. The VRML 3D-environment is implemented on a 200 MHz PC which communicates with the SensorGlove machine by MIDI. The sound generation is realized on a 200 MHz MAC compatible 604PPC which is controlled by the Sensor Glove machine as well by MIDI. The training of a the neural network was executed on a Silicon Graphics Indigo2.

The MIDI connections were chosen because of their variability in combining the possible external tools as well as because of their easy handling. Due to the high data rates produced by the Sensor Glove, a more powerful way of transmission has to be considered. Ethernet TCP/IP or Firewire connections are possible solutions, since the C/JAVA environment can be connected with standard software tools over these connections.

For growing neural networks and additional processing of the control data, e.g. detection of beats or cyclic structures etc., a wider distribution of the computing tasks has to be considered.

Besides, the sound synthesis engine profits from a parallel architecture. Promising tests with a dedicated communication protocol for parallel distributed audio processes MIDAS have been conducted.

9 Conclusions

Based on our experiments we come to the following results and conclusions.

The subgestural concept for deriving symbolic and parametric gestures is a good approach for integrating gesture recognition into a performance situation. The neural network pattern recognition is combined with flexible and intuitive possibilities of altering material.

Specific control changes as well as intuitive overall changes can be achieved.

The proposed categories of subgestures offer the performer a comprehensive way to design the behavior of the sound generation. This provides a powerful alternative to the one to one mapping of single parameters.

The proposed dictionary of gestures provides the performer with an intuitive way for musical expressiveness and meaningful variations.

Behaving Virtual Musical Objects integrated in a virtual 3D world offer a promising way for novel visual representation of abstract sound generation algorithms. This includes specific control of a sound scene, but also facilitates memorizing and recalling of a sound scene and inspires the user to new combinations.

The combination of the proposed gesture mapping with the BVMOs constitutes a powerful environment not only for interactive performances, but also for the design of sounds and sound scenes.

10 References

- Hofmann, F.G, Hommel, G.: Analyzing Human Gestural Motions Using Acceleration Sensors., Proc. of the Gesture Workshop '96 (GW'96), University of York, UK, im Druck
- Hommel, G., Hofmann, F., Henz, J.: The TU Berlin High-Precision Sensor Glove. Proc. of the Fourth International, Scientific Conference , University of Milan, Milan/Italy, 1994
- Modler, Paul, Interactive Computer-Music Systems and Concepts of Gestalt, Proceedings of the Joint International Conference of Musicology, Brügge 1996.
- Modler, Paul, Zannos, Ioannis, Emotional Aspects of Gesture Recognition by Neural Networks, using dedicated Input Devices, in Antonio Camurri (ed.) Proc. of KANSEI The Technology of Emotion, AIMI International Workshop, Universita Genova, Genova 1997
- Mulder, Axel, Virtual Musical Instruments: Accessing the Sound Synthesis Universe as a Performer, 1994
<http://fassfu.ca/cs/people/ResearchStaff/amulder/personal/vmi/BSCM1.rev.html>
- Schmidt-Atzert, Lothar, Lehrbuch der Emotionspsychologie, Kohlhammer, Stuttgart Berlin Köln, 1996
- SNNS, Stuttgarter Neural Network Simulator, User Manual 4.1, Stuttgart, University of Stuttgart, 1995.
- Waibel, A., T. Hanazawa, G. Hinton, K. Shikano, and K.J. Lang, Phoneme recognition using time-delay neural networks, IEEE Transactions On Acoustics, Speech, and Signal Processing, Vol 37, No. 3, pp. 328-339, March 1989.
- Wassermann P.D., Neural Computing, Theory and Practice, Van Nostrand Reinhold, 1993.
- Wundt, W., Grundzüge der Physiologischen Psychologie, Verlag von Wilhelm Engelmann Leipzig, 1903
- Zell, Andres, Simulation Neuronaler Netze, Bonn, Paris: Addison Wesley, 1994.

Sound Sculpting: Performing with Virtual Musical Instruments

AXEL G. E. MULDER

*School of Kinesiology, Simon Fraser University, Burnaby, BC, Canada V5A 1S6.
Email amulder@sfu.ca*

S. SIDNEY FELS

Dept. of Electrical and Computer Eng., University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Abstract

We have prototyped an environment for designing and performing with virtual musical instruments. It enables a sound artist or musician to design a musical instrument according to his or her wishes with respect to gestural and musical constraints. Sounds can be created, edited or performed by changing parameters like position, orientation and shape of a virtual input device, that can only be perceived through its visual and acoustic representations. We implemented the environment by extending a realtime, visual programming language called Max/FTS, with software objects to interface datagloves and 6DOF sensors and to compute human movement and virtual object features. Our pilot studies involved a virtual input device with a behaviours of a rubber balloon and a rubber sheet for the control of sound spatialization and timbre parameters. It was found that the more physical analogies are used for a mapping the faster the mapping can be learned. We also found that we need to address more manipulation pragmatics in the mapping and more carefully identify movement features to map to virtual object parameters. While both hands can be used for manipulation, left-hand-only sound sculpting may be a useful replacement for and extension of the standard keyboard modulation wheel.

1 Introduction

We report in this paper on work in progress to develop gestural interfaces that allow for simultaneous multidimensional control. An example of simultaneous multidimensional control can be found in music composition and sound design, which task involves the manipulation of many inter-dependent parameters simultaneously. For any sound designer or composer and certainly a performer the control of these parameters