1 Abstract

In this paper we present a system that maps hand gestures into musical parameters in an interactive computer music performance and virtual reality environment.

In the first part of the paper we comment on our view of emotions. Thereafter, the technical background will be introduced.

We show that in a performing situation the expression of emotion is strongly related to intuitive and interactive aesthetic variations.

Our focus on the mapping of these gestural variations into relevant musical parameters leads to the concept of symbolic and parametric subgestures. This allows the generation of emotional and aesthetic variations.

We use data obtained from a sensor glove, a high end input device for digitizing hand and finger motions into multi-parametric data.

The processing of this data is done by a neural network recognizing symbolic subgestures combined with the extraction of parametric values.

We present a dictionary of symbolic and parametric subgestures which is categorised with respect to complexity.

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 so called "Behaving Virtual Musical Objects".

2 Emotion and Interactive Computer Music Systems

The expression of emotion in a musical performance can roughly be divided into static and dynamic parts. We assume that altering given material of both parts is a fundamental way of expressing emotions. Emotion is a term which inherently refers to subjective experience (Metzinger, 1995). Therefore, it is very difficult to find a commonly accepted definition of "emotion".

For the purposes of our work we assume that the term emotion is related to the aspects sketched below.

As commonly assumed, emotions refer to something which can be experienced, or which can be expressed (Schmidt-Atzert 1996). According to Wundt (1903) there are 3 basic bipolar components of emotions from which all other emotions derive (i.e. enjoyment / dislike, excitement / calming, tension / soothing).

Emotions are assumed to result from basic experiences like sadness, loneliness, joy, love, hate, etc. For the purposes of creativity, for example in a musical performance, it is important to understand how these emotions can be evoked or how they can be diminished in a musical system.

The concept of emotion is assumed to be special because of several reasons. One of them is the changing behavior of emotion: The same thing can be experienced as sad or joyful, depending on the psychological condition of the recipient. Accordingly, a given structure can be performed with different emotional reactions and perspectives.

Emotional aspects of musical systems can be classified according to their temporal variation:

- emotional symbols: 'static' colors, smile, low sound, high sound
- emotional movements: 'time-variant' accelerando-ritardando, de-/ crescendo

The symbolic level is closely related to the macro-structural components of music, that is, composition. The movement level is closely related to the micro-structural components of music, that is, interpretation.

In musical interpretation, movements are commonly used to model emotional aspects. Movements in this context means the changing of parameters related to the music. Examples are dynamic or timing variations a performer uses to add expressiveness.

In computer music, symbolic aspects seem to be easy to handle. Sound and environment settings can be prepared and recalled at performance time.

Compared to the symbolic level, movement variations are much harder to realize in a computer music environment. Simply playing a sequence of events doesn't seem to fascinate an audience. Instead, there is a demand to compose movements, which here represents emotional expression, or model them by an algorithm, which arises from non real-time and non-interactive systems.

With the possibility of changing sound or event parameters just in time, interactive computer music systems offer new potentials of realizing creative movements. These instruments can handle the demand of expressing emotions by...
rapidly and intuitively changing the parameters directly controlled by the performer.

3 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.

To give an example:

A straightened index finger indicates mouse down (symbolic subgesture) and moving the whole hand determines the alteration of the position of the mouse (parametric subgesture).

Or a straightened index finger selects a sound object (symbolic subgesture) and determines the volume and pitch of the object through the hand position (parametric subgesture).

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 establishing 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.

3.1 Categorisation of Symbolic Subgestures

A set of 15 to 25 gestures was selected and used as a prototype dictionary. For a 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.

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

3.2 Categorization of Parametric Subgestures

Besides the symbolic gestures a set of parametric gestures has been selected for building a dictionary for classification in which following categories for subgestures are available:

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.

4 Components of the System

The interactive computer music system we assume comprises the following components (Picture 1) 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. It is included in the JAVA environment
5 Digitization of Hand Movements

The sensor glove developed by Frank Hofmann at the Technical University of Berlin is used as an input device (Picture 2). By tracking 24 finger ankles 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.

6 Pattern Recognition of Gestures by Neural Networks

6.1 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.

6.2 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.

6.3 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 offered good results, i.e. a recognition rate of about 90%.
6.4 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 symbolic parameters. The combination of both parameters produced good results in both recognition of a subgesture as well as altering the overall gesture by changing the parametric subgesture (e.g. flat hand, fingers moving up and down [slower] combined with translation movements of the whole hand).

6.5 Results

The proposed combination of gesture recognition of symbolic subgestures with parametric ones brought up good results. In other words, they promise to promote and 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.

7 Visual Representation of Virtual Musical Instruments

7.1 Animated Hand Model in a Virtual World

As a feasibility study, we have created a visual representation of a hand and Virtual Musical Objects in VRML language (cf. Picture 3).

The VRML language is a standardized toolkit which provides possibilities 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 animated hand model can then be introduced into unknown worlds (e.g. downloaded from somewhere else).

Picture 3 VRML World with Animated Hand Model and Virtual Musical Objects

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

7.2 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 together 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.

8 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.

9 References


