Using Music to Interact with a Virtual Character

Robyn Taylor
Department of Computing Science
University of Alberta
T6G 2E8, Edmonton, Alberta, Canada
robyn@cs.ualberta.ca

Daniel Torres
Department of Computing Science
University of Alberta
T6G 2E8, Edmonton, Alberta, Canada
dtorres@cs.ualberta.ca

Pierre Boulanger
Department of Computing Science
University of Alberta
T6G 2E8, Edmonton, Alberta, Canada
pierreb@cs.ualberta.ca

ABSTRACT

We present a real-time system which allows musicians to interact with synthetic virtual characters as they perform. Using Max/MSP to parameterize keyboard and vocal input, meaningful features (pitch, amplitude, chord information, and vocal timbre) are extracted from live performance in real-time. These extracted musical features are then mapped to character behaviour in such a way that the musician's performance elicits a response from the virtual character. The system uses the ANIMUS framework to generate believable character expressions. Experimental results are presented for simple characters.

Keywords

Music, synthetic characters, advanced man-machine interfaces, virtual reality, behavioural systems, interaction techniques, visualization, immersive entertainment, artistic installations

1. INTRODUCTION

We have created a system which enables a musician to interact with life-sized virtual characters within an immersive virtual environment. These virtual characters "listen" to the musician's performance and modify their behaviour accordingly. The musician can perform specific musical gestures that have been predefined to trigger specific behavioural responses from the virtual characters or he or she may simply perform spontaneously and watch the characters' responses as they unfold.

This system may be used for many purposes, ranging from the recreational (users may simply be entertained by the ability to control virtual characters through a musical interface) to the educational (character responses could be used to encourage users to develop particular skill sets) to the artistic (scores and animations could be choreographed to produce a cohesive virtualized performance).

In this paper, we will describe the motivation behind the creation of such a system. We will then describe the ANIMUS [9] framework which enables virtual characters to respond to user input by using an accumulation of basic behaviours. We will address the real-time extraction of musical features from live musical performance and describe how these extracted features can then be used to control character behaviour. In addition, we will discuss the user interface and design decisions which must be considered when creating this type of application. Finally, we will describe the current status of the system as well as future plans for expanding it to include more complex characters.

2. MOTIVATION

Modern virtual reality environments are capable of producing compelling visualizations which draw the user into a virtual world bounded only by its designer's imagination. These high-end visualization systems are capable of displaying stereoscopic images on wall-sized display screens in such a way that the observer feels truly immersed in virtual surroundings that appear life-sized and three-dimensional. The primary objective of this project is that these visualization tools be used to enhance the experience of live musical performance by allowing both the performer, and his or her audience, to enter a virtual world populated by believable characters that react in response to real-time musical input.

There exist several examples of previous research into the concept of immersive music visualization. Notably, Jack Ox's "Color Organ" [6] uses immersive virtual reality systems (in particular, the Cave Automatic Virtual Environment) to create visual abstractions of music, allowing the user to explore the complexity of a composition in three dimensions. Also, the "Singing Tree" [5], designed by Oliver *et al.* at the MIT Media Laboratory, uses digital displays as well as sculptured set pieces to enclose participants in a virtual environment which responds to their vocal utterances. Masataka Goto and Yoichi Muraoka's Virtual Dancer, "Cindy" [4] provides visual enhancement to musical "jam" sessions by dancing in rhythm to musicians' playing.

Additionally, immersive music visualization techniques have been used successfully in live performance. Golan Levin and Zach Lieberman's audiovisual performance piece, "Messa di Voce" [3], utilizes large scale projection screens along with motion tracking technologies to allow live vocalists to integrate their physical bodies into a projected virtual environment, visualizing their vocalizations with relation to their physical locations on the performance stage.

We strive to similarly reduce the barrier between the musician and the virtualized environment, enabling us to cre-

Figure 1: The User in the Virtual Environment

ate a new performance space and medium. This system allows a musician to perform a piece of music in a natural fashion: playing a digital piano and singing into a microphone. The visualization system processes the musical input in real-time and stimulates visibly responsive behaviour in synthetic characters existing within the virtual space. The virtual environment is displayed on a large stereoscopic screen. It allows the participant to experience the virtual space at an appropriate scale, as if she or he was part of this space. Any observers will also perceive the performer and the virtual characters to be of similar size, increasing the realism of the performer/virtual character interaction (see Figure 1).

3. THE ANIMUS FRAMEWORK

The ANIMUS framework was developed in a previous project by Torres and Boulanger [9], to facilitate the creation of "believable artificial characters". They [11] describe the notion of a believable artificial character as follows:

"A believable artificial character shows its own virtual thoughts and emotions, plans and beliefs; it can be tricked; it is not necessarily intelligent or efficient, nor always makes the best decisions, but acts within the bounds of its role and personality. Its animation may not be realistic, but succeeds at expressing its internal state and makes it interesting to an audience."

Our goal is to create instances of believable artificial characters which respond to live music according to their own individual personalities.

In the ANIMUS framework, responsive behaviour is constructed by three separate layers:

- **Perception Layer:** The characters must perceive data in their surroundings
- **Cognition Layer:** The characters must assess the data they receive and determine how they should respond to it, based on their internal cognitive processes
- **Expression Layer:** The characters must exhibit visible behaviours to indicate their cognitive state

We need to create a character, based on these three layers, which responds to user input received in the form of a real-time musical performance. In order to do this, we are adding a fourth layer to the ANIMUS architecture:

- **Musical Perception Filter Layer:** A bank of Max/MSP[1] objects must extract musical features from the real-time performance and send them to the perception layer to be processed

3.1 Perception Layer

The ANIMUS Project uses a method of information organization known as a "blackboard system". All data that is perceivable by the ANIMUS characters is entered on the blackboard, and this blackboard is monitored and modified by each character. Our system must add information about the live musical performance to this blackboard in order for the ANIMUS characters to perceive the music as an input stimulus. To do this, our perception layer interfaces with the specialized Musical Perception Filter Layer which identifies specific features within the live performance. See Section 4 for further details on the musical feature extraction process.

3.2 Cognition Layer

To create ANIMUS characters which respond in a believable fashion to musical input, we need to create a cognition layer that assigns relevant meaning to the musical features received by the perception layer. This is the most complex step of the responsive behavioural process since this is where the character's illusion of "personality" is constructed. In order to create an ANIMUS character that exhibits a distinct personality in addition to appearing to be aware of the music in its surroundings, we must create a complex cognition layer which maps the incoming musical data to modifiers affecting the virtual character's unique internal state which controls its expression layer. For discussion of how musical input is used to influence character behaviour, see Section 5.

3.3 Expression Layer

An ANIMUS character expresses its internal state by executing an appropriate sequence of animations produced by the expression layer, using key frame based animation. These animations are created at run-time by using an interpolation scheme that takes predefined key frames and generates intermediate transitions.

4. MUSICAL PERCEPTION FILTER LAYER

A live musical performance contains a wealth of real-time information which we as humans perceive as meaningful. In order for our synthetic character to assign any meaning to a stream of auditory information, it must receive this stream in a simplified format. We have chosen to create an additional layer to interface with the ANIMUS framework, the Musical Perception Filter Layer. In this layer we parse the stream of incoming auditory data in order to identify important musical features to send as input to the ANIMUS system's perception layer.

To extract features from live musical performance we use Max/MSP to monitor input from a microphone and MIDI controller.

We determine the user's vocal pitch and amplitude using Puckette, Apel and Zicarelli's `fiddle~` object [7]. Additionally we make use of the `fiddle~` object's ability to describe the harmonic spectra of the user's voice. Our system examines the raw peak data output by `fiddle~` and generates a numerical description of the vocal tone based on the weighting of tone amplitude at the fundamental frequency versus multiples of the fundamental frequency.

Figure 2: The Alebrije character

We are also interested in identifying any specific chords present in the user's keyboard playing. To extract information about any chords the user might play, we created a Max patch to monitor input from the MIDI controller and compare it to a list of known chords in order to report any identified matches.

These extracted musical features are then communicated to the ANIMUS engine via a specially created Max external object called `vrpnserver` which integrates the Virtual-Reality Peripheral Network (VRPN) [8] communication library to standardize the information transfer between the computer in charge of sound analysis and the rendering computers. For further discussion of the how the system is structured, please see Section 6.

The data commnicated by the `vrpnserver` object serves as input to the perception layer of an ANIMUS character. An ANIMUS character may be designed which is aware of pitch information, for example, so any pitch information extracted from the live user's performance will be perceived by this character's perception layer and sent to its cognition layer.

5. MAPPING MUSIC TO BEHAVIOUR

The ANIMUS cognition layer is where a character's internal state is defined. This internal state is then visualized by the animations generated automatically by the expression layer. In order to use ANIMUS as a tool for character-based music visualization, we must create a character with cognition and expression layers sophisticated enough to produce behaviours which the user and observers will recognize as being responsive to the musical input.

ANIMUS characters must monitor perceptual data in their environment and decide how to respond. Each character can be customized to notice particular types of data and ignore others. Some characters may be interested in the pitch and amplitude of any singing voices in their vicinity, while others may only care about major and minor chords played on the keyboard. These decisions are made by the program designer who designs the characters to be attentive to particular entries on the virtual world's "blackboard".

To prototype this concept, a simple ANIMUS character who responds to musical input was developed. This character is a re-incarnation of Torres' lizard-like character, Alebrije [9] (see Figure 2). We have modified Alebrije so that he is aware of vocalized pitches and their directional source within his environment.

When the user sings a pitch that is higher than middle C, a perceptual filter, specialized for this event, writes of its occurence to the ANIMUS engine's blackboard. Alebrije monitors the blackboard for pitch information. When this event is perceived, Alebrije's attention immediately becomes focused on what he believes to be the origin of the sound - another ANIMUS character in the scene. This is expressed by an animation which shows Alebrije turning to look at this other character. If Alebrije hears no singing for an extended period of time, he becomes bored and goes to sleep. He will only wake up again if the user chooses to vocalize another appropriate pitch, upon which time his attention will once more be directed toward the small noise-emitting character.

This type of behavioural control is called a driver system. Similar to the behavioural system made widely familiar by the Sims [2], a driver system monitors the current level of a certain aspect of a character's internal state. In this example, the particular aspect of Alebrije's internal state that we are concerned with is boredom. When Alebrije's "boredom" driver reaches a certain threshold, he succumbs to sleep. His "boredom" level is reduced when he hears the user vocalizing and it is increased when the user is silent.

This is an example of a very simple mapping between user input and character behaviour. It would be much more interesting if Alebrije were to exhibit more complex behaviour. He could turn his head slowly towards a low, breathily sung pitch, or sharply shift his attention to a sound voiced with a more aggressive tone quality. He could even snidely cover his ears at the sound of a piercing high C. These behaviours are controlled by the designer. The designer chooses which animated poses to translate between based on the character's evaluation of the musical input, and also chooses how sharply or smoothly these transitions should take place based on the general mood of the music.

As discussed in detail in the works of Torres [9] and Torres and Boulanger [11] [10], the ANIMUS engine facilitates complex responsive behaviour, encouraging the development of a believable synthetic character. The ability to connect Max/MSP to the ANIMUS engine makes it possible to create mappings between live musical performance and real-time animated character response.

6. USER INTERFACE AND SYSTEM STRUCTURE

This system is designed to leverage the computing power of more than one machine; thereby reducing system load and enabling both musical analysis and visual rendering to run in real-time. All music processing work is done on a Macintosh G5 system running Max/MSP. The features extracted from the live musical input are transmitted across the network to a PC running the ANIMUS engine using standard networking capacities of 100Mb/s. This is facilitated by the creation of a Max external which integrates a flexible library of virtual reality networking tools called the Virtual Reality Peripheral Network (VRPN). This external object which brings VRPN functionality into Max/MSP enables Max/MSP data to be represented in any "front-end" visualization engine which supports a VRPN connection.

In keeping with our desire to allow the user to experience immersive interaction with virtual characters, we use a desktop PC to render the virtual environment onto high-end display screens which provide life-sized stereoscopic images.

This allows both the user and the observers to experience the illusion of the physical interacting with the virtual in a three dimensional life-sized landscape.

It is also important that the user be able to interact with the virtual character using a natural means of expression - in this case, musical expression. Therefore, we have taken care to ensure that the system is designed in such a way that once the program is launched, the user need only interact with the program by providing musical input to the MIDI controller and to the microphone. The microphone and MIDI controller are positioned in such a way that the user can easily view the virtual character's responses as she or he performs live music.

It is our hope that our large-scale stereoscopic display and our natural means of user input help reduce the distance between the physical and the virtual, and allow the user and observers to experience a compelling sense of true interaction between the musician and the virtual character.

7. DISCUSSION AND FUTURE WORK

We have created a functional system that is capable of visualizing a live musical performance by triggering responsive behaviour in synthetic characters. Currently, we have created simple ANIMUS characters that respond to musical input in order to illustrate our system's potential. We will now focus on the creation of more complex, "believable" characters.

It is our goal that characters' responses to music serve to illustrate the music's emotive properties. This requires that not only must our characters visibly respond to the music, but also that these responses must be of sufficient complexity that a user would perceive them as portraying a sense of emotional understanding.

In order to achieve this, we must create an ANIMUS character with a cognitive layer far more complex than the ones we have created so far. Our character must make decisions not only based on the factual musical features extracted by the Musical Perception Filter Layer, but also based on a simulated understanding of the emotional content of the music. To do this, our character's cognitive processes must be developed to include a relationship between the perceived musical data and a believably simulated emotional state. Our character should be able to portray a simulated emotion of "unhappiness" when "sad" music is played, or "fear" when "scary" music is played. We must conduct further research to determine how we will choose to define these emotional contexts, both in terms of how we identify these characteristics in a musical work, and how we express these emotions via character behaviour.

Additionally, in order for our animated characters to achieve "believability" as defined by Torres and Boulanger [10], we must also design a more varied set of poses for use in the expression layer so that our character may perform more complex animations that convey a sense of personality to the observer. We aim to collaborate with a visual artist so as to obtain a greater number of poses from which our expression layer may generate character animation.

As stated previously in this paper, we hope that through the creation of ANIMUS characters which respond believably to musical stimulus, we may use this system to create works of recreational, educational, and artistic merit.

8. ACKNOWLEDGMENTS

The use of the VRPN library was made possible by the NIH National Research Resource in Molecular Graphics and Microscopy at the University of North Carolina at Chapel Hill.

9. REFERENCES

[1] Cycling '74. Max/MSP.

[2] Electronic Arts. The Sims, 2000.

[3] G. Levin and Z. Lieberman. In-situ speech visualization in real-time interactive installation and performance. In *Proceedings of The 3rd International Symposium on Non-Photorealistic Animation and Rendering*, pages 7–14. ACM Press, 2004.

[4] Masataka Goto and Yoichi Muraoka. Interactive Performance of a Music-Controlled CG Dancer. http://staff.aist.go.jp/m.goto/PROJ/ip-j.html.

[5] W. Oliver, J. Yu, and E. Metois. The Singing Tree: design of an interactive musical interface. In *DIS '97: Proceedings of the conference on Designing interactive systems : processes, practices, methods, and techniques*, pages 261–264. ACM Press, 1997.

[6] J. Ox. 2 performances in the 21st Century Virtual Color Organ. In *Proceedings of the fourth conference on Creativity & Cognition*, pages 20–24. ACM Press, 2002.

[7] M. Puckette, T. Apel, and D. Zicarelli. Real-time audio analysis tools for Pd and MSP. In *Proceedings of the International Computer Music Conference*, pages 109–112. International Computer Music Association, 1998.

[8] R. M. Taylor II, T. C. Hudson, A. Seeger, H. Weber, J. Juliano, and A. T. Helser. VRPN: A device-independent, network-transparent VR peripheral system. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 55–61. ACM Press, 2001.

[9] D. Torres. The ANIMUS Project: A framework for the creation of synthetic characters. Master's thesis, University of Alberta, 2003.

[10] D. Torres and P. Boulanger. The ANIMUS Project: a framework for the creation of interactive creatures in immersed environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 91–99. ACM Press, 2003.

[11] D. Torres and P. Boulanger. A perception and selective attention system for synthetic creatures. In *Proceedings of the Third International Symposium On Smart Graphics*, pages 141–150, 2003.