Proceedings of the Stockholm Music Acoustics Conference, August 6-9, 2003 (SMAC 03), Stockholm, Sweden

EPS - AN INTERACTIVE COLLABORATIVE GAME USING NON-VERBAL COMMUNICATION

Marie-Louise Rinman[1], Anders Friberg[2], Ivar Kjellmo[3], Antonio Camurri[4], Damien Cirotteau[5], Sofia Dahl[2], Barbara Mazzarino[4], Bendik Bendikesen[3], Hugh McCarthy[5]

[1] Centre of User Oriented IT-design, KTH, Stockholm, rinman@nada.kth.se

[2]Speech Music and Hearing, KTH, Stockholm, [andersf, sofia.dahl]@speech.kth.se

[3]Octaga / Telenor, Oslo, [ivar.kjellmo, bendik.beniksen]@octaga.com

[4]InfoMus Lab, DIST – University of Genoa, Genoa, [music, bunny]@dist.unige.it

[5] CSC – Center of Computational Sonology, DEI – Dept of Information Engineering, University of Padua, [Cirotteau, maccarthyhugh]@csc.unipd.it

ABSTRACT

The interactive game environment EPS (expressive performance space), presented in this short paper, is a work still in progress. EPS involves participants in an activity using non-verbal emotional expressions. Two teams use expressive gestures in either voice or body movements to compete. Each team has an avatar controlled either by singing into a microphone or by moving in front of a video camera. Participants/players control their avatars by using acoustical or motion cues. The avatar is navigated /moving around in a 3D distributed virtual environment using the Octagon server and player system. The voice input is processed using a musical cue analysis module yielding performance variables such as tempo, sound level and articulation as well as an emotional prediction. Similarly, movements captured from the video camera are analyzed in terms of different movement cues. The target group is children aged 13-16 and the purpose is to elaborate new forms of collaboration.

1. INTRODUCTION

Participants involved in the EPS game use (emotional) non-verbal communication in music and dance to control and interact with an avatar. Expressive Gesture, as applied in EPS, is using recent models of expressive communication in music and dance developed by Camurri and others [3,] [1], [4] and [8] in the on-going EU project MEGA. These models are based upon a large body of research in expressive communication [9], [5].

The aim of the EPS game, and presented in this paper, is to elaborate new forms of collaboration for children as well as to test the applicability and intelligibility of expressive gesture in a new setting. First, could expressive gesture, e.g. music and dance, be used as a basis for (audience) collaboration and participation? Second, could it be used as an input control/means of communication in an interactive game environment? And third, the benefits of using a game format when young people are involved.

2. THE EPS GAME

Main features and aims

A multi-user multi-relationship real-time activity including multisensory interaction and multimodality.

Non-verbal communication:

Both in the interaction with the 3D world projected on a screen and between audience and players. Non-verbal expression in voice and body movements.

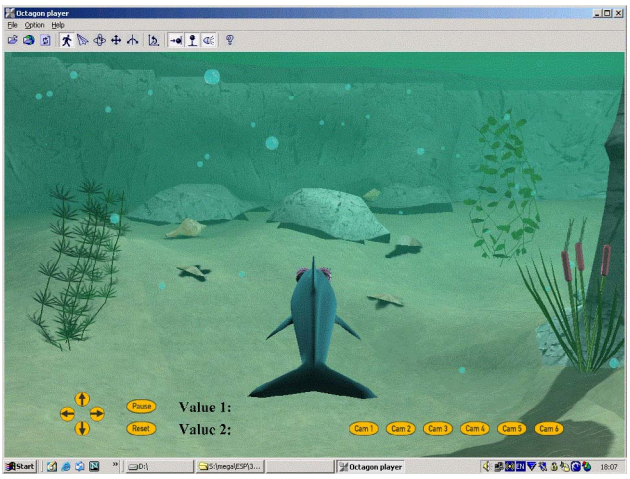

Fig. 1. *Screenshot from the players' point-of-view of one of the teams.*

Collaboration:
Between team members as well as between audience and teams. The audience influences the speed of the avatar and creates sound through their movements.
Mixed-reality features:
Face-to-face meetings between players in the teams as well as between teams and public.
Interaction with a virtual world. Each team has an avatar controlled by one of its players though expressive sounds and movements.
Task:
The main task is to find three caves in the virtual world and solve the riddles/tasks presented in each cave assisted by the audience. To get into the caves teams have to collaborate by solving the problem how to open the gates. The team that finds to the third and last cave first may enter without waiting for the competing team – and wins.

EPS allows participants to interact with each other using verbal communication as well as interacting with a 3D environment using non-verbal communication. The participants are divided into two competing teams (Team 1 and Team 2. There are three players in each team) and members of each team use expressive gestures in voice and in body movements. One team member in each controls an avatar by singing into a microphone or by moving in front of a video camera. Meanwhile the other two team-members may discuss alternative solutions helping the navigator executing his or her task. The avatar changes position and (in a later version of the game) colour, shape and size depending on the acoustical and motion cues conveyed by the participants/players. The first challenge for each team is to figure out how to use the gestures to control the position (and shape) of the avatar in order to play. An example of the controllers viewpoint is shown in Figure 1.

Members of the audience are divided into two groups and invited to take part in the game (Audience group 1 and Audience group 2). Their task is to serve as extra helpers for each team. The audience groups generate a music background by moving and dancing. They may follow the game via the two screens onto which the representations of each team are projected. They also have access to a screen that shows the full overview of the 3D environment. Figure 2 gives an overview of the game setup. The two screens on stage show a third-person view of the virtual world. The audience screen is placed above the stage showing an overview of the complete virtual world.

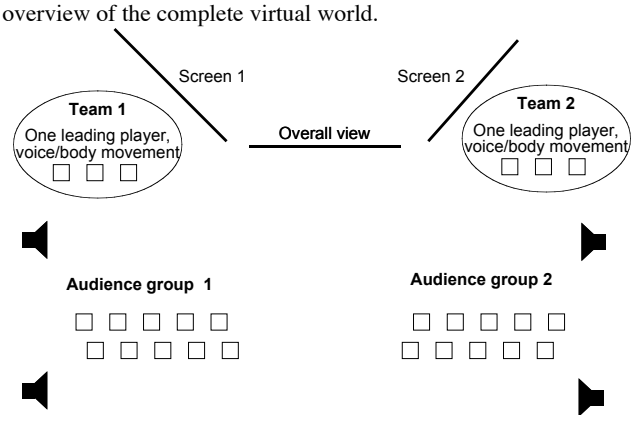

Fig. 2. Overview of game setup.

3. EXPRESSIVENESS IN MUSIC AND DANCE

Music performances can be described by their performance variables such as tone interonsets (IOI), tone sound level, articulation, and intonation. Overall measures of these performance variables over a section, henceforth as cues, have a strong coupling to the performed and perceived emotional expressive content in the performance [7, 8]. For example, happiness is communicated by fast tempo, staccato articulation, high sound level, and fast tone attacks, while sadness is communicated by slow tempo, low sound level, legato articulation, and slow tone attacks. These rather general descriptions have been collected from a large number of investigations and are surprisingly consistent for different music examples, performers or listeners.

Dance performance may be analysed in a similar way using overall motion cues such as quantity of motion, velocity, amount of motion/rest, and fluency [2]. Here too, there is a coupling between these cues and the expressive character of the dance according to Camurri, Lagerlof and Volpe [3].

Recently, real-time computer tools for analysing expressive gestures in music and dance have been developed within the MEGA project (www.megaproject.org). These tools have been developed within the software platforms EyesWeb and pd and are used as a starting point in the non-verbal input modules in the game. The body gesture analysis package in EyesWeb has a large number of different gesture cues and is using imaging algorithms applied to video input. The musical gesture analysis is using signal processing algorithms applied on audio input and outputs both the basic musical cues such as tempo and sound level, but also the emotional intent [2, 6].

4. TECHNICAL IMPLEMENTATION

The virtual 3D environment is run in the Octagon server and player system. This system is a distributed virtual reality system, that uses MPEG4 and VRML file for-mats. Multiple users are represented in the virtual world by their own figure or avatar. Here they are present in the same shared virtual environment and can interact with each other. The distribution goes through a network using the MPEG4 protocol. (More on Octaga and the Octagon system at: www.octaga.com)

There are two types of expressive input: voice or movements. An avatar can be controlled by either voice or movement input using expressive features in each do-main. The voice of the player is captured by a microphone connected to a computer.

The player is supposed to sing or make sounds and different music performance variables are analysed using the musical cue analysis software [6]. The movement of the player is captured by a video camera connected to a computer. The player is supposed to move/dance and different overall motion cues are analysed using a cue analysis patch implemented in the EyesWeb software.

EyesWeb is connected to each Octagon client so the clients can receive expressive parameters. In this case the avatars receive the parameters and change according to them.

5. USING CONVENTIONS IN GAMES AND NARRATIVES TO ENGAGE PARTICIPANTS AND ORGANISE PARTICIPATION

The aim of the activity is to encourage collaboration between young people coming from different cultural and social backgrounds taking specifically language and communication problems into consideration. The reason why a game format is used is that it is familiar to the target group and may thus serve the purpose of establishing a common ground [10].

How could participation in complex settings involving young users be facilitated? Conventions used in games (rules for instance), narratives (a protagonist – an avatar; a story line; organizing the story and participants roles and actions) and drama (three caves, three acts; limitations in time and space; beginning, middle and end) could be used: a) to design participants' actions and responses; b) to create engaging tasks for the target group and allow a certain amount of free space for them to interpret and solve them in a meaningful way; c) to encourage collaboration. The properties of networked computer games are familiar to most 13 – 16 year olds. A large number consists of a mix of teamwork, competition and action (such as *Capture The Flag,* a networked multi-player version of the computer game Quake).

Children having problems with expressing themselves verbally due to language problems may interact with the 3D world in EPS as well as with co-players using other ways of expression such as singing, dancing, and/or rhythmic sounds (clapping their hands etc.). They may solve riddles and overcome obstacles individually as well as together with other players. The use of non-verbal communication in expressive gesture (movements and sound) will hopefully encourage participants to use other forms of communication such as the written and spoken language.

Finally the use of a theatre and game metaphors influences the designer's approach to the user. In an interactive game environment the user becomes a member of an audience, a participant or a player. They are made participants and become co-writers, co-directors and co-actors with a possibility to influence the outcome of the game [10]. An ideal situation would be to involve young participants at an early stage in the design process on order to find out what parts of the game/narrative they find engaging and attractive. The task and challenge for the designers then would be how to put all the bits and pieces together. As a suggestion a game setting may be constructed as a conventional play or story having a beginning, middle and end, containing obstacles, riddles and tasks to be solved, all put together into a meaningful whole.

6. ACKNOWLEDGEMENTS

This work was partially supported by the EU project MEGA IST-20410.

7. REFERENCES

[1] Camurri A (2002). *Interactive Systems Design: a KANSEI-based Approach*, in NIME-02 Intl. Conference on New Interfaces for Musical Expression, Dublin, Ireland

[2] Camurri A, Coletta P, Mazzarino B, Trocca R, &Volpe G (2002). *Improving the man-machine interface through the analysis of expressiveness in human movement*. In Proc. Intl. Conf. IEEE ROMAN-2002, Sept.2002, Berlin. IEEE CS Press.

[3] Camurri A, Lagerlof I & Volpe G (in press). *Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques*. International Journal of Human Computer Studies.

[4] Canazza S (in press). *An abstract control space for communication of sensory expressive intentions in music performance*. Computer Music Journal.

[5] Friberg A & Battel G U (2002). Structural Communication. In (R. Parncutt & G. E. McPherson, Eds.) *The Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*. New York: Oxford University Press, 199-218

[6] Friberg A, Schoonderwaldt E, Juslin, PN & Bresin, R (2002). *Automatic Real-Time Extraction of Musical Expression*. In Proceedings of the International Computer Music Conference 2002, San Francisco: International Computer Music Association, 365-367.

[7] Juslin P N (2001). *Communication of emotion in music performance: A review and a theoretical framework*. In P. N. Juslin & J. A. Sloboda (Eds.), Music and emotion: Theory and research (pp. 309-337). New York: Oxford University Press.

[8] Juslin, P N, Friberg A & Bresin, R (2002). *Toward a computational model of expression in performance: The GERM model.* Musicae Scientiae specal issue 2001-2002, 63-122.

[9] Juslin P N & Sloboda, J A, eds. (2002). *Music and emotion: Theory and research.* New York: Oxford University Press.

[10] Rinman M-L (2002). *Forms of Interaction in Mixed Reality Performance – A study of the artistic event Desert Rain.* Licentiate thesis, Royal Institute of Technology (KTH), Stockholm.