Exploring Emotions and Multimodality in Digitally Augmented Puppeteering

Lassi A. Liikkanen, Giulio Jacucci,
Eero Huvio, Toni Laitinen
Helsinki Institute for Information Technology HIIT
P.O. Box 9800, FI-02015 TKK, Finland
{firstname.surname}@hiit.fi

Elisabeth Andre
University of Augsburg
Eichleitnerstr.30
86159 Augsburg, Germany
andre@informatik.uni-augsburg.de

ABSTRACT

Recently, multimodal and affective technologies have been adopted to support expressive and engaging interaction, bringing up a plethora of new research questions. Among the challenges, two essential topics are 1) how to devise truly multimodal systems that can be used seamlessly for customized performance and content generation, and 2) how to utilize the tracking of emotional cues and respond to them in order to create affective interaction loops. We present PuppetWall, a multi-user, multimodal system intended for digitally augmented puppeteering. This application allows natural interaction to control puppets and manipulate playgrounds comprising background, props, and puppets. PuppetWall utilizes hand movement tracking, a multi-touch display and emotion speech recognition input for interfacing. Here we document the technical features of the system and an initial evaluation. The evaluation involved two professional actors and also aimed at exploring naturally emerging expressive speech categories. We conclude by summarizing challenges in tracking emotional cues from acoustic features and their relevance for the design of affective interactive systems.

Categories and Subject Descriptors

H5.2. **[Information Interfaces and Presentation]:** User Interfaces – *Input devices and strategies*, *Evaluation/methodology*, *Interaction styles*

General Terms

Design, Experimentation, Human Factors.

Keywords

Gestural interaction, Affective computing, Interactive Installations

1. INTRODUCTION

Natural interfaces can enable users to interact with advanced visual applications in a more embodied and expressive way. The latest development in multimodal processing concerns the tracking of expressive and emotional cues. These new interface technologies hold promise for providing tools to build more empathetic, surprising, and engaging applications. They could lead to innovative applications in which media are not just created and browsed but are also augmented in real time using multimodal and emotionally intelligent inputs. Our vision here is to support performative interaction [7] that encourages users to animate rich media and facilitate the genesis of new formats or practices in the new media field. Initial evidence of the relevance of these practices can be found in naturalistic trials of large multi-touch displays that make possible picture browsing and collage in a bodily way [13] or on systems that support the easy creation of comic strips from mobile pictures [16]. In this area the key research challenges for multimodal and emotion interface technologies include identifying the modalities to be used as input, investigating expressive features in each modality, and finally using them to create engaging interaction loops that motivate users to communicate more expressively.

This paper explores how to use multimodal emotional and expressive cues in digitally augmented puppeteering. The work is organized as follows: 1) reviewing related systems and input components found from the literature; 2) presenting an exemplar application, PuppetWall, that provides a medium for digital puppeteering with editable scenes, props, and puppets, and 3) providing feedback from initial evaluative activities regarding how inexperienced users perform with the help of the system. We conclude by summarizing our findings for future research.

2. RELATED WORK

The previous literature, which we will first consider concerns the development of independent input components for multimodal systems which could also be relevant for digital puppeteering interfaces. The interface can be a data glove and a custom sign language, which directly control the behavior of the digital character (for example, see [2]), without tracking and exploiting expressions or emotions. A more complete approach is I-Shadows [11], an interactive installation which utilizes Chinese shadow puppetry concept for kids creating narrative.

The use of emotion tracking for various kinds of interactive applications has been investigated. These studies may be valuable in showing how to decode or alter the affective states of a user. Existing interactive systems track affective states to influence in a direct or indirect way the essential contents of an interactive application. McQuiggan and Lester [9] have designed agents that are able to respond empathically to the gaming situation of the user. AffectivePainting [18] supports self-expression by adapting instantaneously to the perceived emotional state of a viewer, which is recognized from his or her facial expressions. Some empathic interface agents apply physiological measurements to sense users' emotional states

AVI'08, May 28-30, 2008, Napoli, Italy.

[15]. Gilleade et al. [5] measure users' frustrations to drive the adaptive behavior of an interactive system. There are also systems that extend the concept of empathy to account for the relation developed between the user and a virtual reality installation [8]. Cavazza et al. [3] present multimodal actors in a mixed-reality interactive storytelling application in which the positions, attitudes and gestures of the spectators are tracked, influencing the unfolding of the story.

Camurri et al. [2] introduce what they call multisensory integrated expressive environments as a framework for mixed-reality applications for performing arts and culture-oriented applications. They report an example where the lips and facial expression of an actress are tracked and the expressive cues are used to process her voice in real time. SenToy [12] allows users to express themselves by interacting with a tangible doll that is equipped with sensors to capture the users' gestures.. Isbister et al. [6] study uses 3D shapes to communicate emotions to the system and to the design team. However, we are not aware of any work which has applied the tracking of expressive cues from actors to animate or control puppet-like virtual characters.

3. PUPPETWALL

PuppetWall is a multi-user, multimodal installation for collective interaction based on the concept of traditional puppet theatre. When interacting with PuppetWall, users hold a wand in their hands that controls a puppet on a large touch screen in front of them. The touch screen is used to manipulate the playground, which consists of characters, props, and a background. The aim is to provide a platform for exploring emotion and multimodality with an interactive installation. Here we report on the design and details of the first prototype application.

3.1 System Overview

The PuppetWall system includes several input modalities for explicit and implicit control and a large multi-touch screen to visualize and edit the visual animations and scenes. An overview of the system is shown in Figure 1. The hardware of the prototype consists of both standard equipment and custom-made devices. The application runs on a single relatively high-performance PC (as of 2007) workstation and utilizes a Linux operating system. The workstation is equipped with IEEE1394 (FireWire) ports and a 3D accelerated graphics adapter. Input/output devices include a standard stereo microphone, pair of active speakers, a video projector (DLP, 1280 x 768 pixels), and three high-speed, high-resolution FireWire digital cameras, one of them equipped with an infra-red (IR) filter and a wide-angle lens (see 3.2.3). Interaction with the system is based on three inputs: hand movements via the detection of the MagicWand movements (see 3.2.1), direct manipulation through a touch screen (visualized in Figure 2; see 3.2.3) and voice input for tracking acoustic features of speech (see 3.2.2). The application reacts to these inputs to produce a visual 2.5-dimensional representation of virtual puppet theatre playground.

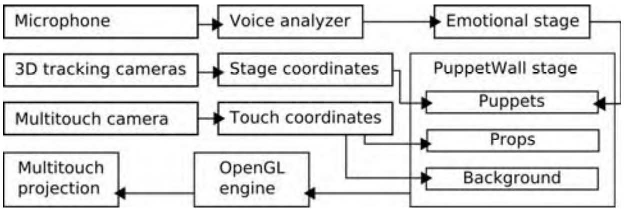

Figure 1. PuppetWall System overview. Three input components. Speech for tracking emotional state, 3D tracking of character control, and a touch screen to interact with objects or to edit puppets.

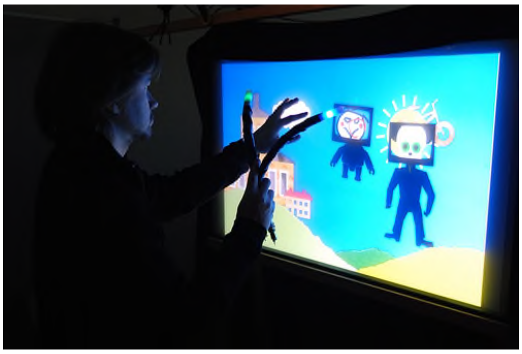

Figure 2. Prototype of PuppetWall and a user holding a MagicWand in their right hand and interacting with a prop with their left. There are two characters on the stage, partially hiding two props (then sun and a bicycle).

3.2 Input modalities

3.2.1 MagicWands for 3D hand tracking

The characters on stage are controlled with custom-made wands (MagicWands) which incorporate a light source, a single LED of variable color. This concept is similar to that of VisionWand [20]. Characters are moved and rotated according to the motion of the illuminated end of the wand and the users can mover one or more wands to control the motion of the puppets. A MagicWand is a plastic stick approximately thirty centimetres in length, consisting of a power source and a super-bright LED at the top end. Wands were assembled using standard electronic components. The super-bright LEDs are detected using a pair of digital cameras operating at 30 frames per second and mounted above the display. The camera image is used to calculate the 3D position of each wand. This happens by comparing the location of a bright spot on both of the camera images and the imaginary normal lines of the cameras. The movement is then interpreted into two-dimensional movements relative to the screen. All three coordinates can be used to control the character and different colours in the LED of the MagicWands are used to differentiate the characters. The wands are equipped with a power switch.

3.2.2 Emotional speech recognition

One essential requirement of the system is to be able to detect and respond to the user's emotions. Currently we are attempting to achieve this using emotional speech recognition. The user's voice, an essential element in building the narrative in this interactive storytelling environment, is captured using a single stereo microphone which feeds into a speech classifier. The classifier, called EmoVoice, is based on Naïve Bayesin classification of reduced feature sets (see [19]). This means that in the target language it has been trained to categorize the component should be able to discriminate between the defined emotional categories from arbitrary spoken input. The training of the initial version of the classifier was carried out with an extensive enacted Finnish emotional speech corpus including six emotion categories (see [17]). In an off-line state, it achieves some 45% accuracy. The preliminary setup is intended for testing (see Initial Evaluation below) and the hardware and the training corpus are subject to change in the future.

3.2.3 *Touch screen for direct manipulation of objects and characters*

A multi-touch screen (1 m wide) is used for displaying the PuppetWall playground and allows objects (props) and characters to be manipulated directly. The system enables multiple hand-tracking and individual hand posture and gesture tracking. Interfacing the screen is based on detecting changes in the IR luminosity from the screen surface, relying on a high-resolution, high-frequency camera and a robust computer vision algorithm. This concept is similar to HoloWall [10]. The technique requires an IR illumination of the screen from behind to level the incoming background IR signal. The movements on the screen surface are captured by the camera, which is also located behind the screen. A diffusing surface which is attached to the back of the screen surface blurs the object's IR image of the object, but when a user touches the screen it will show as a bright sharp spot in the IR camera image. These technological features create the conditions for a multi-user and multi-touch installation appropriate for public spaces (cf. [13]).

3.3 Visual Outputs

The main view of PuppetWall interface is called a playground and is comprised of characters, props, and the background (see Figure 2). All visuals are created with a custom-made 3D graphical engine based on OpenGL libraries. The interface presented on the touch screen is created with one, two or four projectors. The resulting screen resolution is a multiple 1280 x 768 pixels. The current prototype employs one projector.

3.3.1 *PuppetWall basic view*

Puppets are moved according to the movement detected from MagicWands. Puppets are able to swing around their pivot point so when the MagicWand is moved swiftly rotationally they also do a full rotation. Props – clouds, buildings and vehicles – can be moved and manipulated (resized, transformed, moved) on the fly by touch and gestures. Objects can be re-sized by touching them with multiple fingers and pulling touch points closer or pushing them further away from each other. Vehicle and building props will change into a different, larger one when certain size is reached. As an example: the positions of the sun and moon can be changed by rotating the plane containing them. They are placed on the opposite sides of the plane and the lighting conditions of the stage will change according to the state of the plane; it is lighter when the sun is up. The background elements are currently stationary.

3.3.2 *Character editing mode*

When a puppet is touched, the system enters an editing mode illustrated in Figure 3. In this mode, the user can modify the character by changing the puppet's head or body. They are lined up on the screen and the user can select different ones with a pulling gesture so that the desired shape moves toward the center (gesture-based browsing). Heads can be customized by drawing over the face with a finger, enlarging or shrinking the face or moving its relative position in the head frame.

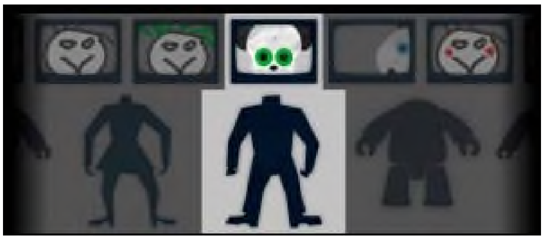

Figure 3. The editing mode. Heads and bodies can be selected by pulling them into the highlighted area.

4. INITIAL EVALUATION

An informal evaluation was organized with two performing arts professionals with a background in improvisational theater. They were involved in an explorative session with the first functional prototype. In the session, they experimented with in improvised and directed story-telling using the system for the first time. The experimental session was videotaped and the audio was additionally recorded with collar microphones to compile a corpus of naturally occurring interaction. The session began with a minimal debriefing and ended up with a structured interview for feedback from the interactive session.

It was observed that the actors could easily utilize the installation with minimal instructions. The actors used the touch interface to modify the puppets and the props, and MagicWands were used successfully to animate the characters. The users seemed to enjoy PuppetWall and created eight short stories with it. The actors, familiar with improvisation, suggested the use of implicit interaction strategies, for example using breathing sensors and adding more surprising elements to the scene. Also the actors complained that while constraints are useful to make things happen in improvisation, they felt too limited by the control of the puppet, as they could not implement all ideas. In addition to new development ideas and usability issues observed from video the analysis, we found that the corpus extracted from the speech naturally elicited during the session, could be meaningfully classified with EmoVoice. The most reliable classification appeared between what could be called 'user' and 'character' voices (68% off-line discrimination). The user voice was low, inactive, and constrained whereas the 'character' voice was active, engaging, and openly emotional.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have introduced PuppetWall, an interactive application for exploring affective interaction and multimodal inputs in an environment intended for multiple, simultaneous users. In addition to providing the technical details, we have described an informal evaluation of the system. Our evaluation demonstrated the feasibility of the concept and also provided a preliminary corpus of affective speech. An important result from the analysis of the corpus with the EmoVoice classifier was the demonstration of how the neutral user and character voices are differentiated along a 'dimension of activation'.

In the future, even if we are able to confront the problem of how to decode user emotions, we still face the additional problem of responding to these emotions. While decoding has received a lot of attention, the other half of the work has barely started and currently, no clear guidelines exist on how to engineer affective responses or to augment emotion. In current HCI, the bestknown collection of techniques is called Emotioneering [4] a set of heuristics for emotional game design. Their problem is their considerable domain dependence. Only a few heuristics, such as the use of symbols, can be transferred to other domains. Additional examples from the literature show context-dependent solutions, for instance analyzing call center requests for later prioritization according to affective status [14] or applications utilizing emotion recognition in the form of a game to help individuals to recognize and manage emotions [1, 14]. One generic approach available in some contexts, as with PuppetWall, might be to recruit professionals in the domain in question to participate in the design process. This co-design can be helpful to exploit the vast knowledge that the experts possess about expressivity in their domain.

In conclusion, the prototype of PuppetWall presented here is the first step in developing a platform studying multimodal and affective interaction techniques. While this step provided useful indications on the feasibility and relevance of the concept, several questions remain open, for instance, which modalities to address and how gestural information could be utilized. However, from the initial evaluation and later co-design (to be documented) we have gained considerable knowledge and many ideas for future development and user research that will, we hope, highlight PuppetWall as a state-of-the-art example of collocated, emotionally augmented interactive installation.

6. ACKNOWLEDGMENTS

This work has been funded by the European Union (EU) 6th Framework Research Programme project CALLAS (ref. 034800). We are grateful to Tommi Ilmonen and John Evans, who designed and implemented the prototype, and thank Jérôme Urbain and Stephen W. Gilroy for providing some references.

7. REFERENCES

[1] Bersak, D., McDarby, G., Augenblick, N., McDarby, P., McDonnell, D., McDonald, B. and Karkun, R. Intelligent biofeedback using an immersive competitive environment. In Proceedings of the Designing Ubiquitous Computing Games Workshop at UbiComp (2001).

[2] Camurri, A., Volpe, G., De Poli, G. and Leman, M. Communicating expressiveness and affect in multimodal interactive systems. IEEE Multimedia, 12, 1 (Jan-Mar 2005), 43-53.

[3] Cavazza, M., Charles, F., Mead, S. J., Martin, O., Marichal, X. and Nandi, A. Multimodal acting in mixed reality interactive storytelling. IEEE Multimedia, 11, 3 (Jul-Sep 2004), 30-39.

[4] Freeman, D. Creating emotion in games. The craft and art of emotioneering. New Riders, Indianapolis, IN, 2003.

[5] Gilleade, K. M. and Dix, A. Using frustration in the design of adaptive videogames. In Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology (Singapore, 2004). ACM Press.

[6] Isbister, K. and Hook, K. Evaluating affective interactions. International Journal of Human-Computer Studies, 65, 4 (Apr 2007), 273-274.

[7] Jacucci, G. Interaction as Performance. Doctoral dissertation, University of Oulu, 2004.

[8] Lugrin, J. L., Cavazza, M., Palmer, M. and Crooks, S. AI-Mediated Interaction in Virtual Reality Art. In Proceedings of the Intelligent Technologies for Interactive Entertainment: First International Conference (INTETAIN 2005) (Madonna di Campiglio, Italy, 2005). Springer-Verlag.

[9] McQuiggan, S. W. and Lester, J. C. Modeling and evaluating empathy in embodied companion agents. International Journal of Human-Computer Studies, 65, 4 (Apr 2007), 348-360.

[10] Nobuyuki, M. and Jun, R. HoloWall: designing a finger, hand, body, and object sensitive wall. In Proceedings of the 10th annual ACM symposium on User Interface Software and Technology (UIST) (Banff, Alberta, Canada, 1997). ACM.

[11] Paiva, A., Fernandes, M. and Brisson, A. Children as affective designers - i-shadows development process. Humaine WP9 Workshop on Innovative Approaches for Evaluating Affective Systems(2006), (accessed,

[12] Paiva, A., Prada, R., Chaves, R., Vala, M., Bullock, A., Andersson, G. and Höök, K. Towards tangibility in gameplay: building a tangible affective interface for a computer game. In Proceedings of the 5th international conference on Multimodal interfaces (Vancouver, BC, 5-7 November, 2003).

[13] Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A. and Saarikko, P. "It's mine, don't touch!": Interactions at a large multi-touch display in a city center. In Proceedings of the CHI2008 (to appear, 2008).

[14] Petrushin, V. A. Emotion Recognition In Speech Signal: Experimental Study, Development, And Application. In Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000) (Beijing, China, 2000).

[15] Prendinger, H. and Ishizuka, M. Human physiology as a basis for designing and evaluating affective communication with life-like characters. IEICE Transactions on Information and Systems, E88D, 11 (Nov 2005), 2453-2460.

[16] Salovaara, A. Appropriation of a MMS-based comic creator: from system functionalities to resources for action. In Proceedings of the SIGCHI conference on Human factors in computing systems (San Jose, CA, April 28-May 4, 2007). New York, NY: ACM Press.

[17] Seppänen, T., Toivanen, J. and Väyrynen, E. MediaTeam speech corpus: a first large Finnish emotional speech database. In Proceedings of the Proceedings of XV International Conference of Phonetic Science (Barcelona, Spain, 2003).

[18] Shugrina, M., Betke, M. and Collomosse, J. Empathic painting: interactive stylization through observed emotional state. In Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering (NPAR 2006) (Annecy, France, 2006). ACM Press.

[19] Vogt, T. and Andre, E. Comparing Feature Sets for Acted and Spontaneous Speech in View of Automatic Emotion Recognition. Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on (2005), 474-477.

[20] Xiang, C. and Ravin, B. VisionWand: interaction techniques for large displays using a passive wand tracked in 3D. In Proceedings of the 16th annual ACM symposium on User Interface Software and Technology (UIST) (Vancouver, Canada, 2003). ACM.