Emote to Win: Affective Interactions with a Computer Game Agent

Jonghwa Kim, Nikolaus Bee, Johannes Wagner and Elisabeth André

Multimedia Concepts and Application, Faculty for Applied Computer Science
University of Augsburg, Eichleitnerstr. 30, 86159 Augsburg, Germany
{jonghwa.kim, andre}@informatik.uni-augsburg.de

Abstract: In this paper[1], we introduce a game interface that is based on affective interactions between a player and a computer pet. As opposed to many traditional computer games, users are not expected to manipulate control devices in a skillful manner to win the game. Instead the basic idea is to elicit certain reactions of the pet via appropriate emotive user behaviors. For improved accuracy of emotion recognition, we employ a combined analysis of signals from two emotive expression channels: affective speech and physiological reactions.

1 Introduction

With advances in the area of sensor technology and signal processing, a new generation of computer games based on biofeedback and affective computing is emerging. In order to increase a player's level of immersion and engagement (see [Fr03]), a number of affective games encourage a user to express his or her emotive states and dynamically adapt actions and events to them. In Girland, young girls may use an emotion wheel to input their emotive states, see http://www.girland.com/. In SenToy, the user interacts with a tangible doll to communicate one of six emotions through gestures (see [PCP+03]). Games relying on biofeedback aim at helping the user to gain control over his or her bodily reactions (e.g. see [LSB+03]). For instance, in the game "The Wild Divine" the player needs to achieve a state of relaxation or concentration to manipulate devices of the game environment, see http://www.wilddivine.com/. Another example of an affective interface is the computer game FinFin where the user may influence the emotional state of a half bird, half dolphin creature via talking and waving.

In this paper we propose an affective user interface to a computer game based on a new paradigm called "Emote to Win". The basic idea is to influence the behavior of a virtual pet - Tiffany, the snail - by appropriate emotional reactions of the user. While interacting with the snail, the user is monitored by means of bio sensors, measuring skin conductivity, heart rate, respiration and muscle activity. In addition, we collect information on the user's emotional state by analyzing his or her speech input. Tiffany, the snail, is not supposed

[1]We would like to thank the students of the Multimedia Student Project "Spiel mit Emotionen" for their work on the implementation of Tiffany.

to understand the user's speech input. However, similar to a real pet, it is sensitive to the user's emotive state and tries to respond to it accordingly. As in "The Wild Divine", there are several options to play the game. For instance, the user may aim at getting Tiffany to perform certain actions within a limited time (to have a kind of winning condition as in traditional games) or communicate with it in an open-end mode in order to learn how to control his or her emotional response.

The central components of our game environment are: modules for the recognition of emotions from biosignals and speech, a decision module for fusing the results from the single input channels and finally a module for determining and animating the behaviors of the virtual snail. For signal processing, Matlab was used. To enable real-time interactions, Tiffany's Java-based animation engine was linked with the emotion recognition system via Simulink/Matlab. Figure 1 provides an overview of the system.

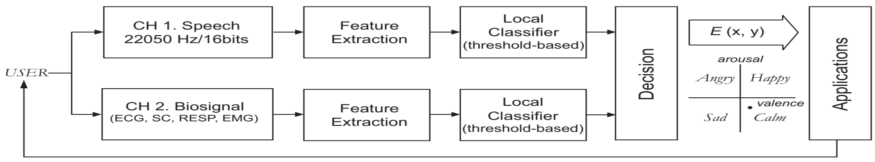

Figure 1: Conceptual Overview of the Affective Game Environment

2 The Underlying Emotion Model

Two represent emotions, we follow a dimensional approach [La02] which characterizes emotions in terms of several continuous dimensions, such as arousal or valence. Arousal refers to the intensity of an emotional response. Valence determines whether an emotion is positive or negative and to what degree. A given emotional state is then characterized by a point in this two dimensional space. Emotion dimensions can be seen as a simplified representation of the essential properties of emotions. For instance, anger can be described by high arousal and negative valence.

Apart from the ease of describing emotional states that cannot be distributed into clear-cut fixed categories, the two dimensions valence and arousal are well suited for emotion recognition. As a first step, we aim at recognizing four distinct emotional states that may be associated with one of the four quadrants of our dimensional emotional model: angry (valence: negative, arousal: high), calm (valence: positive, arousal: low), sad (valence: negative, arousal: low) and happy (valence: positive, arousal: high).

3 Two-Channel Emotion Recognition

Recently, there has been a signicant amount of work on the recognition of emotions (see [CDCT+01] for an overview). Most approaches to emotion recognition so far concentrate on a single modality and do not take advantage of the fact that an integrated multimodal analysis may help to resolve ambiguities and compensate for errors. An exception is the work by Chen and colleagues [CTH+98] who present an approach to emotion recognition from speech and video, showing that higher recognition rates can be achieved by a bimodal analysis. The great challenge of our work is to recognize emotional signals from multiple channels while the player interacts with the application and to respond to them in real-time. There are several advantages of using biosensor feedback in addition to affective speech. First of all, we can continuously gather information on the user's emotive state through biosensors while the analysis of emotions from speech should only be triggered when the microphone receives speech signals from the user. Secondly, it is much harder for the user to deliberately manipulate biofeedfack than external channels of expression.

3.1 Data Collection

To identify features from biosignals and speech whose values may be updated with reasonable computational cost, we collected a corpus containing physiological and speech data. In particular, we presented four subjects with a number of video sequences that are supposed to evoke one of the four emotions mentioned above and asked them for verbal feedback.

While the subjects watched the video sequences, their physiological feedback was recorded using four biosensors. The ECG sensor (electrocardiogram) was used to measure the heart rate. The EMG sensor (electromyogram) determines the activity of muscles in the forehead to detect whether or not the user is frowning. The RESP sensor (respiration rate) captures abdominal breathing. The SC sensor (skin conductivity) measures sweat secretion, which was taken at the index and ring finger of the non-dominant hand. During the experiment, we also gathered the speech signals resulting from spontaneous verbal responses to the video sequences.

3.2 Feature Selection and Classification

There has been a great amount of work in the development of offline recognition algorithms based on discriminative or generative classifier models, such such as SVM (support vector machine), HMM (hidden markov model, and MLP (multilayer perception), see [CDCT+01]. Unfortunately, these algorithms cannot be directly adapted to the challenging task of online recognition where the emotional state needs to be recognized as soon as possible (see [VP99] for discussion of these problems). As a first step, we decided to start from simple threshold-based classification methods. Secondly, we relied on a fixed set of pre-selected features based on an offline analysis of the collected data.

From the speech signal, we extracted 23 features derived from pitch, harmonics and energy. From the biosignal, we extracted 25 features including derivations of mean energy from SC/EMG, standard deviation from SC/EMG, and four subband spectra from

RESP/ECG. In Figure 2, examples of some features are shown with raw signals.

Table 1 shows how the recognition of emotions may profit from a combined analysis. For instance, skin conductance seems to be a good indicator for arousal. Determining the valence of a emotion seems to be much tougher. However, the voice harmonics may help to distinguish, for example, positive emotions with high arousal from negative emotions with high arousal.

4 Affective Behavior of the Game Character

Information on the user's emotive state is forwarded to Tiffany in terms of two coordinates that represent a point in the dimensional emotion model (x: valence, y: arousal). To respond to the user's emotional state, information on the user's emotive state is mapped onto appropriate facial and body movements of the snail (see Fig. 3). In addition, the snail is able to express emotions via sound. For instance, the user may scare the snail away by shouting at her angrily. As a consequence, Tiffany will hide behind a bush. In case Tiffany does not receives any utilizable input from the recognition module, for instance if the user does not show any interpretable reaction or if the emotion analysis does not lead to any consistent results, certain idle time behaviors, such as eye blinking, will be elicited. Note that there is a bi-directional flow of emotions between the user and the virtual pet. The snail does not only respond to the user's emotive behaviors, but in turn evokes emotions in the user as well. For example, the user may get irritated or bored if Tiffany refuses to emerge from behind the bush.

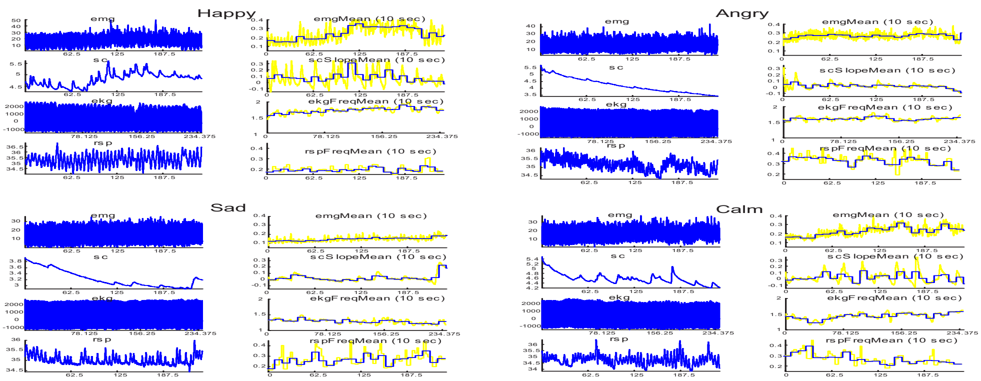

Figure 2: Biosignal Examples with Selected Features

	ECG (heart rate)	RESP (Resp. Freq.)	SC (Mean Value)	EMG (Mean Value)	Pitch (F0)	Harmonics (Fk-Cand.)	Energy (Mean Value)
Happy	+	+	+	+	++/-	-	+
Angry	++	+	++	+	++	+	++
Sad	+	+/-	-	-	-	- -	- -
Calm	- -	-	- -	-	+/-	+/-	-

Table 1: Emotional Cues in Speech and Biosignal

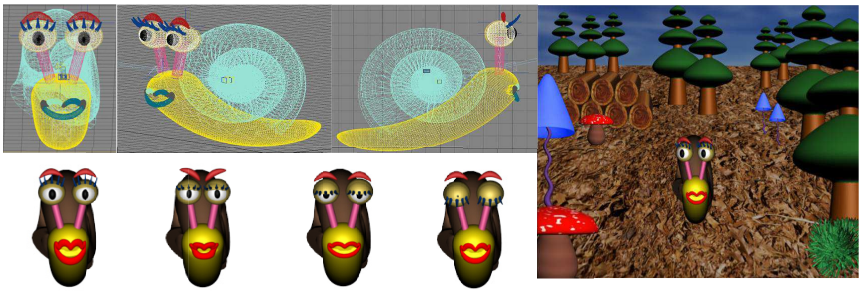

Figure 3: The Emotion Character "Tiffany"

5 Discussion and Conclusion

In this paper, we presented a first prototype of a gaming application for exploring new intuitive control strategies based on biophysiological feedback and affective speech. The game is based on the player's ability to express him- or herself emotionally.

Our experience has shown that affective speech and biofeedback can be integrated in a computer game in a natural manner. As opposed to the user interfaces of many traditional computer games, the user did not have to learn complex control sequences for manipulating the game environment. Nevertheless, we identified a number of usability problems. Even though the game logic was quite simple, the behavior of the virtual snail was not always obvious to the user. By their very nature, sensor data are heavily affected by noise and especially bio sensor systems are, among other things, very sensitive to motion artefacts. As a consequence, it was hard to acquire reliable information on the user's emotive state. To improve the recognition rate of our system, we are currently collecting a more comprehensive corpus of bio sensor and speech data. In addition, we are performing experiments with selected statistical clustering methods for offline analysis and investigate in how far they can be adapted to online analysis.

We also noticed that some users had problems with the control mechanisms of the game since the mapping between their emotive input and Tiffany's resulting action was not always clear to them. For instance, some of the users expected Tiffany to reflect their own emotions as opposed to responding to them. Others had problems to identify Tiffany's emotions from its bodily and facial behaviors. To remedy this, we intend to increase the transparency of the interface, e.g. by providing the user simple instructions on how to influence Tiffany's behavior. Furthermore, we will exhibit Tiffany with more expressive animations and make more extensive use of sound.

We conclude that "Emote to Win" is an interesting new paradigm that can help users to increase their awareness of their affective behavior and to deploy affect consciously to communicate (e.g., to control a game character).

References

[CDCT+01] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., und Taylor, J.: Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*. 18(1):32–80. 2001.

[CTH+98] Chen, L., Tao, H., Huang, T., Miyasato, T., und Nakatsu, R.: Emotion recognition from audiovisual information. In: *Proceedings, IEEE Workshop on Multimedia Signal Processing*. S. 83–88. Los Angeles, CA, USA. 1998.

[Fr03] Freeman, D.: *Creating Emotion in Games: The Craft and Art of Emotioneering*. New Riders. 2003.

[La02] Lang, P. J.: The emotion probe: Studies of motivation and attention. *American Psychologist*. 50(5):372–385. 2002.

[LSB+03] Lyons, G., Sharma, P., Baker, M., O'Malley, S., und Shanahan, A.: A computer game-based emg biofeedback system for muscle rehabilitation. In: *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*. volume 2. S. 1625–1628. Sept. 2003.

[PCP+03] Paiva, A., Chaves, R., Piedade, M., Bullock, A., Andersson, G., und Höök, K.: Sentoy: A tangible interface to control the emotions of a synthetic character. In: *Proc. of AAMAS 2003*. S. 1088–1089. 2003.

[VP99] Vyzas, E. und Picard, R.: Online and offline recognition of emotion expression from physiological data. In: *Workshop on Emotion-Based Agent Architectures at the International Conference on Autonomous Agents*. Seattle, WA, USA. 1999.