VIRTUAL REALITY AND AFFECTIVE COMPUTING TECHNIQUES FOR FACE-TO-FACE COMMUNICATION

Hamza Hamdi[1,2], Paul Richard[1], Aymeric Suteau[1] and Mehdi Saleh[2]
[1] *Laboratoire d'Ingénierie des Systèmes Automatisés (LISA), Université d'Angers, 62 Avenue ND du Lac, Angers, France*
[2]*I-MAGINER, 8 rue Monteil, Nantes, France*

Keywords: Virtual reality, Human-computer interaction, Emotion recognition, Affective computing, Job interview.

Abstract: We present a multi-modal affective virtual environment (VE) for job interview training. The proposed platform aims to support real-time emotion-based simulations between an ECA and a human. The first goal is to train candidates (students, job hunters, etc.) to better master their emotional states and behavioral skills. The users' emotional and behavior states will be assessed using different human-machine interfaces and biofeedback sensors. Collected data will be processed in real-time by a behavioral engine. A preliminary experiment was carried out to analyze the correspondence between the users' perceived emotional states and the collected data. Participants were instructed to look at a series of sixty IAPS pictures and rate each picture on the following dimensions : joy, anger, surprise, disgust, fear and sadness.

1 INTRODUCTION

There is an increasing interest in developing intelligent human-computer interaction systems that can recognize user affective states. Affective Computing (AC) aspires to narrow the communicative gap between humans and computers by developing computational systems that recognize and respond to the user's affective states. Emotions constitute a privileged support to model Embodied Conversational Agents (ECAs) that are able to communicate verbally but also through gestures, facial expressions, postures, and speech. Different systems have been developed in the last decade using ECAs (Woolf et al., 2009) (Helmut et al., 2005). However, none of these systems allow realistic immersive multi-modal emotion-based dialogue between an ECA and a human.

In this paper, we describe a multi-modal affective virtual environment (VE) for job interview training. The proposed platform aims to train candidates (students, job hunters, etc.) to better master their emotional states and behavioral skills using an Embodied Conversational Agent (ECA).

In the next section, we survey the related work concerning the classification and the recognition of human emotions. In section three, we present the platform architecture and the human-machine interfaces. In section four, we describe a preliminary experiment based on IAPS protocol.

2 RELATED WORK

Emotions are recognized as involving components such as cognitive and physiological changes, trends in the action and motor expressions. Darwin postulated the existence of a finite number of emotions present in all cultures and having a function of adaptation (Darwin, 1872). This postulate was subsequently confirmed by Ekman which divided the emotions into two classes: the primary emotions (joy, sadness, anger, fear, disgust, surprise) which are natural responses to a given stimuli, and secondary emotions that evoke a mental image which correlates with the memory of a primary emotion (Ekman, 1999). Emotions can be represented by discrete categories (e.g. "anger") or defined by continuous dimensions such as "Valence", "Activation", or "Dominance". These three dimensions were combined in a space called PAD (Pleasure, Arousal, Dominance) originally defined by Mehrabian (Mehrabian, 1996).

2.1 Emotion Recognition

Several approaches based on facial expression recognition have been proposed to classify human

Hamza H., Richard P., Suteau A. and Saleh M..
VIRTUAL REALITY AND AFFECTIVE COMPUTING TECHNIQUES FOR FACE-TO-FACE COMMUNICATION.
DOI: 10.5220/0003377203570360
In *Proceedings of the International Conference on Computer Graphics Theory and Applications* (GRAPP-2011), pages 357-360
ISBN: 978-989-8425-45-4

emotional states (Pantic and Rothkrantz, 2003). Tian (Tian et al., 2000) has attempted to recognize Action Units (AU), developed by Ekman and Friesen in 1978 (Ekman and Friesen, 1978) using permanent and transient features of the face and lips, the nasolabial fold and wrinkles. Hammal proposed an approach based on the combination of two models for segmentation of emotions and dynamic recognition of facial expressions (Hammal and Massot, 2010).

Several approaches aimed to recognize emotions from speech (Pantic and Rothkrantz, 2003) (Scherer, 2003). For example, Roy and Pentland classified the emotions by using a Fisher linear classifier (Roy and Pentland, 1996). Using short sentences, they have recognized two kinds of emotions: approval and disapproval. They conducted several experiments with characteristics extracted from measurements of height and power, obtaining a precision going from 65% to 88%.

The analysis of physiological signals is another possible approach for emotion recognition (Healey and Picard, 2000) (Picard et al., 2001). Several types of physiological signals can be used to recognize emotions. For example, heart rate, skin conductance, muscle activity (EMG), temperature variations of the skin, variation of blood pressure are signals regularly used in this context (Lisetti and Nasoz, 2004) (Villon, 2007).

2.2 Multi-modal Approaches

Multi-modal emotion recognition requires the fusion of the collected data. Physiological signals are then mixed with other signals collected through human-machine interfaces such as video or infrared cameras (gestures, etc.), microphones (speech), brain computer interfaces (BCIs) (Lisetti and Nasoz, 2004) (Sebe et al., 2005) (Busso et al., 2004). Multimodal information fusion may be performed at different levels. Usually the three following levels are considered : Signal level, Feature level, and Decision or Conceptual level.

Fusing information at the signal level actually means to mix two or more signals before extracting features required by the decision maker (Paleari and Lisetti, 2006). Fusing information at the feature level means to mix together the features issued from different signal processors. Features extracted from each modality are fused before being transmitted to the decision maker module (Pantic and Rothkrantz, 2003). Combining information at the conceptual level does not mean mixing together features or signals but directly the extracted semantic information. Decision level fusion of multi-modal information is preferred by most researchers. Busso (Busso et al., 2004) compared the feature level and the decision level fusion techniques, observing that the overall performance of the two approaches is the same.

Our goal is to propose a model allowing to analyze the behavior of the various signals and to build a system of emotion detection in real-time using multimodal fusion. We aim to identify the six universal emotions listed by Ekman and Friesen (Ekman and Friesen, 1978) (anger, disgust, fear, happiness, sadness and joy), to which we add despise, stress, concentration and excitation (Calvo and D'Mello, 2010).

2.3 International Affective Picture System (IAPS)

Different methods have been used to investigate human emotions, ranging from imagery inductions to film clips and static pictures. The International Affective Picture System (IAPS) is one of the most widely used stimulus sets (Lang et al., 1999). This set of static images is based on a dimensional model of emotion. It contains various pictures depicting mutilations, insects, attack scenes, snakes, accidents, etc. IAPS based experiments have also shown that discrete emotions (disgust, sadness, fear, etc.) have different valence and arousal ratings, and can be distinguished by facial electromyography, heart rate, and electrodermal measures (Bradley et al., 2001).

Figure 1: Snapshot of interview simulation.

3 SYSTEM OVERVIEW

3.1 System Architecture

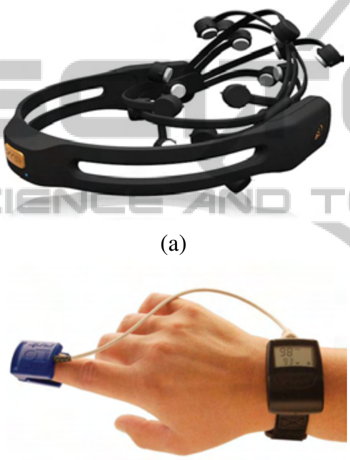
The proposed platform aims to support real-time simulations (Fig. 1) that allow emotion-based face-to-face dialogue between an ECA and a human. The

ECA can have specific personality and behavior (gentle, aggressive, passive, etc.).

3.2 Human Computer Interfaces

The main challenge is to identify and classify behavioral and emotional states of the participant using non-intrusive human-computer interfaces :

- Brain computer interface: Emotiv EPOC (Fig. 2 (a));
- Biofeedback sensor: Nonin (Fig. 2 (b));
- Speech and facial recognition devices: microphone, webcam.



(a)

(b)

Figure 2: Human-machine interfaces: (a) Emotiv EPOC, (b) Nonin.

Different signals and input modalities are been considered:

1. Physiological signals:
   - Facial Electromyography (EMG) ;
   - Electrocardiogram (ECG) ;
   - Electroencephalography (EEG) ;
   - The galvanic skin response (GSR) ;
2. Speech:
   - The user's emotional state is estimated through speech analysis (pitch, tone, speed).
3. Text:
   - The user's emotional state is estimated through textual content.
4. Gestures:
   - The user's emotional state is estimated through static and dynamic gestures.

4 PRELIMINARY EXPERIMENT

In order to allow the development and the integration of behavioral and emotional models in the platform, we carried out a preliminary experiment. We seek to analyze the correspondence between the users perceived emotional states and the data collected from the biofeedback sensor (Nonin Oxymeter) and the brain-computer interface (Emotiv EPOC). Participants were instructed to look at a series of sixty IAPS pictures and rate each picture on the following dimensions: joy, anger, surprise, disgust, fear, sadness. They were all equipped with Nonin Oxymeter and Emotiv EPOC as illustrated in Figure 3. Furthermore, a camera was used during the experiment to take some snapshots of the participants.

Figure 3: Subject during the preliminary experiment.

5 CONCLUSIONS

A multi-modal affective virtual environment (VE) has been presented, aiming to support real-time emotion-based simulations between an ECA and a human. The first goal is to train candidates to better master their emotional states and behavioral skills. Human-machine interfaces and biofeedback sensors are used to assess users' emotional and behavior states. A preliminary experiment was carried out. The goal was to analyze the correspondence between the users' perceived emotional states and the collected data. Participants were instructed to look at a series of sixty IAPS pictures and rate each picture on the following dimension : joy, anger, surprise, disgust, fear, sadness.This work opens the way to new possibilities in different areas such as professional or medical applications and contributes to the democratization of emotion-based human-machine interfaces for face-to-face communication and interaction.

REFERENCES

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and motivation i: Defensive and appetitive reactions in picture processing. *Emotion*, 1(3):276–298.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Lee, S., Neumann, U., and Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 205–211, New York, NY, USA. ACM.

Calvo, R. A. and D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transaction on Affective Computing*, 1:18–37.

Darwin, C. (1872). *The expression of emotion in man and animal*. University of Chicago Press (reprinted in 1965), Chicago.

Ekman, P. (1999). *Basic emotions*, pages 301–320. Sussex U.K.: John Wiley and Sons, Ltd, New York.

Ekman, P. and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for Measurement of Facial Movement*. Consulting Psychologists Press Palo Alto, California.

Hammal, Z. and Massot, C. (2010). Holistic and feature-based information towards dynamic multi-expressions recognition. In *VISAPP 2010. International Conference on Computer Vision Theory and Applications*, volume 2, pages 300–309.

Healey, J. and Picard, R. W. (2000). Smartcar: Detecting driver stress. In *In Proceedings of ICPR'00*, pages 218–221, Barcelona, Spain.

Helmut, P., Junichiro, M., and Mitsuru, I. (2005). Recognizing, modeling, and responding to users' affective states. In *User Modeling*, pages 60–69.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1999). International affective picture system (iaps). *Technical manual and affective ratings*.

Lisetti, C. and Nasoz, F. (2004). Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP J. Appl. Signal Process*, 2004:1672–1687.

Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292.

Paleari, M. and Lisetti, C. L. (2006). Toward multimodal fusion of affective cues. In *Proceedings of the 1st ACM international workshop on Human-Centered Multimedia*, pages 99–108, New York, NY, USA. ACM.

Pantic, M. and Rothkrantz, L. (2003). Toward an affect-sensitive multimodal human-computer interaction. volume 91, pages 1370–1390. Proceedings of the IEEE.

Picard, R., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191.

Roy, D. and Pentland, A. (1996). Automatic spoken affect classification and analysis. automatic face and gesture recognition. In *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96)*, pages 363–367, Washington, DC, USA. IEEE Computer Society.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(7-8):227–256.

Sebe, N., Cohen, I., and Huang, T. (2005). *Multimodal Emotion Recognition*. World Scientific.

Tian, Y., Kanade, T., and Cohn, J. (2000). Recognizing lower face action units for facial expression analysis. pages 484–490. Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00).

Villon, O. (2007). *Modeling affective evaluation of multimedia contents: user models to associate subjective experience, physiological expression and contents description*. PhD thesis, Thesis.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., and Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *Int. J. Learn. Technol.*, 4(3/4):129–164.