Enhancement of Human Computer Interaction with facial Electromyographic sensors

Guillaume Gibert
MARCS Auditory Laboratories
University of Western Sydney
Locked Bag 1797
Penrith South DC
NSW 1797 Australia
g.gibert@uws.edu.au

Martin Pruzinec
Cognitive Systems Lab
University of Karlsruhe (TH)
Adenauerring 4
76131 Karlsruhe
Germany
pruzinec@ira.uka.de

Tanja Schultz
Cognitive Systems Lab
University of Karlsruhe (TH)
Adenauerring 4
76131 Karlsruhe
Germany
tanja@ira.uka.de

Catherine Stevens
MARCS Auditory Laboratories
University of Western Sydney
Locked Bag 1797
Penrith South DC
NSW 1797 Australia
kj.stevens@uws.edu.au

ABSTRACT

In this paper we describe a way to enhance human computer interaction using facial Electromyographic (EMG) sensors. Indeed, to know the emotional state of the user enables adaptable interaction specific to the mood of the user. This way, Human Computer Interaction (HCI) will gain in ergonomics and ecological validity. While expressions recognition systems based on video need exaggerated facial expressions to reach high recognition rates, the technique we developed using electrophysiological data enables faster detection of facial expressions and even in the presence of subtle movements. Features from 8 EMG sensors located around the face were extracted. Gaussian models for six basic facial expressions - anger, surprise, disgust, happiness, sadness and neutral - were learnt from these features and provide a mean recognition rate of 92%. Finally, a prototype of one possible application of this system was developed wherein the output of the recognizer was sent to the expressions module of a 3D avatar that then mimicked the expression.

Author Keywords

EMG, Facial expressions, Gaussian models

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): H.5.2 User Interfaces: Input devices and strategies.

INTRODUCTION

In the field of Human-Computer Interaction (HCI) there is major interest in creating computer systems that are able to understand all natural communication channels of a human user. In Human-Human communication only a small amount of the actual information is exchanged through spoken language. Through channels like body language and facial expressions, additional information is transferred. Including this information into HCI will contribute to creating affective systems that can change their status and react automatically according to the state of the user. In the same way, human to human communication through machines (such as in games) can be improved by detecting and transferring additional information to a display for example to a 3D avatar in virtual environments.

(Ekman and Friesen 1971) described six primary emotions that reflect distinguishable expressions: anger, disgust, fear, happiness, sadness and surprise. These emotional displays are often referred to as basic emotions. They are recognized by humans independently of their cultural background and their ancestry. A large amount of work has been devoted to recognise these facial expressions based on video inputs. Two main approaches have been developed using appearance models (Bartlett, Littlewort et al. 2006) or by determining the locations of fiducial points (Tian, Kanade et al. 2001). These systems are in general sensitive to the user location, the lighting conditions and the viewing angle. Few studies have provided insights in the recognition of facial expressions using EMG signals. (Ang, Belen et al. 2004) proposed a recognition system using 3 EMG electrodes able to distinguish between 3 basic expressions: happiness, anger and sadness. They reached an accuracy of 94.44% using 4 features (mean, standard deviation, Root Mean Square (RMS) and the power density spectrum) and a minimum-distance classifier. (Chin, Ang et al. 2008) investigated the classification of facial expressions from electroencephalogram (EEG) and EMG signals using the Filter Bank Common Spatial Pattern. They used 32 sensors recording EEG and EMG data and obtained an accuracy of 86% in recognizing 6 types of facial expressions. By reducing the number of sensors (6 frontal electrodes for a headband implementation), the accuracy decreased significantly to 78%.

In the present study, we describe a light expression recognition system based on 8 facial EMG sensors placed on specific muscles able to discriminate 6 expressions. We first address the experimental description of the recordings of one subject performing 6 facial expressions. Then, we explain the method used to extract features from the 8 EMG signals and to build a classifier. Results of classification in terms of recognition rate and similarity

OZCHI 2009, November 23-27, 2009, Melbourne, Australia.
Copyright the author(s) and CHISIG
Additional copies are available at the ACM Digital Library (http://portal.acm.org/dl.cfm) or ordered from the CHISIG secretary (secretary@chisig.org)
OZCHI 2009 Proceedings ISBN: 978-1-60558-854-4

of the distributions of the expressions are provided. Finally a prototype of an online scenario using this system is described.

METHODS

Subject

One healthy male volunteer took part to make the recordings. The protocol of this experiment was approved by the University of Western Sydney Human Ethics Research Committee.

Data acquisition

EMG activity was recorded continuously from 8 Ag/AgCl bipolar electrodes grounded to the right wrist. EMG signals were amplified and digitized at a rate of 600 Hz using a Varioport III amplifier (Becker 2003). The EMG was collected and stored using UKA EMG/EEG Studio (UKAv2) software from the ITI Waibel labs at the University of Karlsruhe.

Sensors positions

According to the commonly used guidelines for human electromyographic research (Fridlund and Cacioppo 1986) the electrodes were placed on height major mimic muscles mainly on one side of the face (see Figure 1).

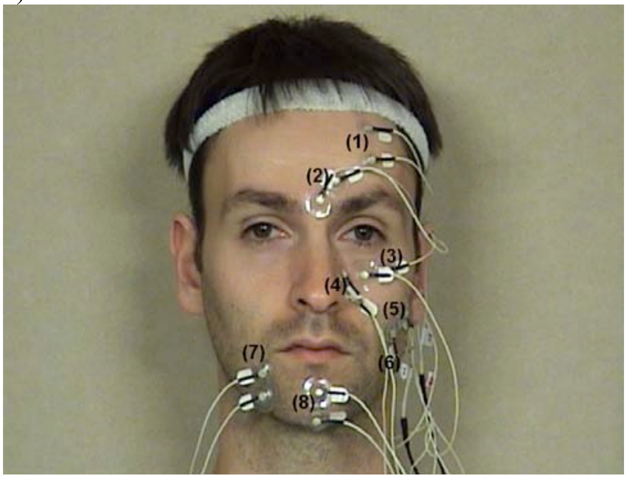

Figure 1: Sensors position used during the recordings

The muscles from which activity was recorded and their functions are:

1. **Venter frontalis** pulls the eyebrow up;
2. **Corrugator supercilii** pulls the eyebrow to the medial corner and down;
3. **Orbicularis oculi** constricts skin around the eye;
4. **Levator labii** wrinkles the nose, stretches nasal wings and rises upper lip;
5. **Zygomaticus major** pulls mouth corners upwards and laterally;
6. **Masseter** raises jaw and presses teeth together;
7. **Depressor anguli oris** controls shape and size of mouth opening;
8. **Mentalis** pushes skin above chin upwards and curves lips upwards.

Procedure

The subject was seated in a sound insulated research booth. The stimuli presentation was handled by a laptop situated at 1.2 m from the subject. The electrodes were attached according to the above mentioned positions.

Each run started with the display of a black cross in the middle of the screen during 1 second. Then, the expression production was evoked by showing the name of the expression during 3 seconds. Then, a grey bar was displayed during 3 seconds. The subject was instructed to relax during this phase. Following this procedure, the subject performed the 6 following expressions: Surprise, Happiness, Anger, Disgust, Sadness and Neutral. Each expression was performed 3 times by the subject.

In addition to the EMG recordings, a frontal view video was recording at 25 Hz. For synchronization reason, a pure tone (400 Hz) of 500 ms duration was played at the beginning and the end of the expression phase.

EMG Signal processing

Each facial expression was automatically epoched during the recordings. A baseline correction was applied to each epoch using the first 300 ms of signal. Note that the literature suggests that a subject takes about 500 ms to mimic the facial expression seen in static images (Dimberg and Thunberg 1998). Moreover, visual inspection of the signal confirmed this hypothesis for our data and the reliability to use this period of signal to compute the baseline. Then, the absolute value of the signal was computed for every sensor. A low-pass filtered (Butterworth, 6th order, 10 Hz) was applied to the absolute values previously computed. This leads to a pattern of activation of each sensor. These characteristic parameters associated with each expression were collected and simple Gaussian models were estimated for each expression. The subject was instructed to maintain the expression during 3 seconds but we only selected the features values comprised between 0.5 s and 1.5 s after the beginning of the performance to build the Gaussian models. The a posteriori probability for each frame to belong to each of the 6 expression models can then be estimated at any time.

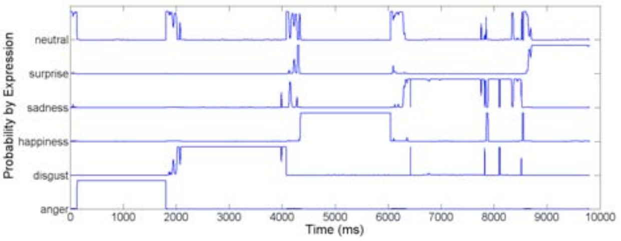

Figure 2: Time course of the probability functions over the execution of a series of expressions by the subject: anger, disgust, happiness, sadness, and surprise

RESULTS

The recognition rate calculated over all the expressions is 92.19%. This is quite high given the simplicity of the model. We exhibit in Figure 2 an example of the time course of these probability functions over the execution of a series of expressions by the subject. The probability of the target expression is stable during all the performance as we can note for the expression “anger”. Some errors are made by the classifier at the beginning and the end of each performance and mainly mislabeled as “neutral” expression.

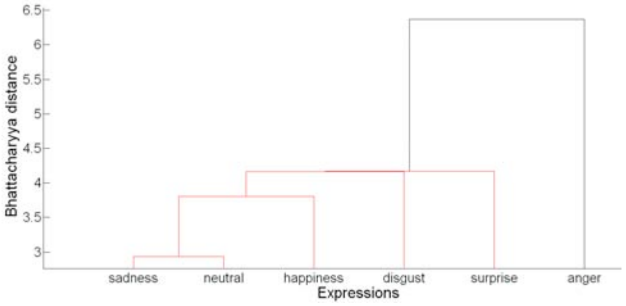

Figure 3: Hierarchical cluster tree computed for every expression

We computed the Bhattacharyya distance between the Gaussian models to measure the similarity of the probability distributions of the 6 expressions. This distance is normally used to measure the separability of classes in classification. A dendrogram computed from these distances can be seen in Figure 3. This dendrogram is a hierarchical cluster tree showing the similarities between the distributions of the expressions. It exhibits that the classes “sadness” and “neutral” are the closest one. Therefore, errors in classification between these 2 classes are expected. This is verified in the example of time course in Figure 2: “sadness” is the expression which probability is the less stable during the performance with several mislabels as “neutral”. On the other side, “sadness” and “anger” are the 2 more dissimilar distributions and few errors are expected between theses 2 classes. This is consistent with the previous remark about the stability of the “anger” expression.

APPLICATION

This recognition system is very simple and light in terms of computing. The only delay between the EMG recordings and the response delivered by the recognition system is due to the low-pass filter performed on the absolute values. That means this system is able to deliver a recognition response in real-time i.e. every 1.66 ms whereas video-based systems could deliver it every 16.66 ms (frame rate at 60 Hz) in the best case. It may be suitable to control the expressions module of a 3D avatar in virtual environments for example.

We developed a prototype of one possible application of this system for HCI. We built an online paradigm where the expression recognized is sent to a 3D avatar which mimics this expression. Examples of 2 expressions recognized and mimicked by the avatar are displayed on Figure 4. We integrated the responses from the recognizer over time window of 500 ms and then sent it to the expressions module of the avatar. This solution was used to improve the accuracy and also to avoid too much jittering faces if mislabelling occurred. Moreover, most of 3D avatars do not have a biomechanical model and are driven by geometric or articulatory parameters and so rely on a specific module to generate expressions.

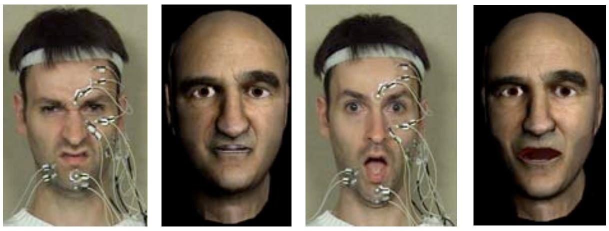

Figure 4: A 3D avatar mimicking the expressions performed by the subject (Left, “anger”; Right, “surprise”)

FUTURE DIRECTIONS

The system we developed reaches good recognition rate but several improvements are possible. The next step in the analysis will be to compare different kind of classifiers such as Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM) used in video-based recognizers and in the Brain-Computer Interface community. We can expect improvements in the accuracy with more complex classifiers. Another way to improve the classification rate will be to test other features as described in (Reaz, Hussain et al. 2006). Wavelets coefficients and autoregressive models should provide additional information useful for the classifiers. Of course, these methods will be run on more subjects to test their consistency. Similar results are expected with the same electrode placements. The stability of the results with fewer sensors is another issue as a real implementation of the system would need a light system.

Another interesting analysis will be to compare the recognition rate and the speed in delivering it of video-based recognizers on our data. Indeed, we recorded synchronously EMG signals and a front view of the subject. We could expect difference in terms of accuracy but also in terms of the size of the time window necessary to reach these accuracies.

Our work focused on the performance of expressions only. We could expect for an HCI application that the user would concurrently speak and use gestures to express emotions. A future study will be focused on detecting the affective states of the user using multimodal data (EMG, Video and Acoustic) while users are interacting with an application. (Busso, Deng et al. 2004) proposed a system based on bimodal data: video and acoustic to recognize the expression of the user. More recently, facial surface EMG sensors have also been successfully used to build continuous speech recognition systems (Wand and Schultz 2009) and speech synthesis systems from voiceless EMG signals (Toth, Wand et al. 2009).

In the proposed application, i.e. changing the expression of an avatar to mimic the expression of the user, we could imagine to make the face shape of the avatar vary directly from the values of the features. In the robotics community, (Tingfan, Butko et al. 2009) developed a robot able to learn to perform expressions using a video-based recognizer as feedback. (Minoru, Chisaki et al. 2006) used EMG signals to control the robotic muscles to make natural facial expressions. In the 3D animation community, (Lucero and Munhall 1999) developed a face model composed by soft tissue with multilayer deformable mesh animated from EMG recorded activity. Their aim was to produce speech. It would be interesting to extend this kind of biomechanical models to generate facial expressions directly from the EMG signals or features extracted from it as initiated by (Morishima 2001).

DISCUSSION

In the research area of HCI, building affective applications that can react on demand to the human user is a very challenging issue. For example, an application could use the expression of the subject to adapt the level of difficulty to the emotional status. Another interesting application could be the integration of expressions into 3 dimensional online virtual environments. These kinds of application rely on the accuracy of the inputs processing. To determine the emotional state of the user and adapt the level of difficulty for example, the traditional input is a video of the face of the user. Video-based expression recognition systems are non-invasive and suitable in certain application areas and certain conditions but are quickly overpassed by human behaviour (moving too fast, out of the camera field, etc.) (Krell, Niese et al. 2009). EMG-based expression recognizers are an alternative for certain kinds of HCI such as in games. They are more invasive than the video-based as the user agrees to wearing sensors on his face. On the other hand, the user has complete freedom of movements and can really interact in an ecologically valid way with the interface without any constraint (in a virtual reality environment for example). Currently, new devices have been developed using wireless and small dry electrodes. The electrodes are integrated into devices such as headphones for MindSet from Neurosky (www.neurosky.com) or helmets for Epoc from Emotiv (www.emotiv.com). The use of such devices will reduce the duration of the preparation of the user which is actually the main limitation of EMG-based recognition systems.

CONCLUSIONS

The results presented in this research show that it is feasible to recognize human expressions with high accuracy by the use of facial EMG sensors. Therefore, the next generation of HCI might be able to use this information as a feedback of the user and adapt the response to improve the performance and the engagement of the user.

ACKNOWLEDGMENTS

This work was supported by the Thinking Head project a *Special Initiative* scheme of the Australian Research Council (ARC) and the National Health and Medical Research Council (NH&MRC) (Burnham, Dale et al. 2006-2011).

REFERENCES

Ang, L. B. P., E. F. Belen, et al. (2004). Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors. TENCON 2004. 2004 IEEE Region 10 Conference.

Bartlett, M., G. Littlewort, et al. (2006). Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia **1**(6): 22-35.

Becker, K. (2003). Varioport™. http://www.becker-meditec.de.

Burnham, D., R. Dale, et al. (2006-2011). From Talking Heads to Thinking Heads: A Research Platform for Human Communication Science. from http://thinkinghead.uws.edu.au/index.html.

Busso, C., Z. Deng, et al. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. Sixth International Conference on Multimodal Interfaces ICMI, State College, PA.

Chin, Z. Y., K. K. Ang, et al. (2008). Multiclass voluntary facial expression classification based on Filter Bank Common Spatial Pattern. Engineering in Medicine and Biology Society, EMBS 2008. 30th Annual International Conference of the IEEE.

Dimberg, U. and M. Thunberg (1998). Rapid facial reactions to emotional facial expressions. Scandinavian Journal of Psychology **39**(1): 39-45.

Ekman, P. and W. V. Friesen (1971). Constants across Cultures in the Face ans Emotion. Journal of Personality and Social Psychology **17**(2): 124-129.

Fridlund, A. J. and J. T. Cacioppo (1986). Guidelines for Human Electromyographic Research. Psychophysiology **23**(5): 567-589.

Krell, G., R. Niese, et al. (2009). Facial Expression Recognition with Multi-channel Deconvolution. Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on.

Lucero, J. C. and K. G. Munhall (1999). A model of facial biomechanics for speech production. Journal of the Acoustical Society of America **106**(5): 2834-2842.

Minoru, H., Y. Chisaki, et al. (2006). Development and Control of a Face Robot Imitating Human Muscular Structures. Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on.

Morishima, S. (2001). Face analysis and synthesis. Signal Processing Magazine, IEEE **18**(3): 26-34.

Reaz, M. B. I., M. S. Hussain, et al. (2006). "Techniques of EMG signal analysis: detection, processing, classification and applications." Biological Procedures Online: 11-35.

Tian, Y. I., T. Kanade, et al. (2001). Recognizing action units for facial expression analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on **23**(2): 97-115.

Tingfan, W., N. J. Butko, et al. (2009). Learning to Make Facial Expressions. Development and Learning, 2009. ICDL 2009. IEEE 8th International Conference on.

Toth, A., M. Wand, et al. (2009). Synthesizing Speech from Electromyography using Voice Transformation Techniques. Interspeech.

Wand, M. and T. Schultz (2009). Towards Speaker-Adaptive Speech Recognition Based on Surface Electromyography. Biosignals.