Assessment of the Emotional States of Students during e-Learning

Leon Rothkrantz, Dragos Datcu, Iulia Chiriacescu, Alin Chițu

Abstract: *Emotions are assumed to have a great impact on our behaviour and also on our learning behaviour. In a face-to-face learning environment emotions can be expressed by (non-)verbal behaviour as way of speaking, facial expressions or words with an emotional loadings. In this paper we research the possibility to assess the emotional state of learners by analysis of nonverbal behaviour as speech analysis and analysis of facial expressions. We developed tools to extract features from sound and video recordings and used classifiers as SVM to label emotional states. We used discrete emotional states but also the well-known 2D valence and arousal score as a continuous score of the emotional state. From our experiments it proves that students show overt emotions under special conditions with strong emotional triggers. Our system was able to assess strong emotions up to some level.*

Key words: *Emotions, nonverbal behaviour, speech analysis, facial expressions analysis, e-learning.*

INTRODUCTION

It is commonly agreed that emotions have a strong impact on our behaviour. Students with well developed abilities and trained skills show the expected behaviour if they get motivated by internal or external emotional triggers. Strong negative emotions as fear and anxiety can block the learning behaviour. Happiness has a positive effect on the learning behaviour. But emotions and their impact on the e-learning behaviour are not well understood and a lot of research is needed. Some students miss the social support and interaction in e-learning environment. A strong motivation, discipline and time scheduled learning is needed to survive in a distant learning environment. For that reason the Open University invested a lot of effort in setting up networks and communities of students that are remote in space and time.

In the past there was a strong impact on the cognitive aspects in the learning theories. Emotions were neglected or researched independently from cognition. Only recently researchers studied the impact of cognitive and emotional aspect on learning. Goleman [8] introduced the concept of emotional intelligence similar to the classical cognitive intelligence. Researchers as Darwin [4], Damasio [5], Averill [1], LeDoux [11] contributed a lot to the understanding of emotions and their impact on learning.

In a face to face learning environment emotions can be assessed from the nonverbal behaviour of students. In distant learning without or with digital teacher the assessment of emotions is more complex. A common way is to use questionnaires. But questionnaires can not be used for continuous real time assessment of emotions. That is the reason we developed a surveillance system using cameras and microphones to record nonverbal behaviour of students and analysed those recordings with our EmoRec system.

After Darwin, Ekman [7] spent a lot of research on the automatic recognition of facial expressions. He claimed that 6 emotions are universal namely happiness, sadness, disgust, anger, fear and surprise. Many systems have been developed to recognise the six basic emotions in speech and recordings of facial expressions. But in daily life most emotions are not pure but blended and of varying intensity. And even the six basic emotions can be displayed in many ways. This can also be expected in an e-learning environment.

Most emotion recognition systems are based on classifiers. To train a classifier a huge database of emotional expressions is needed. Recording and annotating a database of emotional expressions is very time consuming. Most available databases are acted emotions by trained people and not recordings of spontaneous emotions. In case of recognition of facial expressions unsolved problems are varying lighting conditions, changing posture and occlusion. In case of speech, multiple speakers, background noise and reverberation are difficult to handle problems.

The outline of this paper is a follows. In the next section we will report about related work. In section three we will introduce the technologies used to recognize emotions in sound and video recordings. Next we report about our experiments and we will end with a section conclusion and future work.

RELATED WORK

In [14] O'Regan reports studies of online learning in which emotion plays an important role. The Web offers the perfect technology and environment for personalized learning. Martinez [12] studied individual learning on the Web with a focus on emotions. At eduweb [http://www.eduweb.com], the goal of the designers is to develop the most engaging and effective online learning experiences possible. To this end, we engage in research to better understand learning theory, learner preferences and engagement, and educational outcomes. Schaller and his colleagues report about the negative emotions students experience when they have to navigate the first time through learning sites. In [19] Wegerif centres on the sense of isolation that online study may engender among learners, a factor often ignored by many educators, but one that ignored by many educators, but one that may make the difference between a successful and an unsuccessful online learning environment for many students. The importance of a proper appreciation of the learners' social context is stressed, as is the concept of the 'virtual self' that individual learners may choose to portray during online communication. Ng [13] reported about online learners showing fear during electronic communication. Students educated with Twitter, Weblogs and Facebook probably require social and communicating abilities to handle his negative emotions. Hara and Kling [9] studied the frustration on line learners experience with badly designed or non-functioning online learning environments. In [15] Rothkrantz introduced e-learning in virtual environments. A virtual University was designed in Second Life. The focus was on the design of emotion in the social interactions of students represented as virtual characters (Avatars).

FACIAL EXPRESSION RECOGNITION

In the case of video data processing, we have developed an automatic system for the recognition of facial expressions for both still pictures and video sequences. The recognition was done by using Viola&Jones features and boosting techniques for face detection [18], Active Appearance Model – AAM for the extraction of face shape and Support Vector Machines –SVM for the classification of feature patterns in one of the prototypic facial expressions. For training and testing the systems we have used Cohn-Kanade database [10] by creating a subset of relevant data for each facial expression.

The Active Appearance Model – AAM [3] makes sure the shapes of the face and of the facial features are correctly extracted from each detected face. Starting with the samples we have collected from the Cohn- Kanade database, we have determined the average face shape and texture (Figure 1).

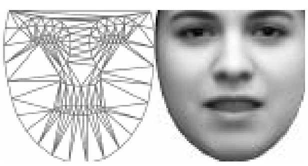

Fig.1. The mean face shape (left) and the mean face texture aligned to the mean shape (right).

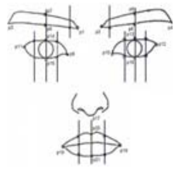

Fig.2. The Facial Characteristic Point FCP model.

Based on the AAM face shape, the recognition algorithm generates during the emotion classification stage a set of features to be used such as distances computed between specific Facial Characteristic Points – FCPs as shown in Figure 2. For the recognition of expressions in still pictures, the distances determined from one face form a representative set of features to reflect the emotion at a certain moment of time. In the case of recognition of facial expressions in video sequences, the features are determined as the variation of the same distances between FCPs as observed during several consecutive frames.

Our algorithm for the recognition of emotions in videos implies an initial processing of the multimodal data. Firstly, the audio-video input data is rescaled by conversion to a specific frame-rate (Figure 3). This process may imply downscaling by skipping some video and audio frames. Secondly, the audio data is processed in order to determine the silence and non-silence segments. The resulting segments are correlated to the corresponding audio data and constitute the major data for the analysis.

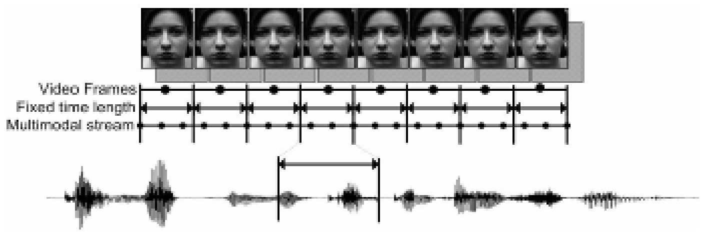

Fig.3. Multimodal frame rescaling algorithm.

In the case of facial expression recognition, within each segment an overlapping sliding window (Figure 3) groups together adjacent video frames. Based on the set of video frames, the recognition of facial expressions determines the most probable facial expression using a voting algorithm and a classifier trained on still pictures. For the video oriented classifier, the most probable facial expression is determined by taking into account the variation of the features extracted from all the video frames in the group. The identification of silence and non-silence segments is realized by using both acoustic and video information. The recognition of emotions is realized differently for silence segments and non-silence segments, namely for silence segments only the video stream is used. For the non-silence segments, the emotion recognition is based on the multimodal semantic fusion of the results of the emotion classification on single modalities. The input features in this case relate to only FCPs from the upper part of the face. The reason for not considering the FCPs of the mouth is explained by the natural influence of the phoneme generation on the mouth shape during the process of speaking.

EMOTION RECOGNITION FROM SPEECH

In the case of emotion recognition from speech, the analysis is handled separately for different number of frames per speech segment. In the current approach there are five types of split methods applied on the initial audio data. Each type of split produces a number of data sets, according to all the frame combinations in one segment. The data set used for emotion analysis from speech is BerlinDB – a database of German emotional speech. The database contains utterances of both male and female speakers, two sentences pro speaker. The emotions were acted by ten native German professional actors (five female and five male). The result consists of ten utterances (five short and five long sentences). The length of the utterance samples ranges from 1.2255 seconds to 8.9782 seconds. The recording sample rate is 16kHz. The final speech data set contains the utterances for which the associated emotional class was recognized by at least 80% of the listeners. Following a speech sample selection, an initial data set was generated comprising 456 samples and six basic emotions (anger: 127 samples, boredom: 81 samples, disgust: 46 samples, anxiety/fear: 69 samples, happiness: 71 samples and sadness: 62 samples). The Praat [2] tool was used for extracting the features from each sample from all generated data sets. According to each data set frame configuration, the parameters mean, standard deviation, minimum and maximum of the following acoustic features were computed: Fundamental frequency (pitch), Intensity, F1, F2, F3, F4 and Bandwidth. All these parameters form the input for separate GentleBoost classifiers according to data sets with distinct segmentation characteristics. The GentleBoost strong classifier is trained for a maximum number of 200 stages. Separate data sets containing male, female and both male and female utterances are considered for training and testing the classifier models.

In Figure 4 we display the architecture of a system that is able to recognize emotions given a speech input in a real time setting.

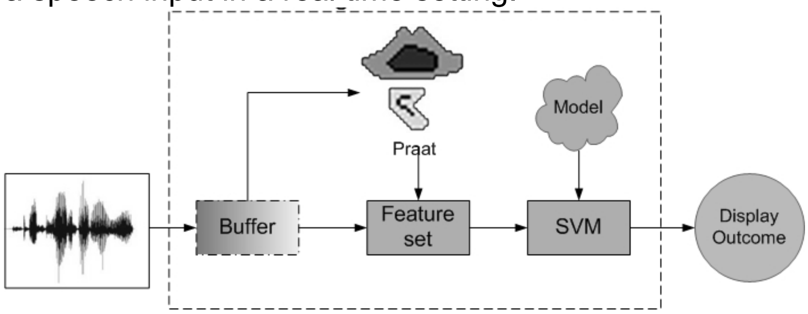

Fig.4. Architecture of the real time emotion recogniser from speech.

EXPERIMENTS

We performed some experiments in an e-learning environment. Twelve MSc-students (8 male, 4 female) in a student lab were connected to the Web via a PC equipped with Webcam and Microphone. The goal was to assess emotions in a multimodal way and to research what triggers those emotions. After the experiments the students were requested to fill in a questionnaire to evaluate the user experiences.

Experiment 1

Design: In the first experiment students took part in a Web-based Course on neural networks. Students were supposed to read some text, answer questions and do some exercises.

Results: Students showed some emotional expressions but we were not able to classify them as one of the basic emotions or as a point in the valence –arousal space. Some facial expressions were either nerve-tics, non-verbal utterances etc. Maybe because students were aware of the fact that their behaviour was recorded they controlled their emotional behaviour.

Experiment 2

Design: The design was similar to experiment 1. But there were many interruptions, such as network timed-out, errors, failures in saving and communicating answers.

Results: We observed different emotional behaviours. Some students showed frustration on their face, use dirty words, and even give up after some time and logged off for a coffee break. Other students showed sarcastic smiling and made funny remarks about the system, administrators and lecturers. All students started interaction to see if other students suffer from the same problems and were looking for social support.

Experiment 3

Design: In the third experiment students in subgroups of three students have to play a game with a race car simulator (TORC). There were three kinds of tracks: one straight line, a curvy track and a track with a lot of billboards with funny pictures or pictures of traffic accidents. The car drivers have to pass the tracks as soon as possible, Car crashes, violating the traffic rules results in penalties. Every member of the group has to drive, every player gets a score and the goal was to reach the high score as a team. The blink rate of the player was assessed by a sensor. The assignment was to find the correlation between the eye blink rate and the track condition.

Fig.5. TORC simulator.

Results: As expected a lot of emotions were shown by the players. The time pressures induced a lot of stress. But most stress was generated by the team mates, pushing the driver player to drive as hard as possible. Every mistake resulted in a lot negative comment of the team members, but also in some positive feedback. Some negative comments resulted in angriness of the drivers. The funny pictures on the billboards evoked smiling faces and laughing, the accidents fear or disgust. But the next rounds the effect of the billboards was minimal, probably the drivers concentrated only on the driving task (tunnel view). Because of the close to mouth microphone and because the driver was focussed on the screen (frontal pictures), we were able to process the video/audio streams.

CONCLUSIONS AND FUTURE WORK

Our conclusion is that emotions play an important role in e-learning environment. But it depends on the educational material, environment and conditions are clearly displayed by students by facial expressions or their way of speaking. Individual learners in traditional face-to-face learning or e-learning conditions show no emotions unless they are triggered by teachers/students or features from the environment or learning material. Especially serious gaming is able to evoke emotional reactions from students.

From the questionnaires we learned that students appreciated e-learning environments, generating emotions. They reported that they were more motivated to take part in the lessons and had positive feelings afterwards. The question is of course of the results of learning have a positive correlation with the amount of emotions.

One of the issues is if the shown emotions are acted or spontaneous. At this moment we repeat experiment three and record EEG signals using a brain cap. It is difficult to fake the brain processes, so we expect to measure the true emotional conditions of the students.

We agree with conclusion of many researchers involved in e-learning that the main challenge is not to imitate face-to-face learning but to go beyond that. Using serious gaming environments enables teachers to generate educational environments which cannot be realised in real life.

REFERENCES

[1] Averill, J *A connectionists view of emotion.* Emotion: Theory, Research and Experience, Vol. 1, *Theories of Emotions,* Academic Press, New York, pp 305-337, 1980.

[2] Boersma, P., Weenink, D., Praat: doing phonetics by computer (version 4.3.14) [Computer Program], 2005.

[3] Cootes, T. F., G. J. Edwards, C. J. Taylor, "Active appearance models", Lecture Notes in Computer Science, vol. 1407, pp. 484–98, 1998.

[4] Darwin, C., *The expressions of emotions in Man and Animals,* The University of Chicago Press, Chicago, 1872/1965.

[5] Damasio, A., *Descartes's Error: Emotion, Reason, and the Human brain*, Papermac, London, 1996.

[6] Datcu, D., L. J. M. Rothkrantz, "The recognition of emotions from speech using GentleBoost Classifier", CompSysTech'06, June 2006.

[7] Ekman, P., *The nature of emotion: Fundamental questions*. Oxford University Press, pp 7-19, 1994.

[8] Goleman, D., *Emotional Intelligence*, Bantam Books, New York, 1995.

[9] Hara,N. , R. Kling, Students' distress with a Web-based distance education course: An ethnographic study of participants' experiences, Center for Social Informatics, Indiana University, 2000.

[10] Kanade, T., J. F. Cohn, Y. Tian, Comprehensive database for facial expression analysis, *Proc. of the 4th IEEE Int. Con. On Automatic Face and Gestures Recognition*., France, 2000.

[11] LeDoux, J., *The emotional Brain: The mysterious Underpinnings of Emotional Life*, Phoenix, London, 1999.

[12] Martinez,M., Key design considerations for personalized learning on the Web, *Educational Technology and Society*, Vol. 4, No.1 pp 26-40, 2001.

[13] Ng, K-C., Using e-mail to foster collaboration in distance education. *Open learning*, Vol. 16, No. 2, pp 191-200, 2001.

[14] O'Regan, K. Emotion And E-Learning. *JALN* Volume 7, Issue 3, 2003.

[15] Rothkrantz, L.J.M. E-learning in virtual environments, *Communication & Cognition*, Vol 42, No. 1&2, pp 37-52, 2009.

[16] Vapnik, V.,*The Nature of Statistical Learning Theory*, Springer- Verlag, New York, 1995.

[17] Vapnik, V., *Statistical Learning Theory*, John Wiley and Sons, Inc., New York, 1998.

[18] Viola, P., M. Jones, Robust Real-time Object Detection. Second International *Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling*, 2001.

[19] Wegerif, R., The social dimension of asynchronous learning networks. *Journal of Asynchronous Learning Networks*, Vol. 2, No. 1, 1998.

ABOUT THE AUTHORS

Prof. Leon Rothkrantz, drs. dr., Department of man machine Interaction, Delft University of Technology, Phone: +332787504, E-Mail L.J.M.Rothkrantz@tudelft.nl

Netherlands Defence Academy, Faculty Military Sciences, Den Helder, The Netherlands.

Dragos Datcu, MSc., Department of man machine Interaction, Delft University of Technology, Phone: +332787504, E-Mail D.Datcu@tudelft.nl

Netherlands Defence Academy, Faculty Military Sciences, Den Helder, The Netherlands.

Iulia Chiriacescu, MSc., Department of man machine Interaction, Delft University of Technology, Phone: +332787504, E-Mail I.Chiriacescu@tudelft.nl

Netherlands Defence Academy, Faculty Military Sciences, Den Helder, The Netherlands.

Alin G. Chitu MSc., Department of man machine Interaction, Delft University of Technology, Phone: +332787504, E-Mail A.G.Chitu@tudelft.nl