



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Experimental Assessment of an Emotion Tracking Software Agent (ETA) for Assisting Communicative Interactions of Multitasking Users in Groupware
Author(s)	Smith, Paul; Redfern, Sam
Publication Date	2011-08-22
Publication Information	Smith, Paul, & Redfern, Sam. (2011). Experimental assessment of an Emotion Tracking Software Agent (ETA) for assisting communicative interactions of multitasking users in groupware GAME'ON International Conference on Intelligent Games and Simulation. Paper presented at the AME'ON International Conference on Intelligent Games and Simulation, Galway, Ireland, 22-24 August.
Publisher	Eurosis
Link to publisher's version	<a href="http://www.eurosis.org/cms/?q=taxonomy/term/4">http://www.eurosis.org/cms/?q=taxonomy/term/4</a>
Item record	<a href="http://hdl.handle.net/10379/4081">http://hdl.handle.net/10379/4081</a>

Downloaded 2024-04-23T10:04:51Z

Some rights reserved. For more information, please see the item record link above.



# Experimental Assessment of an Emotion Tracking Software Agent (ETA) for assisting Communicative Interactions of Multitasking Users in Groupware

Paul Smith and Sam Redfern  
Discipline of Information Technology  
National University of Ireland, Galway  
Galway,  
Ireland.

E-mail: dr.paul.t.smith@gmail.com

E-mail: sam.redfern@nuigalway.ie

## KEYWORDS

CVEs, Automatic Facial Expression Recognition, Serious Games, Virtual Environments, Non-Verbal Communication, CSCW, Neural Networks, Groupware

## ABSTRACT

Interactions such as discussion and negotiation in face-to-face work contexts strongly rely on non-verbal feedback. Such feedback provides indications of important negotiation states such as agreement or disagreement and understanding or confusion. The increasing popularity of groupware and its use by virtual teams for collaborative remote work necessitates the development of appropriate tools to manage the reality of distributed and remote work. Such remote collaboration is often hampered by a lack of social cohesion and such phenomena as participant multi-tasking. This paper outlines the experimental assessment of a proof of concept AI based software agent (Emotion Tracking Agent or ETA) for the real-time tracking of groupware user's facial expressions of emotion during virtual meetings. The software agent is designed as a novel approach to the removal of negative or unwanted effects of user multitasking and attention distracting behaviours in virtual collaboration and meeting environments.

## INTRODUCTION

Groupware for remote collaboration has become increasingly important in recent years due to the growing number of businesses using distributed workers based in different locations around the world (Lojeski, Reilly and Dominick, 2007). Studies have established that companies such as Intel (70%), IBM (40%) and Sun Micro Systems (nearly 50%) already have high percentages of virtual and distributed workers (Lojeski, Reilly and Dominick, 2007). These workers can be based at home or abroad or may travel frequently making remote collaboration software the most convenient and economical means for these individuals to communicate with colleagues and clients.

In virtual and face-to-face meetings, one of the inherent problems described in the literature relates to the negative impact of multitasking in both traditional and virtual work environments (Appelbaum, Marchionni and Fernandez, 2008). Successful management of worker multitasking and

the provision of appropriate solutions to specific problems caused by multitasking or inattention is of key importance to future work practices (Black and Hearne, 2008).

This paper briefly discusses the architecture and use of our prototype Emotion Tracking Software Agent (ETA) in tackling the negative impact of multitasking or task switching behaviours of users in virtual collaboration environments. The ETA is also used as a means to augment a user's ability to communicate his/her non-verbal displays of emotional information to fellow collaborators, enriching their communication in environments which traditionally tend to provide poor support for non-verbal communication (Ekman and Friesen, 1975).

The research area of serious games is a diverse field, encapsulating any usage of game technology for purposes other than its primary entertainment use. Our developed ETA is designed for integration within many forms of virtual interaction environments, many of which are based on existing game technology and graphic engines. The environment which our ETA is integrated for testing is itself created with the Torque Game Engine developed by Garage Games. Further discussion of our agent and its usage in serious games application is described in (Smith and Redfern, 2010).

In the later part of this paper we describe a number of tests designed to assess the ETA's usability and accuracy in recognising user's facial expressions. An implementation study is then described where a number of subjects were tested in a communicative interaction in a Collaborative Virtual Environment (VRCGroups) in which the ETA was integrated, in order to provide feedback on the effect of inclusion of the agent in a groupware environment.

## ETA ARCHITECTURE

The software agent was developed on a Windows operating system and the initial code for the neural network used in the project was taken from (Chopra, 2009) and is open source, written in C++ which has been extended and edited for our purposes. The completed neural network consists of three separate layers: the input layer which receives numerical inputs, a single hidden layer and a final output layer in which a single emotion is identified as the dominant among Paul

Ekman's universal prototypic emotions (Ekman, 1980) (see Figure 1). As illustrated in Figure 1 there are 15 nodes in the input layer representing the angles computed using an input representation scheme, 15 nodes in the hidden layer and 6 nodes in the output layer (one for each of the prototypic emotions proposed by Ekman). The input representation scheme was chosen after a series of experiments were carried out using the marker coordinates recorded from the user's face and an optimal scheme was determined through the use of cluster analysis which took the positions of the markers and computed 15 specific angles between them (see Figure 2).

The finished neural network was trained using 40 instances of each emotion (270 in all) and 10 instances were recorded separately for each emotion and used as validation sets for early stopping and prevention of over-fitting.

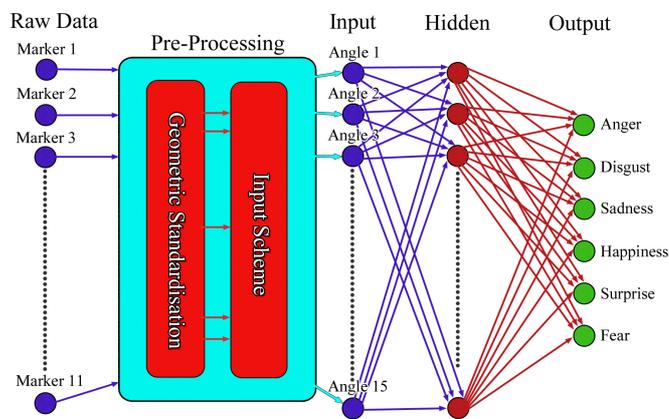


Figure. 1. The basic structure of the ETA neural network, showing how raw marker coordinate data is processed through the network.

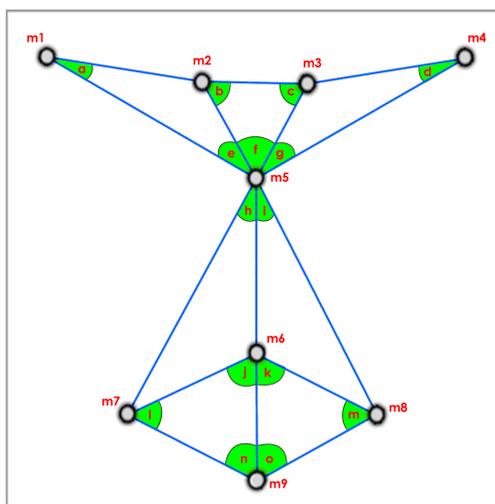


Figure. 2. Illustration of 15 angles used in current analysis for the Angle Input Representation Scheme. Angles are labelled in order from a to o.

The data which is sent to our emotion recognition neural network for identification is recorded in real-time using an Optitrack FLEX:C120 optical motion capture camera (see Figure 3). This camera is an integrated image capture and processing unit which uses a B&W CMOS imager to

captures 120 frames of video per second and an onboard image processor which transfers marker data over standard USB to a computer for display and post processing. The camera then preprocesses the record image frame, removing light from non reflective surface which allows only the reflective markers to be detected in the correct lighting conditions.

Eleven tracking markers are placed on the user's face at specific position to facilitate feature tracking and the coordinates of these markers are computed by the camera and sent to the ETA for categorization. Figure 4 shows the markers placed on a user's face in their correct positions (Zhang, 1999). More information on the neural network behind our ETA is available in (Smith, and Redfern, 2010a)

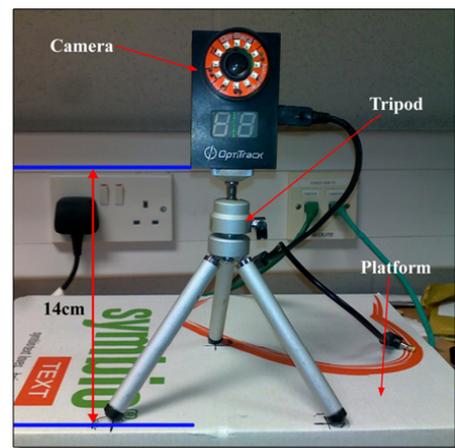


Figure. 3. Our chosen tracking solution, the Optitrack FLEX: C120 developed by Naturalpoint mounted on a standard tripod.

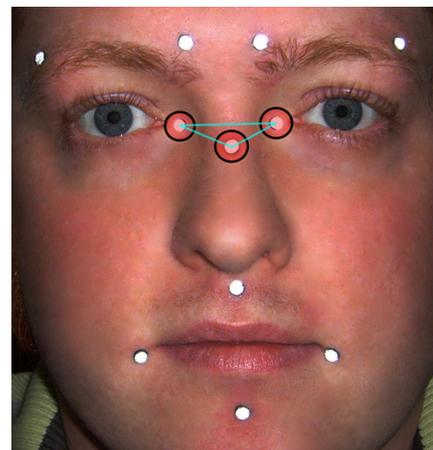


Figure. 4. Tracking markers placed in specific positions on user's face to facilitate feature extraction.

The ETA is developed in C++, and incorporates a Multi-Layer Perceptron (MLP), trained using back propagation (Stergiou and Signanos, 1996). Data is received from the Optitrack camera via the SDK. The MLP with backpropagation was chosen due to its prevalence in the facial tracking literature, and due to the flexibility it afforded us in allowing users to record their own specific emotions and use them as training for the network (Pantic and Rothkrantz, 1999).

For the sake of experimental assessment of the prototype ETA, we integrated the agent within a Collaborative Virtual Environment (CVE) known as VRCGroups (developed previously by our research group) (Redfern et al, 2006). Inside this CVE, the agent tracks the user's facial expressions at all times and prompts participants of the emotional changes of fellow collaborators inhabiting the same environment.

## **ETA INFORMATION MODES**

Since the aim of the software agent is to increase users' emotional awareness and attention in remote meeting or groupware environments, in VRCGroups when ETA functionality is enabled, the users are prompted with information relating to other participant's emotional states in three separate communication modes. These modes are vocal (a recorded message prompts the user of fellow collaborator's emotional states), textual (a chat message is sent to the user informing them of the emotion change of a fellow environment user) and visual (an emoticon icon appears over the head of the user whose emotional state has changed allowing easy identification of the user's avatar). The user also has the ability to specify which co-workers they are interested in receiving information about, from the ETA to ensure that they are not bombarded with unwanted information which may detract from the interaction.

## **EXPERIMENTAL ASSESSMENT OF ETA**

The neural network which forms the heart of the emotion recognition system for our ETA was evaluated in a number of experimental tests designed to gauge its ability to recognise each of the 6 primary emotions. The network's abilities were also compared to the emotion recognition capacity of a group of human participants tested on videos of the same facial expressions as the network, in order to ascertain that it operated with at least the abilities of a human observer. This also ensured its performance was adequate for its real-time usage requirement, as a substitute for a user's eyes, in keeping track of the non-verbal communication of other users outside his/her field of view. The final testing of the software agent took the form of an implementation study where the ETA was integrated in the VRCGroups environment in order to evaluate its contribution to reducing the negative effects of multitasking among users in such groupware.

### **ETA Accuracy Experiments**

In order to determine if the ETA was as (or more) effective at classifying emotion from facial movements than human subjects, two tests were carried out. Firstly, we sought to discover how well a human subject could correctly interpret an emotion from the facial expression of another human being recorded displaying a number of expressions on video. This test entailed the simultaneous recording of a specific user's facial movements using the FLEX:C120 camera and a digital video camera. The experiment sought to evaluate the recognition ability of the ETA neural network in comparison with human subjects. The videos recorded using the digital

camera were displayed to a group of 15 subjects instructed to identify which of the 6 primary emotions (the 6 prototypic emotions proposed by Ekman were Anger, Disgust, Sadness, Happiness, Fear and Surprise) each video portrayed. The data recorded by the motion tracking camera (2D coordinates of facial markers) was sent as input to the ETA neural network in the same order as the videos were given to the experiment subjects. The results were then analysed and compared (accuracy was computed using percentage of correctly identified emotions from data and video).

Analysis of the results for the first stage of this test, which involved human subjects identifying emotions depicted in videos indicated that a high percentage of the test participants had a strong ability for emotional recognition. A 78% recognition rate was recorded during this test which was in line with previous tests carried out in the literature such as the high average recognition accuracy of 89% published in (Susskind et al, 2007) and the relatively low accuracy of 69% published in (Zucker et al, 2008). The results appeared to be in the midpoint between both these previous studies which is verified further by a Chi Squared analysis of the results.

Analysis of the results of the second stage of this experiment, which involved the testing of the ETA neural network on the same data used in stage 1 to test human subjects, indicated that the network is strongest in the recognition of fear, surprise, happiness and sadness. The neural network displayed problems mainly with data sets representing anger and to a far lesser degree disgust. The overall recognition rate recorded during the test was 96% (that is 96% of all 270 facial expressions were correctly identified), indicating that the neural network was superior to the performance of the human subjects in stage 1. This result was again verified statistically using a Chi Squared analysis.

### **Implementation Study**

After accuracy of the ETA neural network was determined, an implementation study was then conducted with the objectives of illustrating the use of the ETA agent in its intended role and performing an evaluation of the completed prototype agent. The study involved evaluation of the provided benefits of the ETA software agent to a multitasking user's understanding of information presented to him/her in the course of the study and his/her perceived effectiveness of the agent in enhancing emotional awareness and attention in the virtual test environment. The study involved the integration of the agent into VRCGroups. VRCGroups is a fully functional LAN based collaborative interaction environment which provides text and voice chat facilities along with application and work sharing tools for user collaboration.

By analysing the experiences of a number of individuals in the VRCGroups environment, both with and without the use of the ETA, we intended to measure their emotional awareness under a multitasking remote meeting situation and evaluate the impact of the agent on their communication abilities in such environments.

The study comprised of two stages, involving a number of participants engaging in a remote meeting requiring them to juggle multiple tasks in parallel while simultaneously paying attention to another user's presentation. The first of these stages required the participants to complete all tasks without the help of the ETA and the second repeated these tasks a month later with the ETA functionality included (a month between stages minimised the learning effects which may have skewed the results of the study). A mixed quantitative and qualitative questionnaire was used to determine a participant's emotional awareness in both stages and to determine the participant's perception of the meaning of presented messages during the study. Another survey was used at the end of the second stage to allow participants to give feedback on the perceived benefits or disadvantages instigated by the introduction of ETA functionality. A side-task in the form of a short ten question IQ-style test was also given to the participants to complete in parallel with the main emotion identification and perception survey to induce multitasking. The results of this side-task were not important but the effect it had on the participants of the study was imperative.

### Implementation Study Results

During the analysis of the study results a number of interesting observations were made regarding the effect of the ETA on subject's attention and awareness in a remote collaborative and communicative interaction. During the study the subjects were required to identify the emotions portrayed in the voice of the presenter (who relayed information to the subjects from the CVE) at set points in the course of the each stage. The subjects were also instructed to describe the meaning of what the presenter was saying at these set points in order ascertain how well they understood the underlying messages portrayed. There were 19 of these points in total during each stage of the study. In the first stage, where the ETA was not used to help participants, the overall average emotion identification accuracy observed was 60 percent, which is considerably lower than the accuracy observed in the previous experiments from facial expressions. The results from the emotion identification portion of this stage are shown in Table 1 in terms of correctly identified, logically misidentified (such as anger mistaken for disgust), entirely incorrect and totally misidentified emotions. Although the number of participants in this study is too few to make any definitive conclusions regarding the ability of humans to recognise emotion from speech, this result still implies that the level of emotion conveyed from a person's voice is considerably lower than that conveyed through facial displays.

### Emotional Awareness and Understanding Observations

It is clear from the results displayed in Table 1 that the best performing participants of the study were Subjects 4 and 5, who incidentally were both of Chinese nationality. From further analysis of their descriptions recorded during this study stage, it became apparent that although their emotion accuracy from vocal tone was superior to the remaining subjects, their understanding of the underlying meaning of

the messages was in fact inferior. This was likely due to the fact that English was not their first language which may have left them at a disadvantage.

Table 1: Compiled results for individual subjects during initial emotion identification stage of Implementation study

	Total Correct	Total Incorrect	Logically Incorrect	Incorrect Entirely
Subject 1	11	8	3	5
Subject 2	8	11	2	9
Subject 3	10	9	3	6
Subject 4	13	6	3	3
Subject 5	15	4	2	2

Comparisons made between these subject's first stage descriptions and those recorded in the second stage, show a considerable increase in understanding, suggesting that the inclusion of the ETA was of significant benefit to these two subjects. Although the small number of subjects used in this study is not enough to make definite conclusions, this result still suggests that there may be possible augmented usefulness of such an emotion tracking agent in online interactions for non native English speakers.

In relation to emotional awareness, a number of subjects remarked in interviews that once the ETA informed them of the emotion being portrayed on the presenter's face during the interaction, they were more aware of that emotion in the presenter's vocal tone and the emotions in his voice seemed clearer and more pronounced. It was also noted that the ETA made it easier for them to recognise and distinguish problem emotions such as the negative emotions of anger, disgust and fear which were identified by subjects as being harder to differentiate or detect from the presenter's vocal tone alone in stage 1 of this study. Positive emotions such as happiness and surprise were noted as easier to recognise from vocal tone than their negative counterparts.

The effect of the information offered by the ETA was evident from the descriptions given by the subjects in the second stage of this study. This was seen in the subject's use of the names of emotions prompted by the ETA in their described meanings recorded in this stage.

Analysis of participants' descriptions in both stages of this study also indicated that certain linguistic modes were unfavourably influencing and complicating the subject's interactions. The most dominant of these linguistic modes is

the manifestation of sarcasm in the vocal tones of the presenter's voice.

### **Attention and Certainty Observations**

The use of more concise and detailed language during the second stage of this study by the majority of subjects also indicated an increase in understanding of the underlying message portrayed by the presenter. The usage of words such as "might be", "seems", "maybe" or "it appears to be" are prevalent in stage 1 giving the impression of a lack of certainty and sureness in the subject's descriptions. During the second stage it was clear from the predominance of terms such as "definitely" or "does not" or "is not" or "is", that there was a clear increase in the subject's confidence and certainty in their answers. This improvement in the certainty levels of the study participants suggests a positive reaction to the inclusion of the ETA which complements the previously observed increases in the subject's understanding and emotional awareness.

It was also observed that in stage 2, subjects tended to more often reference multiple elements of a section in their answers, and appeared to be more capable of describing the presenter's attitude to these individual elements; this was not a phenomenon observed to such a degree in stage 1. An increased ability to break up the message into separate informational strings also indicated an increase in understanding of the underlying meaning which may be due to an improvement in the subject's attention to particular details of the presented information. Many times in stage 1, the descriptions given by subjects, were unspecific and used general terms such as "this statement" to refer to the entire message as a whole. Such language use implied either the participants were not paying full attention or they were unable to adequately understand the information given.

### **Multitasking Observations**

It became evident the further the implementation study progressed, that the experimental setup used may not have simulated multitasking to a sufficient degree since some subjects tended not to task switch, but rather to finish one task before starting another. This may have been due to the amount of concentration required for completion of the main task. This was apparent in the fact that some subjects noted in interview that the first stage without the use of the ETA, caused them to have to concentrate more on the presenters information task but in the second stage with the help of the ETA's prompting, they were capable of concentrating on the side task more due to the abundance of emotional information provided which made the understanding of the meaning of the presented information much easier. Subjects stated that the juggling of two tasks was much easier if one of the tasks was aided by the ETA and it was clear when the increased accuracy of the subject's descriptions in stage 2 is considered, that the overall interaction seems to have benefited significantly.

Another remark made by subjects in interview indicated that concentration was never really a large problem in either stage

but they did note that it was definitely easier to focus and concentrate in the studies second stage. Some subjects revealed that their strategy for dealing with the side-task in the first stage consisted of choosing the easiest of the given questions to perform first, in order to allow more time to concentrate on the main presentation task. This was apparently not a problem during the second stage and most subjects stated that they tended to go from the first question sequentially to the last due to the lack of need to concentrate as much on the main task to the degree required in the initial stage.

### **CONCLUSIONS**

We believe the the ETA solution has been successfully demonstrated to allow remote collaborators to more easily perform multiple tasks when one or more of these tasks involve a communicative interaction supported by the ETA. The results of this work and experiments have illustrated that the application of such an emotion tracking agent to remote collaboration tools has the capability to increase collaborating user's understanding of vocal communication and increase their attention and ability to concentrate on the ETA supported interactions while engaged in side-tasks. Increased communicative abilities between collaborating users will help bridge the gap between real world and virtual world interactions and if such an agent is developed further to a commercially available product built on commonly available hardware then it would represent one step closer to ensuring the continued usage and growth of the remote collaboration industry.

The research related to the tracking and recognition of non-verbal displays of emotion carried out during this research project has demonstrated the importance of facial expression in the portrayal of emotion and meaning in an online communicative interaction between two or more collaborating persons. It was evident from observations made during the implementation study that without the accompaniment of facial expressions during a communicative interaction, the listening parties may exhibit difficulties in extracting the full meaning of vocally communicated messages. Clear improvements to user's understanding and confidence in the correctness of extracted meaning were observed upon the introduction of the ETA prototype into the communicative process. During the implementation study it was also observed that the non-verbal feedback provided by the ETA adds to a user's ability to more clearly understand messages from other participants where the person's facial expressions are not directly in view of the user.

### **FUTURE RESEARCH**

A wide range of fields have been drawn upon and applied to this research and as a result a wide variety of important research problems have been uncovered. A number of these apply primarily to remote collaboration itself, while others are of more general interest to emotion recognition, non-verbal communication interfaces and cognitive science and psychology applications.

Uncovered limitations from assessment of the chosen research methodologies have highlighted a number of additional research studies which if conducted may benefit further assessment of the developed ETA in its use for tackling remote collaborator multitasking. Through analysis of the results, the implementation study appears to have been unsuccessful in its ability to fully simulate remote worker multitasking. To better analyse the agent's benefits to multitasking remote interaction participants, a more detailed case study is proposed in which the ETA is employed in a real collaborative research situation over an extended time period.

In order to further assess the effects of the ETA on multitasking behaviours, it is proposed that a more realistic case study is performed in future work involving the use of the ETA in a full-scale online meeting between participants collaborating on a realworld unscripted task. Such a case study will also provide insight into the nature of current remote collaboration and help measure the extent in which multitasking truly effects such interactions.

Another more obvious direction to take in furthering this research is the replacement of the optical motion capture device with a webcam using robust image processing techniques to provide feature extraction and input for the ETA neural network. This would not only be more practical and less intrusive but also more economical if such an emotion recognition agent is to be eventually developed into a commercial product. Use of the ETA without specialised equipment beyond those required for remote work would minimise the boundaries which may arise from the extra expense and expertise needed to operate specialist technology. Much work has been done in developing efficient image processing algorithms for facial expression recognition in recent years and further research into the application of these algorithms for the ETA would undeniably benefit the future of our work.

It is also evident that future implementations of the ETA should apply greater focus on the support for tracking and recognition of non-universal emotions and states such as confusion, agreement and (dis)interest. The primary emotions used in the proof of concept prototype agent are of limited usefulness in real-life collaborative interactions, and thus research in order to extend the emotion recognition architecture to better deal with such states and emotions would be of great benefit.

## REFERENCES

- Appelbaum, S.H., Marchionni, A. & Fernandez, A., 2008. The multitasking paradox: perceptions, problems and strategies. *Management Decision* 46(9), 1313-1325.
- Black, J. & Hearne, S., 2008. Effective leadership of international virtual project teams, PMI Global Congress, Denver, Colorado, USA.
- Chopra, P., 2009 retrieved on 3 September 2009 from <http://paraschopra.com/>
- Ekman, P., Friesen, W.V., 1975. *Unmasking the Face*. Prentice-Hall Inc, New Jersey.
- Ekman, P., 1980. *The face of man: Expressions of universal emotions in a New Guinea village*. Garland STPM Press, New York.
- Lojeski, K.S., Reilly R. & Dominick, P., 2007. Multitasking and innovation in virtual teams. *Hawaii International Conference on System Sciences* (pp. 44b)
- Pantic, M., Rothkrantz, L.J.M., 1999. An Expert System for Multiple Emotional Classification of Facial Expressions. In: *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*, pp. 113-120
- Redfern, S., Colhoun, M., Hernandez, J., Naughton, N., Noonan, D. 2006. "Distributed Virtual Reality Learning Environments", Chapter 13 in A. Cartelli (ed.) "Teaching in the Knowledge society: new skills and instruments for teachers", Idea Publishing Group, 2006.
- Smith, P., Redfern, S., 2010. Facial Expression Tracking for Remote Conferencing Applications: An Approach to Tackling the Disadvantages of Remote Worker Multitasking. In *Proceedings of the Second International Conference on Games and Virtual Worlds for Serious Applications*, pp. 91-92
- Smith, P., Redfern, S., 2010a. Emotion tracking for remote conferencing applications using neural networks. In *Proceedings of the 21<sup>st</sup> National Conference on AI and cognitive science*, pp.240-249, National University of Ireland, Galway.
- Stergiou, C., Signanos, D., 1996. Neural networks. Final report on neural networks, SURPRISE 96 Journal 4. Department of Computing, Department of Electrical and Electronic Engineering, Imperial College of Science Technology and Medicine, Queen's Gate, London, UK. retrieved on June 28 from [http://www.doc.ic.ac.uk/~nd/surprise\\_96/journal/vol4/cs11/report.html#The Back-Propagation Algorithm](http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#The Back-Propagation Algorithm)
- Susskind J. M., Littlewort G., Bartlett M. S., Movellan J. & Anderson A. K., 2007. Human and computer recognition of facial expressions of emotion. *Neuropsychologia* 45, 152-162.
- Zhang, Z., 1999 Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6): pp. 893-911.
- Zucker, U., Radig, B. & Wimmer, M., 2008. Facial expression recognition-A comparison between humans and algorithms. Retrieved on January 7 from [http://www9-old.in.tum.de/people/wimmerm/lehre/sep\\_zucker/sep\\_zucker.pdf](http://www9-old.in.tum.de/people/wimmerm/lehre/sep_zucker/sep_zucker.pdf)