Lyons, M. J., Kluender, D., & Tetsutani, N. (2005). Supporting Empathy in Online Learning with Artificial Expressions. *Educational Technology & Society*, 8 (4), 22-30.

Supporting Empathy in Online Learning with Artificial Expressions

Michael J. Lyons
ATR Media Information Science Labs
ATR Intelligent Robotics and Communication Labs
2-2-2 Hikaridai, Keihanna Science City
Kyoto, 619-0288, Japan
http://www.irc.atr.jp/~mlyons
mlyons@atr.jp

Daniel Kluender
Embedded Software Group, RWTH Aachen University of Technology
Ahornstr.55, 52074 Aachen, Germany
http://www-i11.informatik.rwth-aachen.de
kluender@informatik.rwth-aachen.de

Nobuji Tetsutani
Department of Information and Arts, College of Science and Engineering
Tokyo Denki University, Hatoyama, Saitama, 350-0394, Japan
tetsutani@ia.dendai.ac.jp

ABSTRACT

Motivated by a consideration of the machine-mediated nature of human interaction in web-based tutoring, we propose the construction of artificial expressions, displays which reflect users' felt bodily experience, to support the development of greater empathy in remote interaction. To demonstrate the concept of artificial expressions we have implemented a system for the real-time visual display of affective signals such as respiration, pulse, and skin-conductivity which, combined with contextual information, may help partners in a learning interaction to estimate one another's level of arousal, stress, or boredom, for example. We have employed this system in a trial learning situation for the remote teaching and learning of Kanji, the Chinese characters used in written Japanese.

Keywords

Affective Computing; Empathy; Artificial Expressions; Web-based Tutoring Systems

Introduction

Human interaction over telecommunication networks necessarily involves an unnatural degree of disembodiment. Despite the burgeoning audio, video, text and graphic media that are now commonly available with chat programs we still do not have the rich, high bandwidth multi-sensorial exchange experienced in face-to-face interaction. Perhaps the key deficit in machine-mediated remote interaction is the much decreased level of non-verbal communication. Non-verbal signals such as gestures, facial expressions, and numerous other forms of body language play an important role in implicit and affective aspects of communication. Extensive studies in various branches of social and communication sciences show that skill in understanding and participating in these modes of interaction forms a significant component of human social expertise [2, 3, 4, 5, 9, 10].

Interactions within the context of education depend not only on the explicit exchange of information but also rely on implicit, affective modes of communication. Indication of salience, emphasis, understanding or misunderstanding, interest, boredom, acceptance, questioning, difficulty, all rely at least partially on non-verbal communication. Learning interactions which are mediated by telecommunications systems suffer from the limitations in non-verbal modes of exchange which do not support the communication of these pragmatic signals effectively. This observation led us to the investigation and development of systems to support the exchange of affective information for web-based learning applications.

Whereas the remoteness of web-based interaction implies a necessary degree of physical disembodiment, the very fact of machine-mediation allows us to consider novel forms of non-verbal communication not previously experienced in natural face-to-face interactions. The goals of this work are therefore as follows:

- propose, define, and describe the concept of artificial expression;
- demonstrate the technological feasibility of artificial expressions by designing and implementing a working hardware and software prototype platform

- explore the plausibility for the benefits which such a system could offer in remote education by conducting a trial of the implemented system within the context of a real learning situations

Our thinking in this paper has at its core the concept of empathy or shared feeling. Our hypothesis is that the increased embodiment of the interaction afforded by such artificial expressions can support the experience of empathy in remote interaction and that this could have numerous beneficial effects for online learning situations.

System Design and Implementation

Much of the prior work in affective computing has been aimed at human-computer interaction (HCI) and especially at machine recognition of affective state [12]. Education is concerned ultimately with human-human interaction. Therefore the approach we have taken here addresses computer-mediated human interaction, or human-computer-human interaction (HCHI). Accordingly, rather than aiming our research at having machines classify emotional states, we seek novel and effective means of gauging, representing, and communicating affective information in such a way as it may be easily interpreted by other humans.

Our approach is motivated by a significant body of work on affect which suggests that the high-level appraisal of emotional events is highly dependent on context and past experience [10, 13]. These studies led us to design systems which leave high level processing of emotional episodes to the cognitive abilities of the users themselves. The approach, which will be outlined below, may seem unambitious to some artificial intelligence enthusiasts who might dream of a machine which can understand a users' every whim. We emphasize that this is a mistaken perception and that our aim is, in fact, quite different. We do not deny the value of the machine recognition of emotion in humans as a research goal. However, The full complexity of the problem is often ignored. An example is that of the recognition of facial expressions by machine vision systems [7]. Extensive work by a dedicated community of researchers has led to considerable progress on a constrained version of the problem of automatic facial expression recognition [7, 11], but nearly all of these studies avoid dealing with some important aspects of the full problem (see, for example, the discussion of the effect of varying viewpoint on expression recognition in [8]). Moreover, there is the complex issue of context to be taken into account. Emotions take on colour and depth only when judged within the context of a situation. Our seemingly more modest goal is in fact a re-framing of this area of research: the goal is to provide new channels, artificial expressions, which allow humans to gauge aspects of core affect [13] (for example, stress or arousal), in machine-mediated communication.

To make the discussion more concrete we now outline a design of the system considered in this work. The system we propose involves several components. First, we use a set of non-intrusive wearable sensors to gauge the affective state of the wearer. Three robustly measurable variables, respiration, pulse, and skin conductance are used. These three signals have been commonly used in the affective computing literature [12] because of the widespread availability of sensors for their measurement and also because there is a relatively well-studied relationship between the subjects affective state and the measured signal. We implement a client/server architecture for sharing the affective signals over the internet in a transparent and user-friendly fashion. The signals are visualized intuitively with simple dynamical graphical displays in such a way that they can be understood without focal attention and with little learning. These *shared affective experience (SAE) displays* are intended to act as computer-mediated *artificial expressions* which give insight to a user's subjective, or felt affective experience. The meaning of these artificial expressions is intended to be learned constructively through social interaction with other users [16]. In other words, if the SAEs are correctly designed it should not be necessary to provide detailed instructions about how to interpret the displays. The user learns a tacit understanding of the SAE by observing the display of their own physiological data and relating it to the experience of their body feelings. This is generalized to the interpretation of the feelings of others. The understanding of the artificial expressions of another are corrected or modified through ongoing interactive behaviour in a variety of situations which arise over the course of time. We study such an interaction in the framework of a specific learning task: a web-based tutoring environment for writing Chinese characters or Kanji.

Implementation

Sensors

Respiration, blood volume pulse, and skin conductance were chosen because they can be measured robustly in a non-invasive fashion using readily available equipment. The relationship between these signals and affective states such as stress or arousal has been extensively studied in the past [10, 12]. For these reasons these signals are the most widely used in wearable sensors used in research on “affective computing” [12].

Figure 1 The sensors used to measure respiration (abdomen), skin conductance (fingers of one hand), and blood volume pulse (thumb of one hand)

The respiration (R) sensor is worn around the chest or abdomen (Figure 1). It produces a voltage in response to expansions or contractions of the abdomen or chest in breathing. Breathing may be roughly characterized by rate (cycles per minute) and depth (or amplitude) as well as pauses in the motion when the breathing is irregular or stops.

These variables are all related to affective state. For example, respiration rate increases and breathing becomes shallower with arousal, whereas with relaxation the breathing slows and deepens. Irregular breathing can result from tense states.

The *blood volume* pulse (BVP) sensor is worn on the tip of a finger or thumb (Figure 1). An infrared LED and light sensor are used to measure the reflectance near the surface of the skin. The amount of light reflected is a function of the volume of blood in the capillaries which varies with the pulse. The signal therefore depends on the pulse rate and blood pressure. A rise in the rate of the beating pulse and increase in amplitude correspond to arousal.

The *skin conductance* (SC) sensor is worn on two fingers of one hand, usually the same hand that the BVP sensor is worn on (Figure 1). This measures the electrical conductance of the skin. The SC signal measures the electrodermal response. This is readily demonstrated by the “startle response” - a small but definite jump in SC is seen in response to an unexpected noise or other stimulus. Changes in affect also result in a change in SC.

With our prototype implementation, the sensors are worn on the non-dominant hand. For the pilot study reported in this paper, the sensors do not interfere with participation in the communicative activity. For use in a wider range of contests, it may be preferable to devise an even less invasive system. For example, the sensors could be embedded in a cycling glove, which leaves the fingers free. Alternatively the sensors could be worn on parts of the body other than the hands, for example embedded in shoes. A wireless interface like Bluetooth could be used to send sensor data to the computer.

Data Acquisition

The Procomp+ system (Thought Technologies Ltd.) is used to convert analogue voltages from the sensors to digital signals that are sent to the computer via the RS-232 serial port. This is a medical grade device that is intended for use in clinical work and in biofeedback training. It samples all three signals at a rate of 20 Hz. The system is connected to the serial port via a fiber-optic cable ensuring complete electrical isolation from the computer – a critically important safety measure since the user is electrically connected via the SC sensors to the data acquisition device.

Client/Server System

A client process running on the local computer receives the samples from the three signals and sends all the incoming data to a broadcasting server using the TCP/IP protocol. A schematic of the client/server architecture of the system is shown in Figure 2. The server sends all received data to every client (hence the name broadcasting server) so each client can access sensor data from all the other clients. A single message containing a single sample from a sensor is 75 bytes in length, including some overhead for error protection. With the global data sampling rate of 20 Hz, this results in a total traffic rate of 36 kbits/s from each client. Our trials with the platform prototype and learning experiments made use of a 100 Mbit local area network. A client-to-client mean transmission time of 19.8 ms (standard deviation 12.8 ms) was measured. This latency does not produce subjectively noticeable delays under the conditions of our studies.

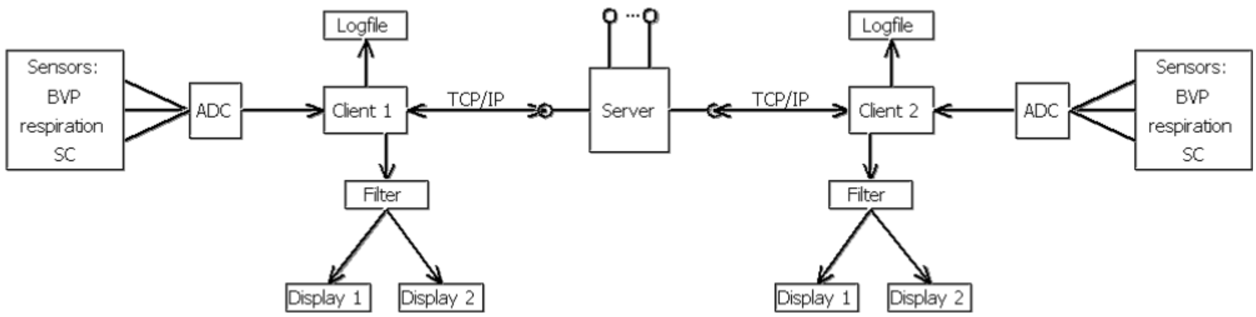

Figure 2. Schematic of the client server architecture used for the shared affective experience (SAE) system

Signal processing

Little or no signal processing is applied to the data from the sensors. Simple linear processing is applied to the SC signal to remove the slowly decaying baseline of the electrodermal response (see the top line in Figure 3). Basically this reflects the fact that after an electrodermal event releasing perspiration from a sweat gland, it takes a finite amount of time before the moisture dries up and the skin conductance returns to baseline. However this slowly varying signal can be removed with straightforward digital filtering techniques. We high-pass filter the SC signal with an IIR filter. Effectively, from each sample x(i) the processed signal d(i) is calculating a weighted running average, A(i), of all previous samples:

$$d(i) = x(i)-A(i)$$
$$A(i) = A(i-1)*0.995+x(i)+0.005$$

Note that the update of the 'baseline', A(i), only requires the last acquired sample, so the history of samples does not need to be stored to calculate the filtering function. The lower trace of Figure 3 shows the high-pass filtered SC signal, d(i). The peaks of this signal correspond to electrodermal events. A higher density of these events indicates a higher rate of sweating corresponding to greater arousal or level of stress. Absence of peaks indicates absence of electrodermal events, corresponding to a greater degree of relaxation.

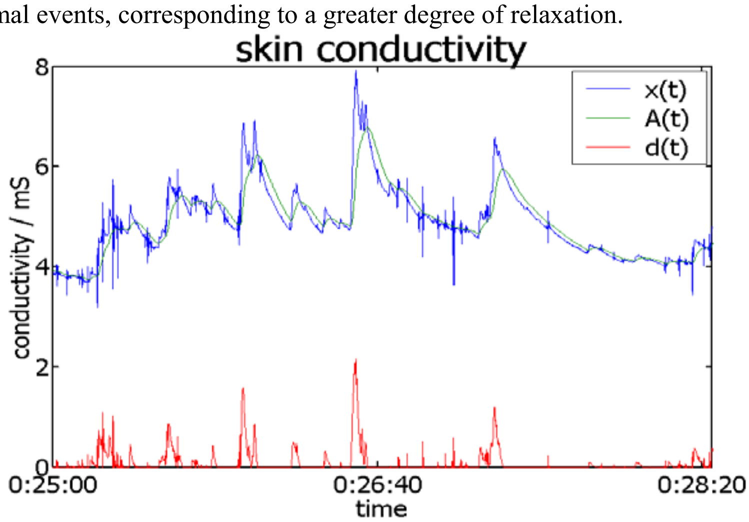

Figure 3. Time series of skin conductivity raw and high-pass filtered signals

Display

The three physiological signals are visualized by using a direct and intuitive display. The aim was to create displays that could be useful without detailed instruction or cognitive effort. Careful planning underlies the simple design of the SAE display shown in Figure 4. Two concepts guided our design. One is the idea of creating a *visual metaphor.* A metaphor is a means, in this case graphical, of expressing one kind of experience in terms of another [6]. The specific shapes, colours, and motions of the dynamic graphical displays were inspired by Arnheim's work on *visual thinking* [1] in that the dynamic features of the displays mirror the signals they represent. For example, as the chest or abdomen rises with an in-breath the blue column rises. This suggests identification of the internal experience of that physiological phenomenon with its concordant visual display. These considerations should illustrate why we consider it important to process the signals as little as possible: salient dynamical characteristics corresponding to felt bodily experiences are preserved.

A further reason for the visual simplicity and intuitiveness of the display is that it allows the information to be taken in at a quick glance, avoiding heavy cognitive demands for processing the artificial expressions. This is necessary if the display is to be used implicitly, in the periphery, as with the peripheral interfaces described by Weiser and Brown on in their work on calm *technology* [17].

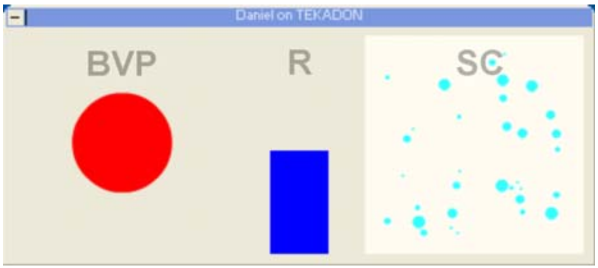

Figure 4 Shared Affective Experience (SAE) Displays

Each graphic is motivated by the physiological function that it represents. The respiration signal is shown as a blue column, with increased height of the column corresponding to increased stretch of the **R** sensor. The slowly rising and falling column is meant to suggest a lung inflating and deflating with the breathing cycle.

The value from the *BVP* sensor is displayed as the radius of a red circle. With the beating of the pulse the red circle expands and contracts like a beating heart.

The processed *SC* signal, d(i) in the equation above, is mapped to the probability of a blue circle suddenly expanding in size in a two dimensional display. This is suggestive of a patch of perspiring skin. At low values of SC the number of suddenly expanding patches is low and activity of the blue circles is sporadic. As *SC* increases the number increases and eventually the patch floods with blue as if the skin is saturated with sweat. To keep the displays as simple as possible no explicit calibration scales are included. These are implied by the geometry of the display. For example, the red circle representing the BVP signal cannot exceed the size of the display. The mapping of input signal to display is calibrated so that the signals operate in a range that that is easily visible to the user.

Kanji Tutoring Environment

We have defined the concept of artificial expressions and we have introduced an actual hardware and software implementation of a platform to support SAE expressions in remote interaction. The third and final goal of the current work is to provide an indication that the SAE displays could play a beneficial role in some computer-mediated learning situations. Furthermore it is natural to ask some practical questions such as whether users can learn to interpret the SAE displays as metaphors for felt experience of a remote tutor or student. We examined such questions empirically by conducting a preliminary study in the context of a web-based platform for remote kanji tutoring. Kanji are the Chinese characters used in written Japanese.

Since HCHI applications are the primary intended domain of application for the SAE displays it was natural to choose a learning situation which involves a two way interaction between users, in this case a learner and a tutor.

This specific learning task was chosen because, at any given time, there are several beginning learners of Japanese as a second language at our laboratory. For these learners, a lesson in writing kanji is an attractive prospect and it was possible for us to recruit a few volunteer "students" for our experiment – usually not an easy thing, working as we do in the environment of an independent research lab.

For the purposes of the experiment a kanji tutoring environment was constructed. It consists of an audio link and shared whiteboards (Figures 5 and 6). The student and teacher each use a large Wacom tablet and stylus to enter handwritten kanji to the shared whiteboards. There was no noticeable latency in the audio and whiteboard links.

Experiment

The preliminary experiment took the form of a structured lesson. First, the basic strokes used in writing kanji were reviewed. Then the tutor introduced five kanji having an increasing level of difficulty. This lesson was followed and concluded by a short quiz. The entire lesson took between 30 minutes and 1 hour according to the individual pace of learning. Four unpaid, but highly motivated, volunteers took part in the experiment - three "students" or beginning kanji learners and one teacher who is knowledgeable about kanji. The primary aim of the experiment was to explore what meanings the visual metaphors we created could take on in a tele-learning situation and whether users make use of the affective displays to gauge each others feelings and thereby provide a basis for more empathetic interaction.

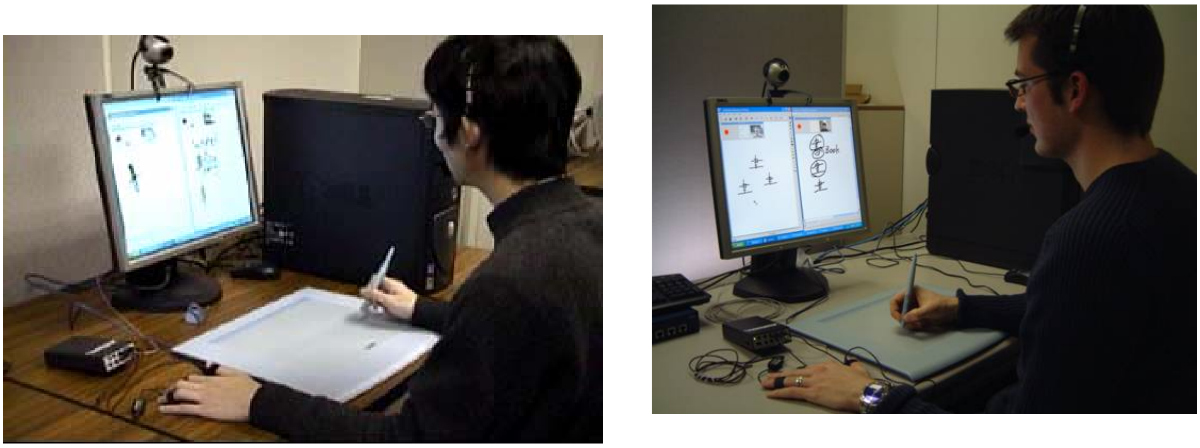

Figure 5. Web-based kanji tutoring platform used in our study of the SAE displays

We evaluated the interaction by interviewing the student and tutor separately after each session. Each was asked whether they found the SAE displays meaningful or useful for gauging their own emotional status or that of the other, during the task. Because both the task and the information displays were novel, the students generally stated that they were not yet able to make extensive use of the SAE displays during the lesson. However, one robust observation was that the skin conductance made the students more aware of their own emotional status. This was obvious almost as soon as the SC sensors were put on – skin conductivity is a sensitive, immediate, and reliable correlate of stress level. Indeed it seemed to take a few minutes before the students became comfortable with viewing their own skin conductivity information displayed in real-time in the presence of the experimenter.

The tutor is a member of our research group and so had greater familiarity with the interaction platform, sensors, and physiological signal displays. The major observation of the tutor, over the course of the several learning interactions, was that the skin conductance quickly became useful in pacing the lessons. By the tutor's account, excessive activity in the skin conductance was taken to imply that the level or speed of the lesson was too high for the student and needed to be relaxed.

It is important to emphasize again that we provided no a priori definition or interpretation of the SAE displays to the teachers or students other than to say that they reported on perspiration, breathing, and pulse. Rather it was intended that the interacting users arrive at their own heuristic understanding of the displays. It is known from previous work that these signals do contain information about affective state. Therefore we expect that our users, with the fine pattern processing capabilities of the human perceptual system and the cognitive ability to combine perceptual data with concepts of context and situation to arrive at meaningful understanding of the novel SAEs.

Our preliminary experiment demonstrated that this kind of process can take place at least for one of the displays – that of the SC. We would like to conduct a much more extensive study of the SAE platform in the context of real learning situations, but this is beyond the scope of the current paper.

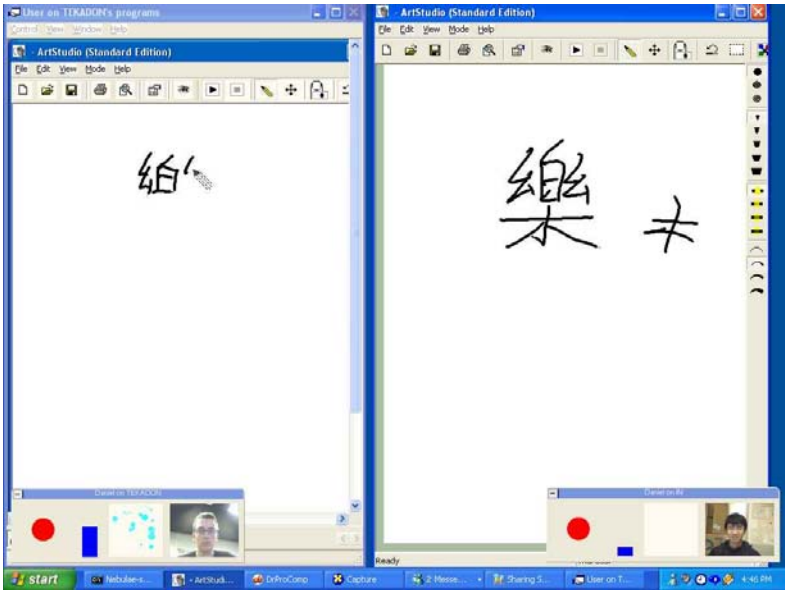

Figure 6. Shared whiteboards with SAE displays used in the Kanji tutoring study

Conclusion

Paying attention to felt bodily sensations can be an aid in recognizing and reducing stress and increasing one's level of relaxed alertness. This can enhance cognitive performance as well as the ability to assess another person's state of mind clearly and objectively. The hypothesis guiding the current work is that the SAE displays, by opening windows into another's affective experience, may furthermore afford increased *empathy*, the awareness both of our own and others' emotional states. We believe that this could have numerous benefits for real-time web based tutoring. Our experiment with the kanji tutoring platform has demonstrated that this is at least feasible though future work will be necessary to further explore the effects the SAE displays have on remote interaction.

The apparent simplicity of the SAE technology we have described in this paper is the result of a careful design process based on the selection of three main design patterns as components of the system, which we may call the *Connection to the Body; Direct and Intuitive Display; and Reciprocity.* Each of these patterns was inspired by precedent work in various fields of research.

The *Connection to the Body* pattern or component of our system was motivated by extensive work on biofeedback and affective computing. We go beyond the biofeedback work in that the sensory feedback of physiological processes takes place in the context of a meaning interaction, a constructive learning process, between users. Our use of the sensory connections to the body is different from that common in the affective computing research community in that we do not attempt to have the machine do all of the intelligent work. Rather, we try to connect our bodies with our own cognitive processes in remote interaction.

Precedent for the *Direct and Intuitive Display* pattern or component of the system comes from studies scientific visualization, gestalt theory of perception, as well as the notion of ambient display technology [17]. We were particularly inspired by Rudolf Arnheim's elegant notions of visual thinking [1] and Weiser and Brown's dictum that "Technologies encalm as they empower our periphery" [17]. The SAEs were designed to work ambiently and require little cognitive load from the users.

Perhaps the most important *pattern* we which found to be useful in the design of the system is *Reciprocity*: the affective displays are the same for tutor and student, and the information about affective data flows freely in both directions. This is expected to be a fundamental condition for the support of empathy or "shared feeling" [14, 15]. An intuitive explanation is as follows. One first identifies felt bodily experience with one's own affective display. In other words the understanding of the SAE is bootstrapped by observing the graphical displays and making a correspondence with feelings as they arise and pass in various contexts over a period of time. This is then generalized to the interpretation of another's feelings from the observation of the SAE of another. This can allows the user to infer the feelings of another by prolonged observation of the others affective display. Combined with contextual information, this can give users insight into how their actions influence another's felt experience and vice-versa. We propose that observation of each others SAE during the course of meaningful machine-mediated interaction could, over the course of that interaction, support users in developing an intuitive sense of each others being.

In summary, we have proposed here that the real-time display of physiological data can serve as artificial expressions of affective state. Sharing this information reciprocally over the internet can engender the experience of empathy in remote, online interaction. We have demonstrated the potential of this idea in the context of a web-based kanji tutoring platform. We consider this demonstration to be preliminary and suggestive and our main hope that it will spark the interest of other researchers working on technology-mediated educational interaction to amplify and extend our preliminary results.

Acknowledgements

We would like to express our gratitude to Mr. Chi-ho Chan and Mr. Gamhewage C. De Silva and for their helpful suggestions and interaction. Thanks also are due to the unpaid volunteers who took part in our study. This work was supported by the National Institute of Information and Communications Technology.

References

Arnheim, R. (1969). *Visual Thinking*, Berkeley CA: University of California Press.

Birdwhistell, R. L. (1970). *Kinesics and Context*, Philadelphia, PA: University of Pennsylvania Press.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (2000). *Embodied Conversational Agents*, Cambridge, MA: MIT Press.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*, London, UK: John Murray.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behaviour: Categories, origins, usage, and coding. *Semiotica, 1*, 49-98.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*, Chicago, USA: University of Chicago Press.

Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic Classification of Single Facial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21* (12), 1357-1362.

Lyons, M. J., Campbell, R., Plante, A., Coleman, M., Kamachi, M., & Akamatsu, S. (2000). The Noh Mask Effect: Vertical Viewpoint Dependence of Facial Expression Perception. *Proceedings of the Royal Society of London, B 267*, 2239-2245, Retrieved October 25, 2005, from, http://www.irc.atr.jp/~mlyons/pub_pdf/noh_mask.pdf.

Mehrabian, A. (1972). *Nonverbal Communication*, Chicago, USA: Aldine-Atherton.

Oatley, K., & Jenkins, J. (1996). *Understanding Emotions*, Oxford, UK: Blackwell.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1424 -1445.

Picard, R. W. (2000). *Affective Computing*, Cambridge MA: MIT Press.

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review, 110* (1), 145-172.

Thompson, E. (2001). Empathy and Consciousness. *Journal of Consciousness Studies, 8* (5-7), 1-32.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge MA: MIT Press.

Vygotsky, L. S. (1980). *Mind in Society*: *The Development of Higher Psychological Processes*, Cambridge MA: Harvard University Press.

Weiser, M., & Brown, J. S. (1996). *The Coming Age of Calm Technology*, Retrieved October 11, 2005, from, http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm.