

Evaluating affective interactions: Alternatives to asking what users feel

Rosalind W. Picard
MIT Media Laboratory
20 Ames Street, 020a
Cambridge, MA 02139
+1 (617) 253-0689
picard@media.mit.edu

Shaundra Bryant Daily
MIT Media Laboratory
20 Ames Street, 120g
Cambridge, MA 02139
+1 (617) 253-0369
sbdaily@media.mit.edu

ABSTRACT

In this paper, we advocate the use of behavior-based methods for use in evaluating affective interactions. We consider behavior-based measures to include both measures of bodily movements or physiological signals and task-based performance measures.

INTRODUCTION

Recent years have seen a large increase in research directed towards adding an affective component to human computer interaction. The ability to measure user affect has become important for intelligent interfaces that aim to either establish believable interactions or alter internal behavior based on the user's affect. Evaluating and interpreting this measure presents a challenge because of many ambiguities related to affect definition, communication, and interpretation.

Classical methods for evaluating affect tend to focus on questionnaires: asking you what you feel now, or interviews, perhaps after the experiment, with a video of your performance in front of you, asking you instant by instant to recall what you felt at each moment during the earlier task. While such "self report" methods are valuable, and we continue to use them in our work, this paper will highlight some alternatives to self-report of feelings. The discussion below is divided into two categories: body measures (e.g. changes in muscle activity), and task measures (e.g. better ability to solve a creative problem).

BODY MEASURES OF AFFECT

The last decade has brought great strides in giving computers affective perceptual abilities, with new sensors

for physiology and for behavior, such as body-worn accelerometers, rubber and fabric electrodes, miniature cameras and microphones, and garment or accessory-type devices, along with new algorithms for recognizing patterns in the sensed signals such as recognition of facial expressions from video or of stress patterns from thermal imagery of the face and other physiological measures. Body measures are not presented here as a replacement for other measures, but rather as additional information that may help combat some of the difficulties encountered with questionnaires and other more subjective methods. Possibly the biggest advantage is that body measurements can be taken in parallel with the interaction rather than interrupting the user or asking him after the task.

An exhaustive list of body-based measures is beyond the scope of this paper, however, Table 1 cites a sample of existing methods (leaving out lots of examples of publications in each of these categories, and also leaving out categories, e.g. EEG and ECG-based measures, and more). Clearly there are lots of possible body measures that may capture aspects of an affective state, including the combination of multiple modalities, which can reduce the uncertainty associated with using a single measure (Mednick et al. 1964; DeSilva et al. 1997; Huang et al. 1998; Picard et al. 2001; Kapoor et al. 2004)

One benefit of these "body" measures is that they can provide additional insight into the user's emotional state without directly relying on his cognitive judgment of his emotional state. Additionally, some of them can be used without the user's knowledge, perhaps with the goal of limiting the amount of misinformation that may arise from his feeling of being monitored. (This can also be seen as a drawback if one is concerned about privacy and about the use of sensing without a person's knowledge).

Table 1. Body-based measures of affect (partial set of examples)

Modality	Sensor	Is it socially communicated?	Comments
Facial Activity	Video (Tian et al. 200; Barlett et al. 1999; Donato et al. 1999; Cowie et al. 2001)	Yes	Facial expressions can differ significantly from genuinely felt feelings
	IR Video (Kapoor et al. 2003)	Highlights pupils & specularities	Usually works better than ordinary video when head moves (better eye detection)
	Thermal Video (Pavlidis et al. 2002)	No	Being explored to detect stress and other changes related to deception and frustration
Posture Activity	Force sensitive resistors (Smith 2000; Mota & Picard 2003; Tan et al. 2003)	Yes, but not as pressure	Good results discriminating level of interest in students in computer learning interactions
Hand Tension & Activity	Force sensitive resistors or Sentograph (Clynes 1986; Reynolds 2001; Qi & Picard 2002; Dennerlein et al. 2003)	Varies; depends on gesture	Can be sensed from handling of mouse, steering wheel, etc., and pressure has been shown to be higher during a frustrating task
Gestural Activity	Electromyogram electrodes (Marrin Nakra & Picard 1998; Dubost & Tanaka 2002)	Visibility varies	Shown for expression sensing in conducting music; other gestures largely unexplored w.r.t. expression
Vocal Expression	Microphone (Banse & Scherer 1996; Cowie et al. 2001; Ang et al. 2002; Breazeal & Aryananda 2002; Fernandez 2004)	Yes	Most methods are great for discriminating arousal but not for valence; limited to times when user is speaking
Language and choice of words	Text analysis tools (Goertzel et al. 2000; Elliott 2002; Liu et al. 2003)	Yes	Can be used with interfaces requiring textual input; promising for valence; trivial to sense scripted dialogue moves
Electrodermal Activity (a.k.a Galvanic Skin Response)	Electrodes (can also be clothing snaps, metallic fabric, etc.) (Picard & Scheirer 2001)	No; except perhaps sweaty palm	Good at detecting changes in arousal but doesn't distinguish positive/negative, and can also be triggered by non-affective changes

TASK MEASURES OF AFFECT

A variety of findings have shown ways that affective states tend to influence various behaviors on subsequent tasks. To the extent that such findings are robust, they can be used to indirectly assess aspects of affect that may have been elicited during an interaction. For example, Isen and colleagues (1987) have demonstrated that positive affect can influence the way cognitive material is organized and have shown that this enables broader forms of thinking, and consideration of less typical solutions, which is useful in creative problem solving (Isen 1987). Using a variety of different techniques such as gifts, comical movies, or refreshments, a positive affect state was induced in the study participants. The slightly positive emotional state benefited subjects' performance on tests such as Duncker's (1945) candle task, the Mednicks' Remote Associates Test and medical decision-making with hypothetical patients (Mednick et al. 1964; Isen 1987; Isen 1991). Subjects also better integrated the material presented to them and exhibited an ability to better organize their protocols as compared to a control group (Isen 1991). Also, like Isen, Schwarz has found that a negative affective state corresponds to a higher level of spontaneous causal reasoning, which fosters bottom-up, data driven processing. Therefore, when involved in an analytical task, it may actually help to be in a sad mood (Schwarz 2002).

Recently, this kind of "task measure" was examined to see if exposing users to reading tasks using two different kinds of fonts impacted it. The hypothesis was that people reading a passage written using good typography would perform better on the candle task and on the remote associates test than readers reading the same content presented with poor typography. While the first such study of this kind was small (N=20), the findings were supportive of this hypothesis (Larson & Picard 2005).

The typography study also examined another indirect task assessment method that we think is of increasing interest for assessing affect. This measure involves asking somebody "how long do you think you spent on that task?" and, its ratio to the actual time spent is known as subjective duration assessment (Czerwinski et al. 2001). Using this measure (of how long they thought they were working on the task before they were interrupted, vs. how long they really were working on it) it has been shown that difficult tasks tend to be overestimated in duration while easy tasks are underestimated in duration. We hypothesize that this measure might also be related to frustration, which predicts it would also be influenced by task difficulty and by other factors such as time pressure and irritating aspects of the task. In two separate typography studies, this measure was found to be significant ($p < 0.05$), each study with N=20: In both studies, subjects using the good typography

underestimated their reading times by a significantly larger amount than did subjects using bad typography. In one of the studies, this difference held even though subjects' self-reports of the quality of the typography did not differ significantly.

We have been interested in the generality of such time-based measures for indirectly assessing affect. Recently, Picard and Liu proposed a new variation, "relative subjective count (RSC)," based on asking people who were interrupted many times during the day by the technology being investigated, "How many times does it *seem* like you were interrupted by this technology?" This perceived number was divided by the actual number of interruptions to obtain the RSC. Comparing two nearly identical systems, which differed mainly in their expressions of empathy, they found that people had a significantly lower RSC when the technology was empathetic. This measure also agreed with self-reported views of how stressful the technology was. We suggest that the RSC might provide a new indirect way of assessing affect related to the stress or irritation associated with an interaction (Liu 2004).

While these kinds of assessment measures are new and require much more investigation before they are fully understood, they potentially offer a nice alternative for exploring certain states such as stress and frustration related to an interaction without having to ask directly about any negative aspects of the user's experience.

In another area of interest, Lerner, et al. have shown that affect has important influences on economic decision-making. Positive affect was shown to reverse the endowment effect (the tendency for selling prices to exceed buying or "choice" prices for the same object), while negative affect eliminated the endowment effect (Lerner et al. 2004). The affect, in this case, was evoked by using the viewing of movies, followed by a writing task in which the subjects attempted to write about how they would describe how they were feeling to a friend.

We think that behavioral task measures such as these may prove powerful for indirectly assessing when a desired positive state has likely been achieved in a group of individuals. While they are not as direct as measuring an individual's emotional bodily reaction, and so far the results are only on populations, and not on individuals, these task-based measures can be accomplished without any special sensors or sophisticated analysis. With a large enough group of individuals, the statistical significance can potentially provide a strong assessment.

CONCLUSIONS

We have highlighted several means of assessing affect beyond directly asking somebody what they are feeling. One can imagine an interface interaction in which the user's facial and electrodermal activity are monitored for valence and arousal information, and the interaction is

followed by an assessment task such as the Duncker's Candle Test, where better success is expected with more positive affect. Thus, information about the user's affect can be gleaned from the physiological sources as well as the task performance. This data could additionally be compared against self-reported measures. There is a lot of room for new methods to be discovered; the ones we have presented here are just a few of the possibilities.

REFERENCES

- Ang, J., R. Dhillon, A. Krupski, E. Shriberg and A. Stolcke (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In Proc. International Conference Spoken Language Processing (ICSLP), Denver.
- Banse, R. and K. R. Scherer (1996). "Acoustic Profiles in vocal emotion expression." Journal of Personality and Social Psychology **70**(3): 614-636.
- Barlett, M. A., J. C. Hager, P. Ekman and T. Sejnowski (1999). "Measuring facial expressions by computer image analysis." Psychophysiology **36**(2): 253-263.
- Breazeal, C. and L. Aryananda (2002). "Recognizing affective intent in robot directed speech." Autonomous Robots **12**(1): 83-104.
- Clynes, D. M. (1986). Sentics. Cambridge, MA, MIT Press.
- Cowie, R., E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. G. Taylor (2001). "Emotion recognition in human-computer interaction." IEEE Signal Processing Magazine **18**(1): 33-80.
- Czerwinski, M., E. Horvitz and E. Cutrell (2001). Subjective Duration Assessment: An Implicit Probe for Software Usability. Proceedings of IHM-HCI 2001 Conference, Lille, France.
- Dennerlein, J., T. Becker, P. Johnson, C. Reynolds and R. W. Picard (2003). Frustrating Computer Users Increases Exposure to Physical Factors. Proceedings of the International Ergonomics Association, Seoul, Korea.
- DeSilva, L. C., T. Miyasato and R. Nakatsu (1997). Facial Emotion Recognition using Multi-modal Information. Proceedings of IEEE International Conference on Information, Communications, and Signal Processing, Singapore.
- Donato, G., M. Barlett, J. Hager, P. Ekman and T. Sejnowski (1999). "Classifying facial actions." IEEE Patter Analysis and Machine Intelligence **2**(10): 974-989.
- Dubost, G. and A. Tanaka (2002). A Wireless, Network-based Biosensor Interface for Music. Proceedings of International Computer Music Conference, Goeteborg, ICMA.
- Elliott, C. (2002). "The Affective Reasoner: A Process Model of Emotions in a Multi-agent System," PhD Thesis, Northwestern University The Institute for the Learning Sciences
- Fernandez, R. (2004). "A Computational Model for the Automatic Recognition of Affect in Speech," PhD Thesis, MIT Media Arts & Sciences

- Goertzel, B., K. Silverman, C. Hartley, S. Bugaj and M. Ross (2000). The Baby Webmind Project. Proceedings of AISB 2000.
- Huang, T. S., L. S. Chen and H. Toa (1998). Bimodal Emotion Recognition by Man and Machine. ATR Workshop on Virtual Communication Environments, Kyoto, Japan.
- Isen, A. M. (1987). "Positive affect facilitates creative problem solving." Journal of Personality and Social Psychology **52**: 1122-1131.
- Isen, A. M. (1991). "The influence of positive affect on clinical problem solving." Medical Decision Making **11**: 221-227.
- Kapoor, A., R. W. Picard and Y. Ivanov (2004). Probabilistic Combination of Multiple Modalities to Detect Interest. International Conference on Pattern Recognition, Cambridge, U.K.
- Kapoor, A., Y. Qi and R. W. Picard (2003). Fully Automatic Upper Facial Action Recognition. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003) held in conjunction with ICCV 2003, Nice, France, October 2003.
- Larson, K. and R. Picard (2005). The Aesthetics of Reading. Appears in Human-Computer Interaction Consortium Conference, Snow Mountain Ranch, Fraser, Colorado.
- Lerner, J. S., D. A. Small and G. Loewenstein (2004). "Heart Strings and Purse Strings." Psychological Science **15**(5): 337-341.
- Liu, H., H. Liberman and T. Selker (2003). A Model of Textual Affect Sensing using Real-World Knowledge. Proceedings of the 8th international conference on Intelligent user interfaces, Miami, FL.
- Liu, K. (2004). "A Personal, Mobile System for Understanding Stress and Interruptions," M.S. Thesis, MIT Media Arts & Science
- Marrin Nakra, T. and R. Picard (1998). The Conductor's Jacket: a Device for Recording Expressive Musical Gestures. Proceedings of the ICMC, San Francisco, ICMA.
- Mednick, M. T., S. A. Mednick and E. V. Mednick (1964). "Incubation of creative performance and specific associative priming." Journal of Abnormal and Social Psychology **69**: 84-88.
- Mota, S. and R. W. Picard (2003). Automated Posture Analysis for Detecting Learner's Interest Level. Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction, CVPR HCI, June 2003.
- Pavlidis, I., N. L. Eberhardt and J. A. Levine (2002). "'Human behaviour: Seeing through the face of deception'." Nature **415**(35).
- Picard, R. W. and J. Scheirer (2001). The Galvactivator: A Glove that Senses and Communicates Skin Conductivity. Proceedings from the 9th International Conference on Human-Computer Interaction, New Orleans, August 2001, New Orleans.
- Picard, R. W., E. Vyzas and J. Healey (2001). "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State." IEEE Transactions Pattern Analysis and Machine Intelligence **23**(10): 1185-1191.
- Qi, Y. and R. W. Picard (2002). Context-sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification. Proceedings of International Conference on Pattern Recognition, August 2002, Quebec City, Canada.
- Reynolds, C. (2001). "The Sensing and Measurement of Frustration with Computers," M.S. Thesis, MIT Media Arts & Science
- Schwarz, N. (2002). Situated Cognition and the Wisdom of Feelings: Cognitive Tuning. The Wisdom in Feeling. F. Barrett and P. Salovey. New York, Guilford Press: 144-166.
- Smith, J. (2000). "GrandChair: Conversational Collection of Family Stories," M.S. Thesis, MIT Media Arts & Sciences
- Tan, H. Z., L. A. Slivovsky and A. Pentland (2003). "A sensing Chair Using Pressure Distribution Sensors." IEEE/ASME Transactions on Mechatronics **3** **6**.
- Tian, Y., T. Kanade and J. F. Cohen (2001). "Recognizing action units for facial expression analysis." Pattern Analysis and Machine Intelligence **23**(2).

BACKGROUND OF THE AUTHORS

Rosalind W. Picard is founder and director of the Affective Computing Research Group at the MIT Media Laboratory. The author of over a hundred peer-reviewed scientific articles in multidimensional signal modeling, computer vision, pattern recognition, machine learning, and human-computer interaction, Picard is known internationally for pioneering research in affective computing and, prior to that, for pioneering research in content-based image and video retrieval. Her award-winning book, Affective Computing, (MIT Press, 1997) lays the groundwork for giving machines the skills of emotional intelligence.

Shaundra B. Daily is a graduate student at the MIT Media Laboratory, working in both the Affective Computing and Future of Learning Groups. She holds a Bachelor's in Electrical Engineering with honors from the Florida State University, and a Master's degree, in Electrical Engineering and Computer Science, from Florida Agricultural and Mechanical University. Her main interests include the design and evaluation of interfaces designed to support affective development through writing, constructing, and other forms of expression as well as technologically supported community and economic development.