M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 444
Appears in: Proceedings of the ICASSP '98

Digital Processing of Affective Signals

Jennifer Healey and Rosalind Picard
Rm E15-389, The Media Laboratory
Massachusetts Institute of Technology
20 Ames St., Cambridge, MA 02139
fenn@media.mit.edu, picard@media.mit.edu

Abstract

Affective signal processing algorithms were developed to allow a digital computer to recognize the affective state of a user who is intentionally expressing that state. This paper describes the method used for collecting the training data, the feature extraction algorithms used and the results of pattern recognition using a Fisher linear discriminant and the leave one out test method. Four physiological signals, skin conductivity, blood volume pressure, respiration and an electromyogram (EMG) on the masseter muscle were analyzed. It was found that anger was well differentiated from peaceful emotions (90%-100%), that high and low arousal states were distinguished (80%-88%), but positive and negative valence states were difficult to distinguish (50%-82%). Subsets of three emotion states could be well separated (75%-87%) and characteristic patterns for single emotions were found.

1 Introduction

Human-computer interface design has become more sensitive to the needs of the user through user centered design practices and "intelligent" assistance programs. However, the computer still has no direct channel to confirm a user's affective reaction to the information or options being presented. People naturally learn from affective reactions. For example, if a human assistant made a suggestion which angered you, they would be hesitant to offer that suggestion again. A computer with the ability to recognize affective response could potentially do a better job of adapting its behavior to you.

Traditionally, the fields of psychology and psychophysiology have sought features which indicate common affective states across large populations, despite the fact that there are wide variations in individual response patterns due to gender, personality type and ethnic background [1]. In this paper, we propose a tailored model that is based on recognizing the physiological signature of a single user's affective patterns over time.

This analysis also differs from that of traditional emotion research in the time window considered. This experiment analyses how the user's affective state changes over an average period of three minutes, far longer than the one-to-ten second reaction usually analyzed in emotion research [2]. In other analysis, we have found that affective measurements for a single individual are easily overwhelmed by physiological factors within such a short window [3]; therefore in this work, we look for larger changes that could be detected in an individual over a three minute window.

2 Experimental Protocol

In this experiment, a single subject intentionally expressed eight affective states toward a computer over a period of 32 days. Each affect state was maintained for an average of three minutes. A total of twenty usable days of data were collected due to sensor failure during some days. Data from four sensors, a skin conductance sensor, a photoplethysmeograph, a respiration sensor and an electromyogram (EMG), was collected and sampled at 20 samples per second using the ProComp unit from Thought Technologies, Ltd.

The subject, an actress trained in guided imagery techniques, was asked to experience and intentionally express eight affective states using a computer controlled prompting system developed by Manfred Clynes for the Sentic Cycles experiment. This experiment was chosen for ease of use and because Clynes reported finding a unique finger pressure signature for each of the eight emotion states used [4]. These states: no emotion, anger, hate, grief, love, romantic love, joy and reverence, are not the most commonly used in human computer design research, but they provide a set of affective states which span the ranges of high and low arousal and positive and negative valence.

Descriptive guidelines on the meaning of each emotion word were given to the subject before the experiment. The subject then reported the images she used to induce each state and the degree to which she found each experience arousing (exciting, distressing, disturbing, tranquil) and the degree to which they felt the emotion was positive or negative (valence). Daily ratings varied in intensity, but the overall character of each state was consistent, as described in Table 1.

3 Feature Extraction

A total of eleven features were used to differentiate affective state: the mean EMG activity, the mean and mean slope of the skin conductivity, average heart rate and heart rate change, and the normalized mean, variance, and four power spectral density characteristics of the respiration signal. Anger was most easily distinguished in many cases; however, the power spectral density analysis for respiration showed a distinct signature for grief. Sample signals from the four physiological signals are shown in Figure 1. The annotations in the top graph indicate the times at which the subject expressed no emotion, anger, hate, grief, love,

Emotion	Imagery	Description	Ar.	Val.
no emotion	blank paper typewriter	boredom vacancy	low	neut.
anger	people who aroused anger	desire to fight	very high	very neg.
hate	injustice cruelty	passive anger	low	neg.
grief	deformed child loss of mother	loss sadness	high	neg.
love	family summer	happiness peace	low	pos..
r. love	romantic encounters	excitement lust	very high	pos.
joy	Ode to Joy	uplifting happiness	med. high	pos.
rev-erence	church prayer	calm peace	very low	neut.

Table 1: A description of the emotional states intentionally induced and expressed in this experiment

Average Features for 20 trials

	ne	an	ha	gr	lv	rl	jy	rv
EMG1	1.0	3.2	1.5	1.9	1.2	1.5	1.7	0.9
SCR1	.56	.68	.60	.54	.54	.56	.60	.57
SCR2	-.36	.76	-.82	.05	-.01	.01	.20	-.44
HR1	78	86	80	81	80	83	80	78
HR2	.20	.01	.04	.00	-.30	.01	-.01	-.07
RT1	.10	.14	.02	-.12	-.01	.01	.00	-0.2
RT2	.02	.02	.02	.08	.02	.02	.06	.01
RF1	.32	.26	.34	.33	.43	.24	.35	.33
RF2	.38	.36	.27	.18	.23	.29	.24	.36
RF3	.07	.12	.07	.07	.04	.11	.06	.06
RF4	.03	.04	.03	.03	.02	.04	.02	.03

Table 2: The average values of the eleven features extracted from the data for the emotion states: no emotion (ne), anger (an), hate (ha), grief (gr), love (lv), romantic love (rl), joy (jy) and reverence (rv). The feature HR2 is multiplied by 100 here for formatting purposes.

romantic love, joy and reverence.

Eleven features were calculated from the four signals. The EMG on the masseter (jaw) muscle captured the presence or absence of activity in the muscle. For this signal the mean activity over each emotion period was calculated (EMG1).

The skin conductivity response (SCR) sensor detected chances in the electrical conductance of the palm of the hand. A long averaging window (25 second hanning) was used to smooth the raw data and characterize the general trend of each affective state. The signal was normalized using the following suggested criteria to account for baseline fluctuations between days [5]:

$$\frac{SCR-mean(SCR)}{max(SCR)-min(SCR)}$$

The mean of this signal for each emotion state was then calculated (SCR1) A second feature, the average change in

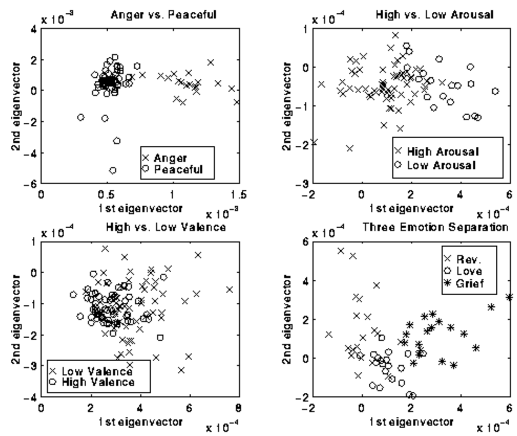

Figure 1: An example of a session's data collected from four sensors. Signals from the EMG on the masseter muscle in microvolts (top), the skin conductance waveform (in micro-Siemens), the heart rate (in beats per minute), and the respiration waveform (in % expansion) are shown. The annotations in the EMG waveform indicate the periods during which the subject was asked to express no emotion, anger, hate, grief, love, romantic love, joy and reverence

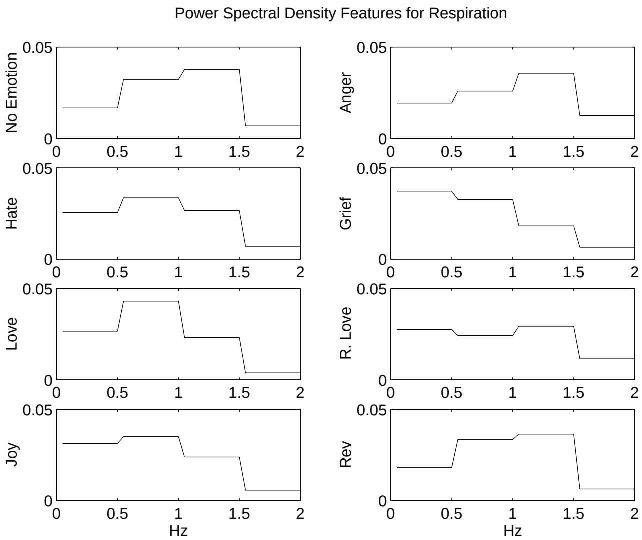

Figure 2: Power Spectral Density

the slope was calculated by taking the mean of the first difference of the smoothed signal (SCR2).

Heart rate was calculated by taking the inverse of the inter-beat interval from the blood volume pulse waveform. For each emotion, the mean of this rate was calculated (HR1). Also, the average acceleration or deceleration of the heart rate was calculated by taking the mean of the first difference of the heart rate (HR2).

The respiration sensor measured expansion and contraction of the chest cavity using a Hall effect sensor attached around the chest with a velcro band. The signal was normalized by subtracting off the overall mean of the data for that day to account for variations in the initial tightness of the sensor placement from day to day.

The mean and variance of this normalized signal were calculated for each emotion state (RT1, RT2) in the time domain. A power spectral density analysis of the signal was also calculated and provided a characteristic signature for some affective states as shown in Figure 2. The features (RF1-RF4) represent the average energy in each of the first four 0.5Hz bands of the PSD (0-2Hz).

The physiological features extracted were hypothesized to reflect the physiological changes in the person experiencing the emotion. The mean of the EMG was originally intended to detect jaw clenching during the emotion anger. However, higher mean activation during joy and grief indicates that smiling or frowning will also be detected with this feature. For the skin conductivity, the slope was more discriminating than the mean. Over the three minute period, the skin conductance tends to gradually increase for arousing emotions and tends to decrease for less arousing states. The mean heart rate for each emotion was more discriminating than the measure of heart rate acceleration chosen. For this time window it may be that the more general measure of overall rate is more descriptive, whereas for a shorter time window, heart rate acceleration has been a more discriminating feature [6]. For the respiration signal, the mean and variance of the time signal together differentiated between shallow regular breathing, shallow breathing punctuated by deep gasps and regular deep breathing. The PSD energy band features give more general characterization of the rhythmic patterns of respiration in each emotion state.

4 Pattern Recognition Results

The techniques to discriminate states based on the eleven features used a Fisher linear discriminant projection [7] and the leave one out test method. For each trial, a single point, x, was excluded from the data set and a Fisher projection matrix, W was calculated for remaining members of the set. The excluded point was then projected using W and classified using quadratic and linear classifiers, in the standard method described by Therrien [8].

The Fisher projection matrix which in some sense maximizes the ratio of the between-class scatter, S_B, to the within class scatter, S_W [7], where these matrices are defined as:

$$S_W = \sum_{i=1}^{c} \sum_{x \in \chi_i}^{n} (x - m_i)(x - m_i)^t$$

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^t,$$

given c is the number of classes, n_i is the number of sample vectors in a class, m_i is the sample mean for class i, m is the total mean, and $x \in \chi_i$ are the 11-dimensional feature vectors comprising class i. The Fisher projection matrix is the matrix W whose columns, w_i correspond to the largest eigenvalues in:

$$S_W^{-1} S_B w_i = \lambda w_i$$

This matrix is then used to project the test point onto the classifier space using

$$y = W^T x$$

The results of the recognition on a number of subsets of data are reported in Table 3. The discrimination is best between anger and a set of more peaceful emotions containing the classes: no emotion, love and reverence. The entire set of eight emotion classes can also be well separated into two categories: high arousal containing anger, grief, romantic love and joy, and low arousal containing no emotion, hate, love and reverence. However, no good discrimination was found for positive valence vs. negative valence emotions.

Although specific patterns were not found to discriminate all eight emotions, certain subsets of three emotions could be well separated as shown in Table 4.

5 Summary

Affect recognition is an imperfect task even between humans. In studies of affect in speech, in absence of context clues, people were able to identify affective states with only about a 60% recognition rate [9] and approval or disapproval with a rate of 65-85% [10].

Computers are now able to begin to recognize affective facial expression through video analysis at rates of 80% -

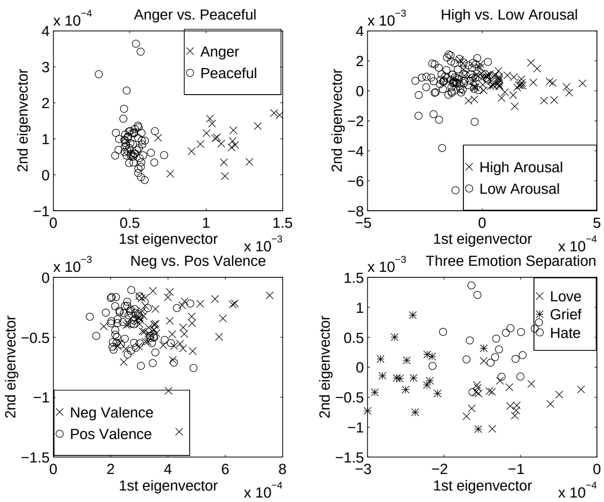

Figure 3: Anger is well separated from more peaceful emotions, in this example the states of No Emotion, Love and Reverence make up the set of peaceful states

Emotion Set	Set Size	Linear error	Linear correct	Quadratic error	Quadratic correct
anger	20	2	90%	0	100%
peaceful	60	1	98%	1	98%
high arousal	80	15	81%	16	80%
low arousal	80	11	86%	10	88%
positive	60	15	75%	11	82%
negative	60	29	53%	30	50%

Table 3: The results of discriminating between subsets of emotions. For these results the peaceful class contains no emotion, reverence, and love; the aroused class contains anger, grief, romantic love, and joy; the calm class contains no emotion, hate, love, and reverence; the positive valence class contains love, romantic love, and joy, and the negative valence class contains anger, hate, and grief. Anger was most easily distinguished, and good discrimination was achieved for the high vs. low arousal states. Positive and negative valence were not well separated.

Emotion Set em1	em2	em3	Linear errors	Linear correct	Quadratic errors	Quadratic correct
no em	joy	rev	1-7-4	80 %	2-5-4	82 %
anger	hate	r.l.	6-4-3	78 %	5-4-3	80 %
anger	hate	rev	5-4-3	80 %	5-3-3	82 %
anger	grief	rev	2-6-2	83 %	3-4-1	87 %
grief	love	rev	6-6-4	73 %	5-5-5	75 %
anger	r.love	rev	4-5-3	80 %	4-6-3	78 %

Table 4: Although all eight classes could not be separated, several subsets of three emotion classes could be differentiated using these features.

98% on small sets of emotions [11], [12], with the best performance on deliberately expressed (mildly exaggerated) emotions.

Processing physiological signals offers yet another avenue to communicating affective state to a computer. These preliminary results show that single emotions such as anger and emotional attributes such as arousal and valence can be identified at a level comparable to human recognition of emotion. This may be used to help the computer learn the user's preferences in the same way the people learn from each other's affective responses.

Perhaps the best results will come when the computer has access to multiple modalities: face, voice, gesture, posture, physiological sensing, and so forth. With information from context such as user location, time of day, workload, proximity to friendly or unfriendly people, and diction analysis of direct user response, affect analysis should become even more precise in building a long term model of an individual's affective responses. This information can then be used to help computers choose more intelligent responses when interacting with people.

6 Acknowledgments

The authors would like to thank Jonathan Klein for his helpful comments on human computer interface design, Rob Gruhl and Dan Gruhl for technical assistance in conducting this experiment, and the MIT Shakespeare Ensemble for acting support.

References

[1] R. W. Picard. *Affective Computing*. The MIT Press, Cambridge, MA, 1997.

[2] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3/4):169–200, 1992.

[3] R. W. Picard and J. Healey. Affective wearables. *Personal Technologies*, (1):231–240, 1997.

[4] Dr. M. Clynes. *Sentics: The Touch of the Emotions*. Anchor Press/Doubleday, 1977.

[5] John T. Cacioppo and Louis G. Tassinary. Inferring psychological significance from physiological signals. *American Psychologist*, 45(1):16–28, Jan. 1990.

[6] W. M. Winton, L. Putnam, and R. Krauss. Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology*, 20:195–216, 1984.

[7] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* Wiley-Interscience, 1973.

[8] C. W. Therrien. *Discrete Random Signals and Statistical Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1992.

[9] K. R. Scherer. Ch. 10: Speech and emotional states. In J. K. Darby, editor, *Speech Evaluation in Psychiatry*, pages 189–220. Grune and Stratton, Inc., 1981.

[10] D. Roy and A. Pentland. Automatic spoken affect analysis and classification. In *Proceedings of the Second International Conf. on Automatic Face and Gesture Recognition*, pages 363–367, Killington, VT, Oct. 1996.

[11] Y. Yacoob and L. S. Davis. Recognizing human facial expressions from log image sequences using optical flow. *Transactions on Pattern Recognition and Machine Intelligence*, 18(6):636–642, June 1996.

[12] I. A. Essa and A. Pentland. Facial expression recognition using a dynamic model and motion energy. In *International Conference on Computer Vision*, pages 360–367, Cambridge MA, 1995. IEEE Computer Society.