BIOSIGNALS ANALYSIS AND ITS APPLICATION IN A PERFORMANCE SETTING
Towards the development of an Emotional-Imaging Generator

Mitchel Benovoy, Jeremy R. Cooperstock
Centre for Intelligent Machines, McGill University, 3480 University St., Montreal, Canada
benovoym@cim.mcgill.ca, jer@cim.mcgill.ca

Jordan Deitcher
Director, E-Motion Project
liminal@total.net

Abstract: The study of automatic emotional awareness of human subjects by computerized systems is a promising avenue of research in human-computer interaction with profound implications in media arts and theatrical performance. A novel emotion elicitation paradigm focused on self-generated stimuli is applied here for a heightened degree of confidence in collected physiological data. This is coupled with biosignal acquisition (electrocardiogram, blood volume pulse, galvanic skin response, respiration, phalange temperature) for determination of emotional state using signal processing and pattern recognition techniques involving sequential feature selection, Fisher dimensionality reduction and linear discriminant analysis. Discrete emotions significant to Russell's arousal/valence circumplex are classified with an average recognition rate of 90%.

1 INTRODUCTION

Emotion classification based on *external* data collection schemes, such as speech analysis and facial-expression recognition from images has been studied extensively. The literature offers numerous examples of relatively acceptable recognition rates (Black et al., 1995; Lyons et al., 1999; Bartlett et al., 1999; Ververidis et al., 2004). However, because these systems require sensors, such as cameras or microphones, focused directly on the subject, they are restrictive in terms of movement and problematic in terms of signal interference from other devices. Moreover, video analysis methods tend to encourage exaggerated physical expressions of emotion that are often artificial and uncorrelated with the actual emotion being experienced by the individual.

In contrast, biosignal analysis, based on skin surface sensors worn by the user, may be a more robust and accurate means of determining emotion. This is because the signals correspond to internal physiology, largely related to the autonomous nervous and limbic systems, rather than to external expressions that can be manipulated easily. However, emotional state recognition by means of biosignals analysis is also problematic. This is due in part to the movement sensitivity of physiological sensors to such signals as electrocardiograms (ECG) and galvanic skin response (GSR). Muscle contractions are induced by electrical neural impulses, which in turn are picked up by the devices designed to measure differences in electrical potential. These may cause noise in the form of signal fluctuations. Furthermore, despite the evidence from psychophysiology suggesting a strong correlation between human emotional states and physiological responses (Watanuki et al., 2005; Cacioppo et al., 1990), determining an appropriate mapping between the two is nevertheless non-trivial.

Our interest in these techniques differs significantly from previous work. Rather than recording and classifying how people respond to external stimuli such as culturally meaningful images, sounds, film clips, and text, we are in the process of developing a biometrically driven

multimedia instrument, one that enables a performer to express herself with artistry and emotional cohesiveness. The goal is to provide a rich, external manifestation of one's internal, otherwise invisible, emotional state. With training, it is our hope that the resulting system, one that is coupled to the performer's emotional intentionality rather than to external gestures, can become as expressive and responsive as a fine musical instrument. Thus, rather than attempt to recognize and label human emotional states, our goal is to investigate the mapping of these states to expressive control over virtual environments and multimedia instruments.

From an artistic perspective, the instrument interface should support the articulation of emotion in a meaningful manner, with acuity and subtlety, allowing it be played with sensitivity and nuance.

We see the development of this instrument as a two-stage process. The first phase, described in this paper, deals with the question of emotion capture, that is, extracting meaningful data from the range of sensors available to us.

The second stage, which we discuss briefly in Section 5, relates these signals to the output of the instrument and how it is designed to be used in a performance setting. Because the instrument is ultimately a highly enriched biofeedback device, a performer's response to anything and anyone she encounters, including the audience, instantly manifests all around her. To bring it under her control, she must first *compose* herself. This involves using the instrument as a feedback device to return to a neutral state from which all emotions are potentially accessible. Once she has done so, she can put the instrument to its true use, directing her emotions outward in the act of creative composition.

The remainder of this paper is organized as follows. Our emotion elicitation method, used to gather the physiological data, is described in Section 3. Next, the recognition engine, including feature selection, reduction and classification, is described in Section 4. Finally, Section 5 concludes with a discussion of some future avenues for research.

2   RELATED WORK

Ekman's emotion classification scheme (Ekman, 2005) included six principal, discrete and universal classes of affect: anger, joy, fear, surprise, disgust and sadness. Russell's arousal/valence circumplex (Posner et al., 2005) introduced a continuous, analog mapping of emotions based on a weighted combination of arousal intensity and emotional valence (negative to positive). Figure 1 depicts this two-dimensional space with an example set of emotions.

For our purposes, both types of representations are useful for "playing" the *instrument* represented by the high-level schematic of Figure 2: discrete states serving as coarse control, with the analog input driving fine-tuned and subtle variations.
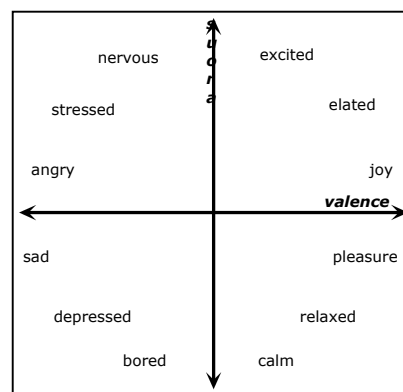

Figure 1: Russell's arousal/valence circumplex (reproduced from Posner et al., 2005)

Previous studies have demonstrated that emotional arousal and valence stimulate different brain regions (Anders et al., 2004) and in turn affect peripheral systems of the body. Significant physiological responses to emotions have been studied, showing, for example, measurable changes in heart rate and phalange temperature in fearful, angry and joyful states (Ekman et al., 1983).

Emotional state recognition using physiological sensors has been investigated by others. Picard (Picard et al, 2001) obtained good recognition results (81.25% accuracy) on eight emotions using one subject stimulated with personally selected images and four physiological sensors: blood volume pulse (BVP), galvanic skin response (GSR), electromyograph, and respiration). Our results, restricted to four emotions, are similar, but the critical difference between our approaches is the elicitation process. While Picard uses images to elicit emotion, we focus on an involved self-generation of affective states. This, we believe, has important implications for real-world theatrical performance, where emotions are continuously varying as opposed to discrete. Capturing the subtle dynamics of emotion is vital to attaining the cognitive and emotive skills required for mastering control of the instrument.
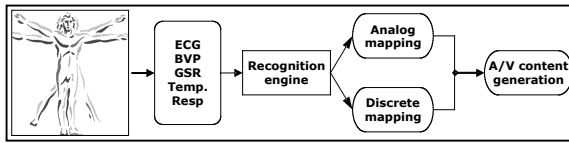
Figure 2: Biosignals-driven Emotional-Imaging Generator.

3   EMOTION ELICITATION

As noted above, we are primarily interested in how self-generated emotional states can be mapped through biosignal analysis to the proposed instrument. Clearly, the performer must be skilled in the art of accessing and articulating emotion. Just as with learning any musical instrument, feedback must be provided that connects her meaningfully both with the appropriate skill level and emotional experience.

As a first step in investigating these issues, we want to capture biosignal data of maximum possible validity. Gaining access to the ground truth of human emotion remains an elusive goal. Nevertheless, we can obtain a better labelled set of input than that available through generic stimuli, as used by other researchers. To do so, we interact directly with the experimental subject to *generate* the stimuli. This avoids the potential problems, articulated by colleagues, of subjects not responding to a particular stimulus as expected, or verbally expressing an emotion "they think the stimulus is supposed to evoke."

Of course, this necessitates that the stimulus be highly personalized and subjective. The benefit is the potentially greater physiological validity of the recorded data that is then used for training (or calibrating) our system. As seen in the results of Section 4, we succeed in obtaining an encouraging correct classification result over four emotions of 90%.

3.1   Experimental Subject

To maximize the validity of our experimental data, we worked with a professional method actor, who was guided by one of the authors (Deitcher), an experienced theatre director. Our subject has had the opportunity to methodically investigate an extraordinarily wide array of characters and situations. Effective emotional solicitation from someone with this kind of experience and flexibility requires the sensitivity to anticipate relevant emotional connections. It also requires the ability to ask the questions and define the exercises that will allow these emotions to emerge. In the broadest of terms, by having the actor play scenes, sing songs, follow guided visualizations and remember events from her own life, we were able to elicit a large and complex range of emotional landscapes. Her focused intentionality was responsible for engendering a high degree of confidence in the collected physiological data.

3.2   Experimental Data Collection

Experiments were conducted in a quiet, comfortable lab environment. The subject either remained seated or standing and was instructed to limit her body movement to minimize motion artefacts in the collected signals. The biosignals were recorded using Thought Technology's *ProComp Infiniti* biofeedback system using five sensor channels: GSR, ECG, BVP, phalange temperature and respiration, all sampled at 256 Hz. Each trial was also videotaped with a synchronization signal to align the video recording with the biosignals.

3.3   Data types

Two types of data were recorded: discrete emotional states and the responses to complex emotional scenarios. Typical trial times of 60 and 300 seconds were used for each type of data, respectively. A fifteen-minute break was taken between each trial so that the subject could return to her *baseline*, emotionally relaxed state.

The discrete class of data afforded a simple labelling of emotions, as expressed by the subject during each trial. These were used primarily for classifier training and validation. During these experiments, the subject was asked to experience four emotional states in turn (joy, anger, sadness, pleasure), while vocalizing what she was feeling. A post-trial questionnaire was used to determine a subjective assessment of the intensity of the sensed emotion, on a numeric scale from one to five. Twenty-five trials of each of the four emotions were recorded.

For the complex scenarios, data segments were recorded while the subject acted out "scenes" of fluid and varying emotional states. Such experiments will be used to study the body's psychophysiological responses during emotional transitions. These scenarios are theatrically dynamic, and thus meaningful in investigating the performance possibilities of our proposed instrument.

4 RECOGNITION ENGINE

Our preliminary investigations deal only with the classification of discrete emotional states to validate our paradigm of emotion elicitation, described in the previous section. The recognition engine comprises two main stages: biosignals processing and classification, both implemented in Matlab.

The emotional state recognition system utilizes five physiological signals: electrocardiogram (ECG), GSR, BVP, respiration and phalange temperature. We employ digital signal processing and pattern recognition, inspired by statistical techniques used by Picard. In particular, our use of *sequential forward selection* (a variant of sequential floating forward selection), as used by Picard, choosing only classifier-optimal features, followed by *Fisher dimensionality reduction*, are similar. For the classification engine, however, we implemented linear discriminant analysis rather than the maximum *a posteriori* used by Picard.

4.1 Biosignal processing

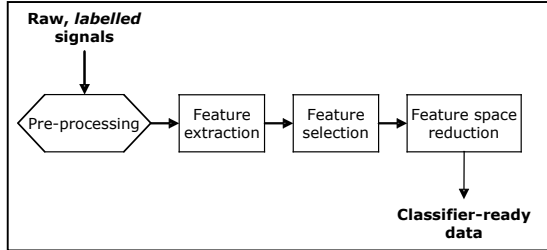The raw, discrete biosignals go through four steps to produce classifier-ready data, as shown in Figure 3.

Figure 3: Biosignal processing engine.

4.1.1 Pre-processing

Emotionally relevant segments of the recordings that are free of motion artefacts are hand-selected and labelled with the help of the video recordings and responses to the questionnaire. High-frequency components of the signals are considered to be noise and filtered with a Hanning window (Oppenheim, 1989).

4.1.2 Feature extraction

We extract six common statistical features from each type of the noise-filtered biosignals, of size N ($X_n, n \in [1...N]$), and its first and second derivatives:

1. Filtered signal mean:

$$\mu_X = \frac{1}{N} \sum_{n=1}^{N} X_n \qquad (1)$$

2. Filtered signal standard deviation:

$$\sigma_x = \left(\frac{1}{N-1} \sum_{n=1}^{N} (X_n - \mu_X) \right)^{1/2} \qquad (2)$$

3. Filtered signal mean of absolute value of the first difference:

$$\delta_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |X_{n+1} - X_n| \qquad (3)$$

4. Normalised signal mean of absolute value of the first difference:

$$\tilde{\delta}_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{X}_{n+1} - \tilde{X}_n| = \frac{\delta_x}{\sigma_x} \qquad (4)$$

5. Filtered signal mean of absolute value of the second difference:

$$\gamma_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |X_{n+1} - X_n| \qquad (5)$$

6. Normalised signal mean of absolute value of the second difference:

$$\tilde{\gamma}_x = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{X}_{n+1} - \tilde{X}_n| = \frac{\gamma_x}{\sigma_x} \qquad (6)$$

Where $\tilde{X}_n$ represents the normalised signal (zero-mean, unit variance):

$$\tilde{X}_n = \frac{X_n - \mu_x}{\sigma_x} \qquad (7)$$

In addition to the previous features, used for each biosignal, other signal-specific characteristics are computed. These include, for example, heart rate mean, acceleration/deceleration and respiration power spectrum at different frequency bands. Combining the statistical and signal-specific characteristics, a total of 225 features are thus computed from the five types of biosignals.

4.1.3 Automatic feature selection

Feature selection is a method widely used in machine learning to select a subset of relevant

features in order to build robust learning models. The aim is to remove most of the redundant and irrelevant features from the data to alleviate the often detrimental effect of high dimensionality and to improve generalization and interpretability of the model.

The greedy sequential forward selection (SFS) algorithm is used to form automatically a subset of the best n features from the original large set of m ($n < m$). SFS starts with an empty feature subset and on each iteration, exactly one feature is added. To determine which feature to insert, the algorithm tentatively adds to the candidate feature subset one that is not already selected and tests the accuracy of a k-NN classifier built on this provisional subset. A feature that results in the highest classification accuracy is permanently included in the subset. The process stops after an iteration where no feature addition causes an improvement in accuracy. The resulting feature set is now considered optimal.

The k-NN classifier used here classifies a novel object r by a majority of "votes" of its neighbours, assigning to r the most common class among its k nearest neighbours, using the Euclidean distance as metric. This type of classifier is chosen because it is a simple and efficient performance criterion for feature selection schemes and is considered more robust than using a single measure of distance, as is the case for many feature selection schemes. It was found through iterative experimentation using $k \in [1,9]$, that a value of $k = 5$ resulted in the best possible selected feature subset.

4.1.4 Feature space reduction

Fisher dimensionality reduction (FDR) seeks an embedding transformation such that the between-class scatter is maximized and the within-class scatter is minimized, resulting in a low-dimension representation of optimally clustered class features. FDR is shown to produce optimal clusters using $c - 1$ dimensions, where c is the number of classes. However, if the amount of training data or the quality of the selected feature subset is questionable, as is the case in many machine learning applications, the theoretically optimal dimension criterion may lead to an irrelevant projection which minimizes error in the training data, but performs badly with testing data (Picard et al., 2001). In our case, a two-dimensional projection resulted in an overall best classification rate using linear discriminant analysis (LDA) to sequentially test with dimensions $d \in [1,3]$. Figure 4 demonstrates the class clustering of four emotional states: joy, anger,

sadness, pleasure (JO, AN, SA, PL), projected on a 2D Fisher space during one of the validation steps. The four emotions were chosen given that they lie in different quadrants of Russell's arousal/valence circumplex (Figure 1).
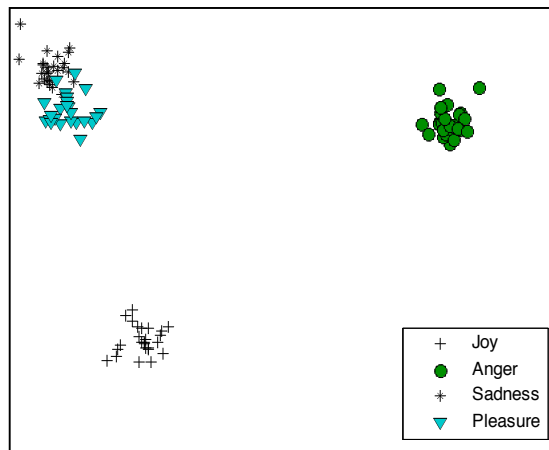

Figure 4: 2D Fisher projection (4 classes)

4.2   Biosignal classification

Three popular classification schemes were tested to classify the four emotional states: LDA, k-NN ($k \in [1,9]$) and multilayer perceptron (MLP). LDA was found to outperform both the best k-NN (k = 7) and MLP by 4% and 11%, respectively. LDA builds a statistical model for each class and then catalogues novel data to the model that best fits. We are thus concerned with finding which discriminant function best separates the emotion classes. LDA finds a linear transformation Φ of the x and y axes (8) that yields a new set of values providing an accurate discrimination between the classes. The transformation thus seeks to rotate the axes with parameter v so that when the data is projected on the new axes, the difference between classes is maximized.

$$\phi = v_1 \mathbf{x} + v_2 \mathbf{y} \qquad (8)$$

Due to the small feature dataset size, leave-one-out cross-validation was used to test the classification scheme. This involves using a single item of the set as the validation data, and the remaining ones as training data. This process is repeated until each item in the dataset is used once as the validation data. At each iteration, SFS and FDR are applied to the new training set and the parameters found (selected features and Fisher projection matrix) are applied to the test set. The

mean classification rate is computed using the result produced at each step. Using this method, our biosignal classification system produced an average recognition rate of 90% on the four emotional states. Table 1 shows the confusion matrix for the classification.

Table 1: LDA classifier confusion matrix

I/O	JO	AN	SA	PL	%
JO	**0.96**	0	0	0.04	**96**
AN	0	**1.00**	0	0	**100**
SA	0.04	0	**.92**	0.04	**92**
PL	0.12	0	0.16	**0.72**	**72**

5   CONCLUSIONS

A novel emotion elicitation scheme based on self-generated emotions is presented, engendering a high degree of confidence in collected, emotionally relevant, biosignals. Discrete state recognition via physiological signal analysis, using pattern recognition and signal processing, is shown to be highly accurate. A correct average recognition rate of 90% is achieved using sequential forward selection and Fisher dimensionality reduction, coupled with a Linear Discriminant Analysis classifier.

We believe that the high classification rate is due in part to our use of a professional method actor as test subject. It is speculated that normal subjects would lead to lower rates because of the high variability of emotion expressivity across a large population pool. It is an avenue of research for us to test the generalization of this type of machine-based emotion recognition.

Our ongoing research also intends to support *real-time* classification of discrete emotional states. Specifically, continuous arousal/valence mappings from biosignals will drive our emotional-imaging generator for multimedia content synthesis and control in a theatrical performance context. In addition, we are exploring the therapeutic and performance training possibilities of our system. Because what we are building is fundamentally an enriched biofeedback device, we anticipate applications ranging from stress reduction for the general population to the generation of concrete emotional expression for those with autism or other communication disorders.

REFERENCES

Anders S., Lotze M., Erb M., Grodd W., Birbaumer N., 2004. *Brain activity underlying Emotional valence and arousal: A response-related fMRI study*. Human Brain Mapping, Vol. 23, p. 200-209.

Bartlett, M.S., Hager, J.C., Ekman, P., Sejnowski, T.J., 1999. *Measuring facial expressions by computer image analysis*. Psychophysiology, Vol. 36, p. 253-263.

Black, M.J., Yacoob, Y., 1995. *Recognizing facial expressions in image sequences using local parameterized models of image motion*. ICCV.

Cacioppo, J., Tassinary, L.G., 1990. *Inferring psychological significance from physiological signals*. American Psychologist, Vol 45, p. 16-28.

Ekman, P., Levenson, R.W., Friesen, W.V., 1983. *Autonomic Nervous System Activity Distinguishes Between Emotions*. Science, 221 (4616), p. 1208-1210.

Ekman P., 2005. *Emotion in the human face*, Cambridge University Press, p. 39-55.

Lyons, M. Budynek, J., Akamatsu, S. 1999. *Automatic Classification of Single Facial Images*. IEEE PAMI, vol. 21, no. 12.

Oppenheim, A.V., Schafer, R.W., 1989. Discrete-Time Signal Processing, Englewood Cliffs, N.J.: Prentice-Hall

Picard, R.W., Vyzas, E., Healey, J., 2001. *Toward machine emotional intelligence: analysis of affective physiological state*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 23, Issue 10, p. 1175 – 1191.

Posner J., Russell J.A., Peterson B.S., 2005. *The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology*. Development and Psychopatholy, p. 715-734.

Ververidis, D., Kotropoulos, C., Pitas, I., 2004. *Automatic emotional speech classification*, IEEE ICASSP.

Watanuki S., Kim Y.K., 2005. *Physiological responses induced by pleasant stimuli*. Journal of Physiological Anthropology and Applied Human Science, p. 135-138.