# Classification of Emotional Arousal During Multimedia Exposure

Adam Anderson
Dept. of Computer Science
University of Maryland
College Park, MD 20742, USA
aander21@umd.edu

Thomas Hsiao
Dept. of Statistics
Rice University
Houston, TX 77005, USA
twh4@rice.edu

Vangelis Metsis
Dept. of Computer Science
Texas State University
San Marcos, TX 78666, USA
vmetsis@txstate.edu

## ABSTRACT

In the study of emotion recognition, relatively few efforts have been made to compare classification results across different emotion induction methods. In this study, we attempt to classify emotional arousal using physiological signals collected across three stimulus types – music, videos, and games. Subjects were exposed to relaxing and exciting music and videos and then asked to play Tetris and Minesweeper. Data from GSR, ECG, EOG, EEG, and PPG signals were analyzed using machine learning algorithms. We were able to successfully detect emotion arousal over a set of contiguous multimedia activities. Furthermore, we found that the patterns of physiological response to each multimedia stimuli are varying enough, that we can guess the stimulus type just by looking at the biosignals.

## CCS Concepts

•Human-centered computing → Empirical studies in ubiquitous and mobile computing; *Empirical studies in HCI;*

## Keywords

Physiological biosignals, emotion recognition, classification, affective computing

## 1. INTRODUCTION

The classification of human emotional states from physiological signals is an active area of research. It has been shown previously that physiological signals can be reliable indicators of affective states [5]. Presently, however, emotion recognition is done mostly in laboratory settings. Widespread applicability of emotion classification techniques based on biosignals would increase dramatically if a robust framework for emotion recognition were developed using a small number of noninvasive sensors.

Computers' ability to recognize emotions in humans can be applied to the domain of human-computer interaction to develop more nuanced communication between computers and users. Computer agents capable of recognizing different emotional states have the ability to react contextually based on how a user is feeling, potentially improving user experience in a variety of applications. One possible use of emotion recognition is in the development of more effective intelligent teaching systems [8]. An emotionally-aware digital teacher would have the ability to recognize frustration or boredom in a student, signaling that it slow down or accelerate the pace of learning accordingly.

Additionally, a reliable emotion recognition scheme could open the door to standardization in the study of emotion theory. This would enable studies improving the diagnosis and treatments of mood disorders, like depression or post-traumatic stress disorder (PTSD).

Studies of emotion recognition often differ in their approaches to labeling affective states, with some using discrete categories [9] while others use continuous spectra. One popular model of emotion uses dimensions of valence (how pleasant the emotion is) and arousal (how strongly the emotion is felt) [6]. One benefit of using a dimensional model is that no specific categories for emotions need to be hypothesized a priori.

One open question in emotion research is whether physiological responses are consistent when different modalities are employed to evoke a particular emotion [10]. Previous studies have separately considered emotion recognition using music [10, 12], images [6, 8], videos [9] and games [4]. In each of these studies, emotional states were successfully discriminated, but no effort was made to compare classification results across different stimulus types.

In this study, we assume the valence-arousal model and attempt to discriminate low and high arousal emotional states using a combination of multimedia stimuli, including music, videos, and computer games. Based on a set of collected physiological biosignals, we are able to successfully discriminate between states of high and low arousal. In addition, we show that we can recognize what type of stimuli was the subject exposed to, just by looking at the collected biosignals. Our methodology follows the standard supervised learning process of feature extraction, feature selection, training and testing, with special attention on the feature extraction stage.

In the following sections, we first explain our experimental procedure and elaborate on our methodology. Subsequently, we present our experimental results. Finally, we conclude with a summary and the benefits of this study.
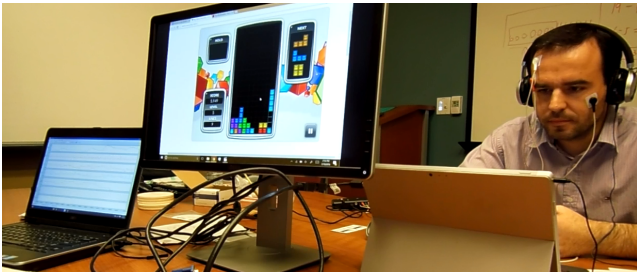
Figure 1: Subject playing Tetris during the experiment. The laptop on the left is running the BioCapture software to record physiological signals while the subject uses a different laptop to play.

## 2. EXPERIMENTAL PROCEDURE

To evoke relaxed and excited emotional states, subject were presented a series of music and video clips. Immediately afterwards, subjects played Tetris and Minesweeper using a keyboard and touchpad. When music was presented, subjects were asked to keep their eyes closed. Five male volunteers participated in the experiment. Figure 1 shows the setup while a subject is playing Tetris.

The order in which the stimuli were presented was as follows:

1. Relaxing Music - *Living Transparent* by Reasoner
2. Relaxing Video - Mountain Stream Video by 4K
3. Exciting Music - *Good Feeling* by Flo Rida
4. Exciting Video - Club Scene from *John Wick* movie
5. Tetris ($\sim$ 4.5 minutes on average)
6. Minesweeper ($\sim$ 4.5 minutes on average)

Music and video stimuli were presented consecutively with a 35-second clip of a flowing creek shown in between each, to bring the subject back to a neutral emotional state. The subjects played each game until their first loss. Stimuli were deliberately presented in order from relaxing to exciting because it was assumed that it would take less time for subjects to become excited when relaxed than it would for subjects to calm down to a relaxed state after being excited.

### 2.1 Data Collection

Biosignals were recorded at a sampling frequency of 250 Hz, from 5 male volunteers of varying ages, using the Great Lakes Neurotech BioRadio [1] device with adhesive pre-gelled electrodes and the pulse oximeter attachment. The following five signals were considered:

**Galvanic Skin Response (GSR):** GSR measures the electrical resistance of the skin, and has been noted to be highly indicative of changes in arousal [7]. Electrodes were placed on the left hand, on the middle and index finger.

**Electrocardiogram (ECG):** ECG measures the changes in electrical activity of the heart that result from heartbeats. One electrode was placed on the left wrist, and the other on the right wrists

**Electroencephalogram (EEG):** EEG measures electrical activity in the brain can be measured through electrodes placed on the scalp. The 10-20 system is a configuration of 64 electrodes on the head to place electrodes for a complete picture of the electrical activity of the brain [14]. In order to maintain a minimal number of electrodes, only one position was selected. The F4 spot located at the top of
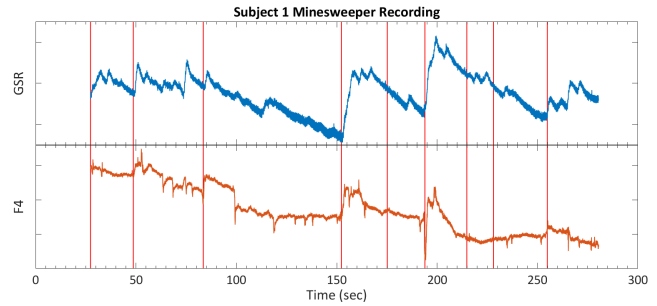


Figure 2: A sample from a subject's GSR and EEG (channel F4) recordings during the Minesweeper session. The vertical markers represent points in time where the subject clicked on a bomb and lost the game. There appear to be some noteworthy spikes in both signals in response to losses.

the right forehead was selected. Electrodes in the frontal lobe have been found to be important in the measurement of emotional activity [11, 13]. Furthermore, a study by Liu et al used an electrode in the F4 position as one of only three EEG electrodes used to recognize emotions [12].

**Electrooculogram (EOG):** EOG measures the difference in electrical potential between the front and back of the eye. The cornea is more positively charged and the retina more negatively charged, creating an electric field that changes as the eye moves [3]. The original purpose of collecting EOG data was to facilitate artifact removal in the EEG signal. However, after some visualization of the signal, EOG features seemed to contain useful information for arousal classification.

One electrode was placed above/to the right of the right eye, and another was placed below/to the left of the left eye.

**Photoplethysmogram (PPG):** PPG measures changes in blood flow in the body. A pulse oximeter collected PPG data from the index finger.

### 2.2 Questionnaire Analysis

Following the experiment, subjects were asked to rate their perception of emotions experienced following exposure to each stimuli. Two questions were asked for each stimulus. The first question asked the subject how strong of an emotion they felt after a particular stimulus on a scale of 1-5, 1 being "Very slightly to none at all" and 5 being "Extremely." The second question asked the subject how positive they felt, 1 being "Negative" and 5 being "Positive." On the arousal-valence scale, the relaxed emotion is considered a moderately high valence and low arousal emotion, while the excited emotion is a moderately high valence and high arousal emotion. A two-sample t-test was performed between the survey results of the relaxed stimuli and those of the excited stimuli, and a statistically significant difference between the arousal ratings at the 0.05 level was found. This result validates the assumption that the relaxing stimuli evoked relaxed feelings in the subjects, and the same for the exciting stimuli.

An interesting fact is that in certain situations, simple visual inception of the biosignals revealed interesting patterns of emotional arousal, e.g. see Figure 2.

# 3. CLASSIFICATION

After the data was collected, two off-line experiments were performed:

1. Arousal Classification - distinguishing between relaxing or exciting stimuli
2. Stimulus Classification - distinguishing stimulus type (music, video, game)

Data was labeled according to the stimulus type, and self-reports of arousal levels. Feature extraction was performed followed by feature selection. The final subset of features were then used to train models and generate predictions. Models were validated using leave-one-subject-out cross validation.

## 3.1 Preprocessing

A linear detrend was applied to all signals measured to normalize each signal around a mean of zero. The normalized mean should correct for differences in sensor placement and skin conductivity profile across different subjects, while the linear detrend should remove an drift in the signal accumulated during the recording.

To give physiological responses enough time to manifest following exposure to a given stimulus, the last 30 seconds of data collected while the subject was exposed to a a particular stimulus were used for classification. This also ensured that there were an equal number of samples from music listening, video watching, and gaming.

Data was split into segments of a fixed length of 20 seconds with a 16-second (80%) overlap between segments. Different values for segment length and overlap were tested, as well as different lengths of time from the end of the signal from which data should be used. The above values were found to yield the greatest accuracy. It is hypothesized that long time slices yielded greater accuracy because physiological signals are subject to instantaneous fluctuations that may result in incorrect classifications of short segments. Longer segments allow the classifier to see a trend in the biosignals that emerges over a longer period of emotional experience. This is an avenue for future research.

## 3.2 Feature Extraction

A total of 98 features were extracted from the data, including features in both the time and frequency domains. Time domain features included included variance, RMS, skewness, kurtosis, and the mean of the first and second derivatives [7]. Frequency domain features included spectral centroid and spectral spread.

Power in the delta (1-3Hz), theta (4-7Hz), alpha (8-12Hz), beta(13-30Hz), and gamma (31-50Hz) frequency bands was calculated using the EEG signal. Feature selection concluded that power in the delta band was useful for both arousal and stimulus classification while gamma band energy was specifically useful for arousal classification while theta band energy was useful for classifying arousal.

Excepting EEG frequency band powers, features were extracted from all channels.

## 3.3 Feature Selection

A wrapper-based sequential forward selection was performed to identify a key subset of features for both experiments.

Originally, a total of 98 features were extracted from the 5 signal channels. Feature selection reduced the number of features to 21 for arousal classification and to 26 for stimulus classification. A total of 9 features were shared between the two classification experiments.

The selected features are summarized as follows:

**Common Features:** *ECGSpecCentroid, ECGEnergy-Entropy, ECGHjorthMobility, EOGSkewness, F4Delta, F4FirstDiff, F4HjorthComplexity, GSRSpectralEntropy, PPGSpectralRolloff.*

**Arousal Classification:** *ECGHjorthComplexity, ECG-Kurtosis, ECGSpecSpread, ECGVariance, EOGFirstDiff, EOG-SpecCentroid, F4EnergyEntropy, F4Gamma, GSREnergyEntropy, GSRVariance, PPGKurtosis, PPGSkewness.*

**Stimulus Classification:** *ECGSecDiff, ECGSpectralEntropy, ECGZeroCrossRate, EOGSpecCentroid, EOGKurtosis, EOGSpecSpread, EOGSpecEntropy, F4Skewness, F4Theta, F4Variance, GSRFirstDiff, GSR-RMS, GSRSecDiff, GSRSpecCentroid, GSRSpecSpread, PPGZeroCrossRate.*

A noteworthy observation about the feature sets selected for both classification experiments is the prevalence of ECG based features. Of the features selected, 9 were extracted from ECG, with the next most useful signal being EOG with 7 useful features across both experiments.

One interesting difference between the arousal and stimulus classification feature sets is that while delta band power from the F4 EEG electrode was found useful for both, the gamma band power was found useful for arousal classification while theta band power was found useful for stimulus classification. In order to validate that one frequency band really contains more useful information about a particular activity, however, future experimentation must be done with more sensitive EEG equipment and more electrodes.

## 3.4 Classification Algorithms

Previous studies have used different classification algorithms to discriminate between emotions. Some have found success using KNN classifiers, decision trees, and SVM classifiers, e.g. [8, 11]. However, there is no current standard algorithm or feature selection procedure in the domain of emotion recognition [2]. As a result, four different classification algorithms were used for both arousal and stimulus classification. MATLAB implementations of the following four algorithms were used: Medium KNN, Complex Tree, Medium Gaussian SVM, and Bagged Trees.

Models were trained and validated using a 5-fold leave one subject out cross validation scheme. Each model was trained using data from four out of the five subjects, and then tested on the data from the remaining subject. This ensured that testing results were not inflated due to the models being overfit to idiosyncrasies in the signals for a particular subject.

# 4. RESULTS

Overall accuracy and recall for arousal and stimulus classification are shown in Table 1 and Table 2, respectively. The reported statistics are for subject independent classification, meaning that during both training and testing, the model did not know what data had been collected from which participant.

Subject independent arousal classification was achieved with 88.9% accuracy and an average recall of 83.3%. Discrimination between the relaxing and exciting videos was achieved with 100% accuracy while accuracy of 78.8% was achieved between the different musical stimuli.

**Table 1: Confusion matrix for Arousal Classification using SVM. Classification accuracy was 88.9%**

| True Class | Predicted Class | | Recall |
|---|---|---|---|
| | Relaxed | Excited | |
| Relaxed | 16 | 8 | 66.67% |
| Excited | 0 | 48 | 100% |

**Table 2: Confusion matrix for stimulus classification using bagged trees. Classification accuracy was 80.6%**

| True Class | Predicted Class | | | Recall |
|---|---|---|---|---|
| | Music | Video | Game | |
| Music | 24 | 0 | 0 | 100.00% |
| Video | 5 | 14 | 5 | 58.33% |
| Game | 0 | 4 | 20 | 83.33% |

It is noteworthy that arousal classification for emotions induced from videos appears to be better than classification for emotions induced from music. There are multiple possible explanations, which present directions for future research.

One possible reason for better classification accuracy of videos is because of salient features extracted from the EOG signal. As was mentioned earlier, subjects had their eyes closed while listening to music, so any features from the EOG signal would likely not be registered during music listening. We see that feature set useful for arousal classification includes the first difference of the EOG signal (EOG-FirstDiff) and the spectral centroid of the EOG signal (EOG-SpecCentroid). Another possible explanation could be that for the specific music and video selections presented, the video clips elicited a stronger emotional response in participants, resulting in more pronounced differences in extracted features. In order to correct for this, future studies may need to present multiple different samples of music and video for each emotional class.

Stimulus classification was achieved with 80.6% accuracy and an average recall of 80.6%.

## 5. CONCLUSION

Our findings show that biosignal-based emotion classification is possible. Consumer grade wearables already exist capable of measuring GSR, PPG and other types of physiological biosignals, and there are some relatively affordable EEG headsets available in the market. EEG and EOG may be able to be embedded in a single head set. Future research could look into using these consumer grade sensors to advance the ways humans interact with computers in their everyday lives. Such research could also consider using other signals these devices collect, such as accelerometer and gyroscope data.

This study focused on classifying arousal level of a subject after exposure to stimuli intended to evoke a particular emotional state. In the future, research is needed to classify valence according to a minimum number of signals.

## 6. ACKNOWLEDGMENTS

## 7. REFERENCES

[1] GLNT BioRadio physiological monitor. https://glneurotech.com/bioradio/.

[2] ARROYO-PALACIOS, J., AND ROMANO, D. Towards a standardization in the use of physiological signals for affective recognition systems. In *Measuring Behavior* (2008).

[3] BULLING, A., WARD, J. A., GELLERSEN, H., AND TROSTER, G. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence 33*, 4 (April 2011), 741–753.

[4] CHANEL, G., REBETEZ, C., BETRANCOURT, M., AND PUN, T. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 41*, 6 (Nov 2011), 1052–1063.

[5] EKMAN, P., LEVENSON, R. W., AND FRIESEN, W. V. Autonomic nervous system activity distinguishes among emotions. *Science 221*, 4616 (1983), 1208–1210.

[6] HAAG, A., GORONZY, S., SCHAICH, P., AND WILLIAMS, J. *Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 36–48.

[7] HEALEY, J. A., AND PICARD, R. W. Detecting stress during real-world driving tasks using physiological sensors. *Trans. Intell. Transport. Sys. 6*, 2 (June 2005), 156–166.

[8] HERAZ, A., RAZAKI, R., AND FRASSON, C. Using machine learning to predict learner emotional state from brainwaves. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (July 2007), pp. 853–857.

[9] JANG, E. H., PARK, B. J., KIM, S. H., CHUNG, M. A., PARK, M. S., AND SOHN, J. H. Emotion classification based on bio-signals emotion recognition using machine learning algorithms. In *Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on* (April 2014), vol. 3, pp. 1373–1376.

[10] KIM, J., AND ANDRE, E. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence 30*, 12 (Dec 2008), 2067–2083.

[11] LIN, Y. P., WANG, C. H., JUNG, T. P., WU, T. L., JENG, S. K., DUANN, J. R., AND CHEN, J. H. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering 57*, 7 (July 2010), 1798–1806.

[12] LIU, Y., SOURINA, O., AND NGUYEN, M. K. Real-time eeg-based human emotion recognition and visualization. In *Cyberworlds (CW), 2010 International Conference on* (Oct 2010), pp. 262–269.

[13] RODRIGUEZ, A., REY, B., CLEMENTE, M., WRZESIEN, M., AND ALCANIZ, M. Assessing brain activations associated with emotional regulation during virtual reality mood induction procedures. *Expert Systems with Applications 42*, 3 (2015).

[14] TRANS CRANIAL TECHNOLOGIES. *10/20 System Positioning Manual*, 1.0 ed. 2410 Fortis Tower 77-79 Gloucester Road Wanchai, Hong Kong, 2012.