Classification of Emotional Arousal During Music Listening, Video Watching, and Gaming

Adam Anderson[1], Thomas Hsiao[2], Dr. Vangelis Metsis[3]
[1]A. James Clark School of Engineering, University of Maryland, College Park, MD, USA
[2]George R. Brown School of Engineering, Rice University, Houston, TX, USA
[3]Department of Computer Science, Texas State University, San Marcos, TX, USA

Introduction

Background: The classification of human emotional states from physiological signals is an active area of research. Widespread applicability of such classification would increase dramatically if a robust framework for emotion recognition were developed using a small number of noninvasive sensors.

Importance: Computers' ability to recognize emotions in humans can be applied to the development of more effective intelligent teaching assistants[1], and possibly to the diagnosis of mood disorders.

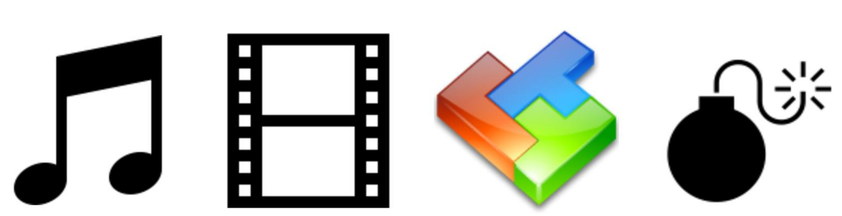
Purpose: A popular paradigm in the field of emotion recognition is the decomposition of emotions into dimensions of valence and arousal[2]. In this experiment, we focus on arousal classification across different stimuli while attempting to minimize the number of sensors used.

Experimental Procedure

Experiment: Subjects were exposed to music and video clips intended to elicit relaxation or excitement. A brief video of a flowing creek was shown between each stimulus to bring the subject back to a neutral emotional state. Each subject then played a game of Tetris followed by a game of Minesweeper. After exposure to the stimuli, subjects answered a questionnaire rating them on five point scales of valence and arousal.

Stimuli:

- Relaxing Music (3 mins)
- Relaxing Video (4 mins)
- Exciting Music (3 mins)
- Exciting Video (6 mins)
- Tetris (5 mins)
- Minesweeper (5 mins)



Data Experiments: Our aim was to classify between two groups:

- Differentiating among the three stimulus types (music, video, games),
- Binary classification between low and high arousal stimuli
  - Low Arousal: Relaxing Music, Relaxing Video
  - High Arousal: Exciting Music/Video, Tetris, Minesweeper

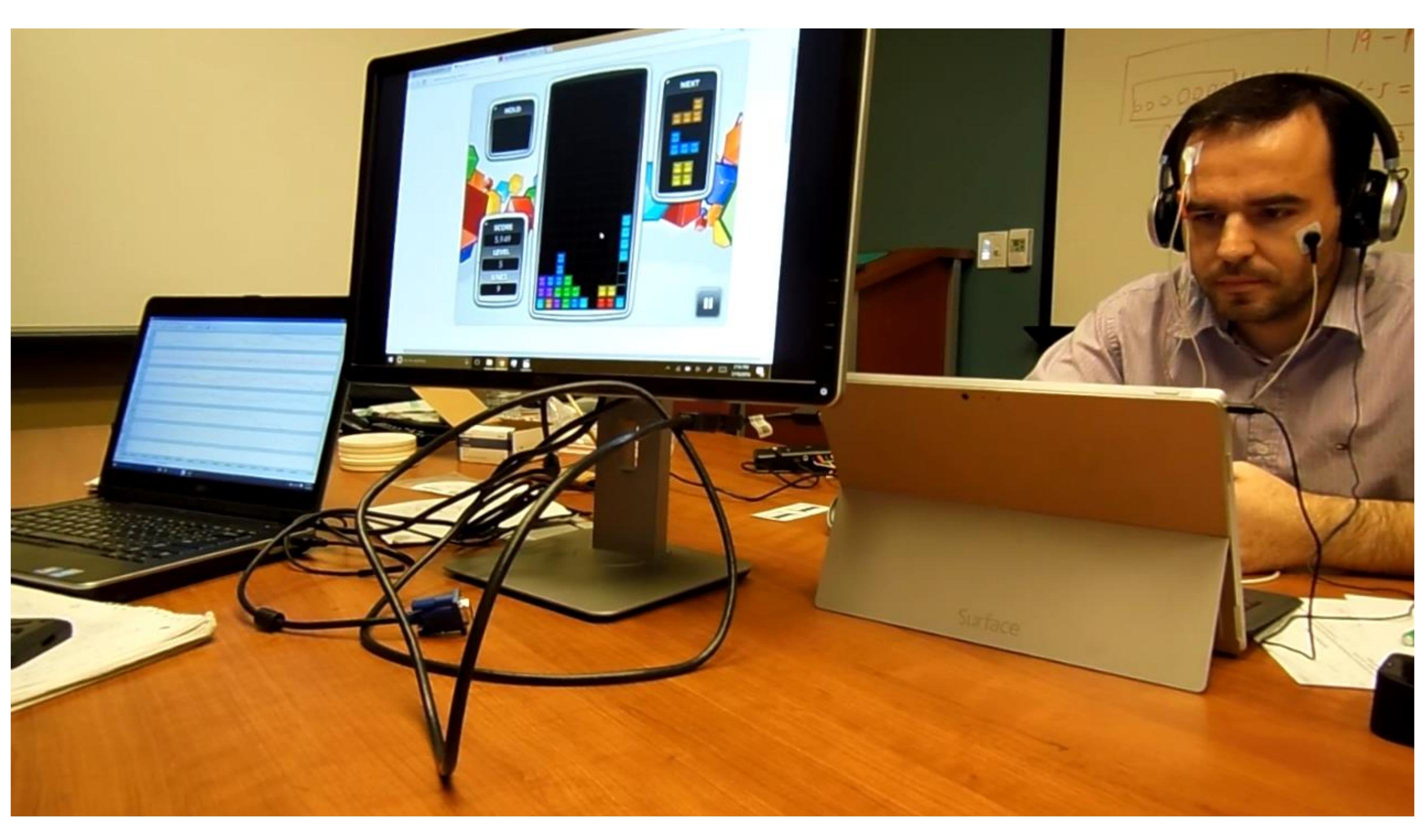


Figure 1. Experimental setup while a subject plays Tetris

Acknowledgements

This research is supported by Texas State University and the NSF REU Program through REU award #1358939.

Physiological Signal Analysis

Recording Setup: Five biopotential channels were recorded using Great Lakes Neurotechnology's BioRadio[3]. Signals measured include:

- Galvanic Skin Response (GSR)
- Electroencephalogram (EEG) at the F4 position
- Electrocardiogram (ECG)
- Electrooculogram (EOG)
- Photoplethysmogram (PPG)

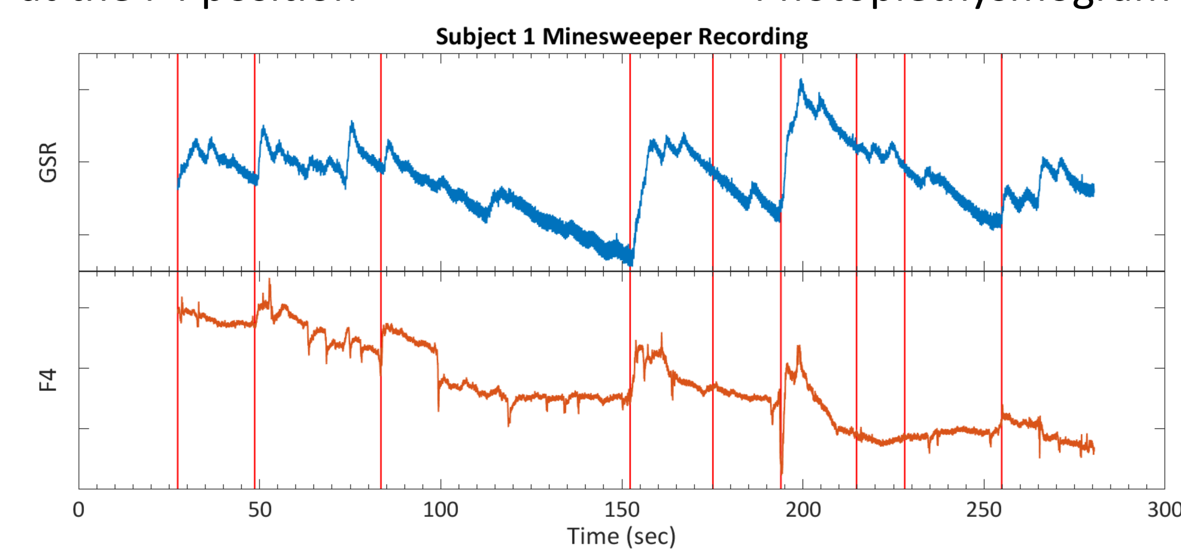


Figure 2. GSR and EEG signals recorded while a subject played Minesweeper. Vertical markers note when the subject lost a game.

Processing: Data was linearly detrended to normalize across subjects and remove drift. Features were extracted from 20 second segments of the signals using a sliding window with 80% overlap. Segments were taken starting 30 seconds from the end of the signal to allow physiological responses to the stimuli to manifest. Processing was done using the MATLAB software by Mathworks.

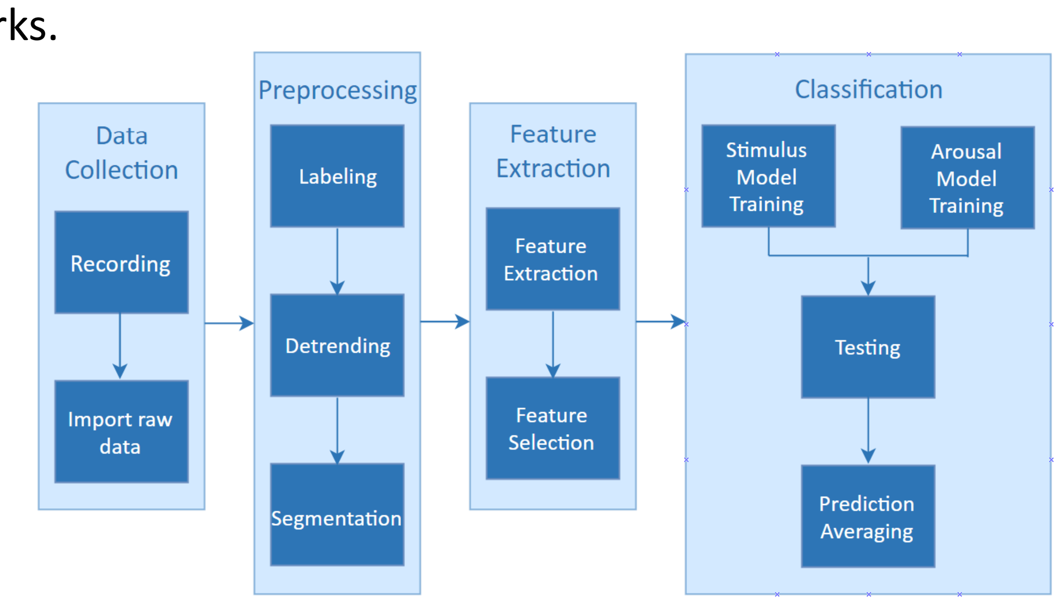


Figure 3. Data processing workflow

Feature Extraction: A total of 98 features were extracted. Different models were trained using four different algorithms (ComplexTree, KNN, SVM, BaggedTrees). Sequential forward selection was applied to identify the most important subset of features. Below are a few of the features useful for particular experiments.

- **Stimulus classification** – EOG spectral centroid, GSR mean of first and second derivatives, EEG energy in the theta frequency band
- **Arousal classification** – ECG average frequency, variance and kurtosis, GSR variance, EEG energy in the gamma frequency band

References

1. A. Heraz, R. Razaki and C. Frasson, "Using machine learning to predict learner emotional state from brainwaves," *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, Niigata, 2007, pp. 853-857.
2. J. Kim and E. André, "Emotion recognition based on physiological changes in music listening," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 12, pp. 2067-2083, Dec. 2008.
3. "Wireless Physiological Monitoring". *BioRadio: Wireless & Wearable Human Monitoring*. N.p., 2016. Web. 27 July 2016.

Results

Testing: Models were tested using leave one subject out cross validation. Performance metrics were averaged across the testing sets, and the best algorithm was selected based on the average accuracy. A sliding window prediction averaging scheme was applied to testing data.

True Class \ Predicted Class	Music	Film	Game	*Recall*
Music	24	0	0	100.00%
Film	5	14	5	58.33%
Game	0	4	20	83.33%
Precision	82.76%	77.78%	80.00%	Accuracy: 80.56%

Figure 4. Confusion Matrix of the Stimulus Classification. Best results were achieved using a support vector machine (SVM) with an overall accuracy of 80.56%. Note the 100% recall for predicting music.

True Class \ Predicted Class	Relaxed	Excited	*Recall*
Relaxed	16	8	66.67%
Excited	0	48	100.00%
Precision	100.00%	85.71%	Accuracy: 88.89%

Figure 5. Confusion Matrix of the Arousal Type Classification. Best results were achieved using a Bagged Trees ensemble classifier with an overall accuracy of 88.89%. Note that 100% recall was achieved in classifying an excited state.

Additional Results: In addition to the experiments shown on the above confusion matrices, classification of arousal was also performed between relaxing and exciting music as well as relaxing and exciting video. Accuracy of 75% was achieved for music discrimination and 100% accuracy was achieved for video classification.

Conclusion

Discussion of Results:

- Achieved reasonable arousal recognition using only five sensors
- Prediction accuracy can be improved with stimulus dependent classification
- Large overlap between data segments allow potential for workflow to be applied to real time classification

Future Work

- Replication of results using commercially available biosensors
- Pursue a reliable way to classify valence while engaging in similar activities with a comparable accuracy