

Augmenting EEG with Inertial Sensing for Improved 4-Class Subject-Independent Emotion Classification in Virtual Reality

Jason Teo¹, Nazmi Sofian bin Datuk Suhaimi², James Mountstephens³

jtwtteo@ums.edu.my¹

Faculty of Computing & Informatics, Universiti Malaysia Sabah
Kota Kinabalu, Sabah, Malaysia. ^{1,2,3}

Abstract : This investigation reports on the promising results obtained from the novel use of inertial sensing data for augmenting electroencephalography (EEG)-based subject-independent classification of emotions generated by virtual reality stimuli in four classes. 31 users were shown various virtual reality scenes to elicit responses in the four-quadrant emotional space according to Russell's Circumplex Model of Affect. Prior studies in emotion classification can be broadly grouped according to two main delineations of investigative methods: (1) whether the classification is binary (i.e. two-class classification) or otherwise (e.g. three-class, four-class classification or more) and (2) whether the training and testing occurs within the same participant (also known as subject-dependent classification) or across different participants (also known as subject-independent classification). Due to the significantly higher level of difficulty in conducting ternary/quaternary, subject-independent classification, the large majority of emotion modeling studies that report high accuracy rates adopts the binary, subject-dependent approach to classification. However, this study attempts the more challenging four-class classification, subject-independent classification. The EEG signals, accelerometer, and gyroscopic data were acquired through a wearable brain-computer interface device called Muse. Raw as well as power spectral density features were extracted from the EEG signals, and together with the first known use of inertial sensing data for emotion classification, were used as input to a deep neural network. Classification results show that without inertial sensing data, inter-subject classification was indeed highly challenging even for a deep learning system with only slightly better than random for 4-class classification at 26-27%. However, the augmentation of inertial sensing data improved the classification accuracy to 40-47%. As such, this work demonstrates the potential of using inertial sensing as an additional modality for emotion modeling.

Keyword : Inertial sensing, emotion classification, virtual reality, 4-quadrant emotion recognition, electroencephalography.

1 Introduction

The next big field of artificial intelligence research is touted to lie within the still nascent domain of emotion engineering [1]. The premise of emotion engineering is that using emotion recognition, made possible through wearable sensors such as those now commonly found on numerous brands of smartwatches, our moods can be regulated. For example, if a worker is detected to be experiencing stress via emotion recognition through sensors on his or her smartwatch, an app on their smartphone could play some relaxing music through their earphones to help alleviate some of the worker's stress.

Emotion classification is typically conducted in a binary fashion, i.e. whether the arousal level is high or low, or whether the valence is positive or negative [2]. Although some useful modeling of user emotions is achieved, nonetheless such classification has rather obvious limitations. For example, was a high arousal detected as a result of a happy or a horrific encounter, or was a negative valence detected as a result of being bored or frightened? As such

rather than binary classification, a four-class classification into the four distinct quadrants of happy, scared, bored and calm provides more specific and useful information compared to two-class classification.

Moreover, the generalizability of most emotion classification studies are limited to the modeling being specific to the user. In most cases, the trained emotion classifier can only be used to detect future emotions of the same user and cannot be readily transferred for use on another user. This is known as subject-dependent emotion classification [2]. A more useful approach particularly if such technology is to be widely deployed for everyday use in emotion engineering applications where users may not be bothered or have time to train their own emotion classifier for use but rather an off-the-shelf, ready-to-use emotion recognition device, would be subject-independent emotion classifier.

The current approaches to emotion recognition involve acquisition of sensory data from different modalities including facial expression, heart rate, galvanic skin response and brainwaves [3,4]. Brainwave data or electroencephalography (EEG) is typically the method of choice for multi-class emotion classification due to its richer set of data acquired through the multiple electrodes located on the skull of the user. Nonetheless, even using EEG data, four-class, subject-independent emotion classification is still extremely challenging.

Inertial sensors refer to the range of sensors that rely on the principle of inertia to conduct its measurements. Two very common forms of inertial sensors are accelerometers and gyroscopes. Accelerometers measure linear acceleration along a particular axis where the most common approach measures across the three dimensions of x,y,z space. Gyroscopes on the hand indicate the orientation in the x,y,z three-dimensional space. Inertial sensors particularly accelerometers and gyroscopes that are commonly found in smartphones and other portable devices are fast becoming the sensors of choice for a wide range of mobile applications including health and fitness tracking apps as well as mobile games [5].

This paper represents the first known report on the use of inertial sensing for emotion classification from stimuli delivered via an immersive virtual reality (VR) environment. This first section has provided an overview of the investigation; the second section will provide a brief literature review of relevant work; the third section will explain the methods adopted in this study; the fourth section will discuss and analyze the outcomes of the data collected from this investigation; the fifth section will conclude this report and provide some brief future lines of investigations.

2 Background

Emotion classification is commonly conducted using still images, video, and music or some combination such as the use of music videos [6]. However, even though virtual reality (VR) is fast becoming one of the key disruptive technologies in the entertainment and computer game industry, there have been very limited number of studies that have been conducted emotion classification using VR environments. The potential of emotion classification serving as an enabler to new genres of entertainment and computer games is vast considering that the real-time ability to passively and non-invasively detect the VR user's state of emotion would open up the possibility of unique subsequent experiences that are tailored specifically for the user's responses to the current content being delivered to the user via VR. Affectively-based personalized entertainment programs and emotion-based procedurally generated game content would now provide user-driven programming that could potentially increase the replay value of entertainment programs and computer games very significantly [7]. This premise is particularly enticing for the VR industry where the last missing piece of the VR simulated experience becoming as real as the non-simulated, real-life experience is the infusion of emotions [8]. VR systems today have graphics that are practically indistinguishable from real-life yet they are completely lacking the emotional aspect completely.

A recent study which used both EEG and electrocardiography (ECG) was conducted using VR for binary, subject-independent emotion classification. Classification accuracies of 75% and 71% were obtained, using Support Vector Machines (SVMs) as the classifier, were obtained for arousal and valence classification respectively [9]. Some early work proposing the use deep learning to detect only the specific emotional quadrant of high arousal, positive valence experienced from VR stimuli based on the presence of goosebumps and galvanic skin response has also been reported although no actual classification results were presented [10]. Inertial sensing is most commonly used to in human activity recognition applications [11]. Accelerometers especially those found in smartphones and smartwatches make up the large

majority of inertial sensing studies [12] while gyroscopes have also been used [13] but much less frequently compared to accelerometers. Recent studies have shown that a deep learning approach to human activity recognition using data obtained from wearable sensors such as accelerometers and gyroscopes is able to produce highly promising classification results [14-16].

Based on the literature review of related research and to the best of our knowledge, no attempt has yet been made to investigate the use of inertial sensing systems for emotion recognition in VR. As such, this forms the basis for this investigation which will leverage on the promising findings that have been reported from the use of deep learning approaches for human activity recognition using inertial sensors. Furthermore, this investigation also presents the first reported results from a four-class, subject-independent emotion classification approach using VR stimuli.

3 Material and Methods

3.1 Hardware Requirements

To deliver the immersive VR stimuli, an off-the-shelf VR Box as shown in Fig. 1(a) was used since it could accommodate the content delivery and app-based EEG acquisition via smartphones. For EEG acquisition, the Muse headset as shown in Fig. 1(b) from Interaxon was used. The Muse headset provides EEG as well as inertial sensing data from its built-in accelerometer and gyroscope. In order for the participants to have an immersive experience while wearing the VR Box, an earphone with noise cancelling capability was used to reduce any unwanted external interference which may disturb the participants' attention towards the VR experience.



Fig. 1. (a) VR Box (left), (b) Muse EEG headset (right).

3.2 Dataset Preparation

In order to collect the data from the Muse EEG headset, a 3rd party software known as Muse Monitor developed specifically for Interaxon's Muse EEG headset was used. The R language is used to run the machine learning classifiers and also for classification analysis. The Muse Monitor was configured to record the EEG and inertial sensing data at 0.5s intervals.

3.3 VR Stimuli

In order to facilitate the immersion experience and to stimulate the senses of the participants according to the Russell's Circumplex Model of Affect, Youtube 360 videos were identified and analyzed to ensure that each of the video would fit into 1 of the 4 quadrants of the AVS model. In order to have an easier grasp of the complex emotions that was observed from the AVS model, each of the quadrants were generalized with the following labelling: Quadrant 1 is Happy, Quadrant 2 is Angry, Quadrant 3 is Bored, and Quadrant 4 is Relaxed.

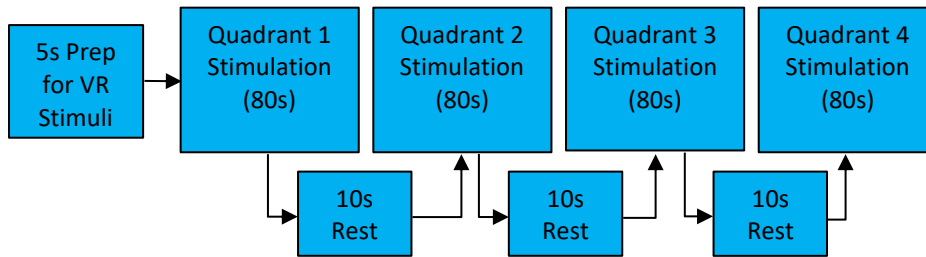


Fig. 2. Generalized AVS Model.

16 videos were collected from Youtube (4 from each quadrant) and were then sliced and stitched together where there are resting periods during initial video as well as transitioning to another quadrant of videos. The videos were presented in 20s block duration per video. Fig 2 below illustrates the setup of the stimuli video for the participants.

3.4 Experimental Setup



Fig. 3. Participant wearing the equipment for the experiment.

31 participants were chosen to participate in this study. The participants were all briefed with regards to the experimental protocol and were asked to indicate their consent to participate in the study. The environment of the experimental setup was prepared to ensure the safety of the participants as well as to avoid any interference from external sources such as noise and objects which may hinder the participants' free movement. The full experimental setup of the hardware including the Muse EEG headset as well as the VR Box as worn by the participants is shown in Fig. 3.

3.5 Feature Extraction

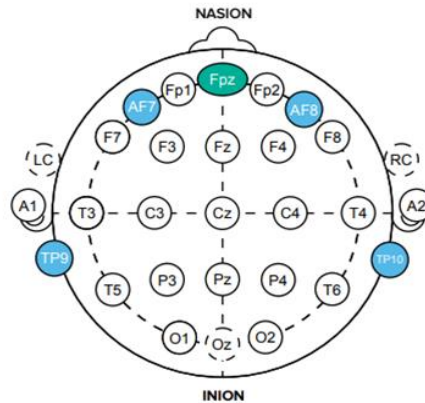


Fig. 4. Muse electrodes according to the standard 10-20 system.

The Muse EEG headset records the brainwaves detected via the scalp of the participant's head. It preprocesses the raw brainwaves into the 5 common power spectral densities of the human brain namely, Alpha (8Hz to 12Hz), Beta (12Hz to 38Hz), Delta (0.5Hz to 3Hz), Theta (3Hz to 8Hz) and Gamma (38Hz to 42Hz) using Fast Fourier Transform. The raw brainwave of the original signals along with inertial sensor information obtained from its accelerometer and gyroscope were also acquired. Muse electrodes records the brain activities on AF7, AF8, TP9 and TP10 and reference at Fpz based on the standard 10-20 system. Fig 4 below illustrates the location of the electrodes attached onto the scalp of the human head. Postprocessing using linear interpolation was utilized to handle missing values in the acquired dataset.

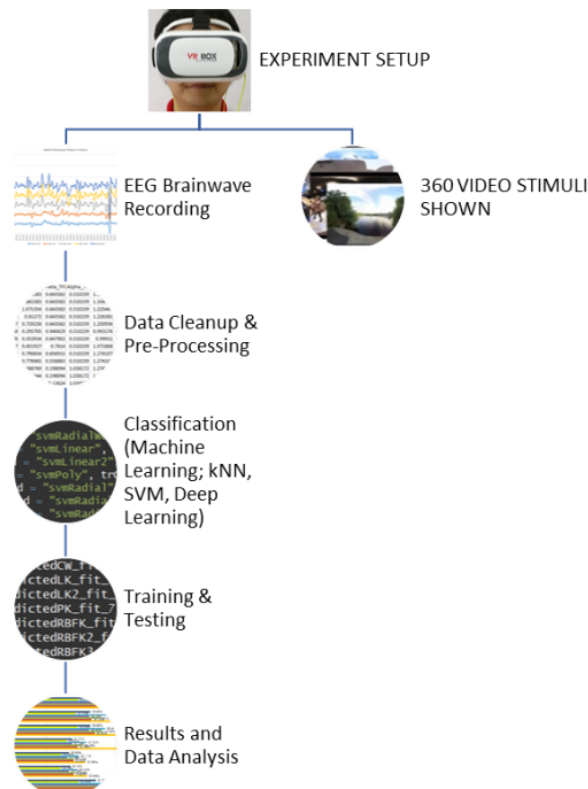


Fig. 5. Muse electrodes according to the standard 10-20 system.

Upon acquisition of the respective signals from Muse, postprocessing was conducted to address artefacts were introduced into the dataset which impacts the overall performance during the classification process. The biggest concern in the data collection was the gap within the

signal recordings of the data where it shows NA (Not Available) or missing datapoints. The process of approximation or “Interpolation” was calculated using R where the method of “Linear” was selected to do the interpolation. The extracted features were then fed into a deep neural network to conduct the emotion classification. Fig 5 shows the overall process of the experiment.

4 Results and Discussion

A preliminary test conducted on the parameter setup for the deep neural networks used for classification yielded architectures that performed best when utilizing two hidden layers with 200 hidden nodes within each layer. 10-fold cross-validation was used for each of the deep neural networks tested and were run using for 10 epochs for each experiment. The initial weights of the deep neural networks were set using the uniform adaptive method. Cross-entropy was set as the error function for training and testing. Rectified linear unit (ReLU) transfer functions set with a 50% dropout ratio and applying an adaptive learning rate were implemented for all the hidden layer nodes. The output layer used a softmax transfer function.

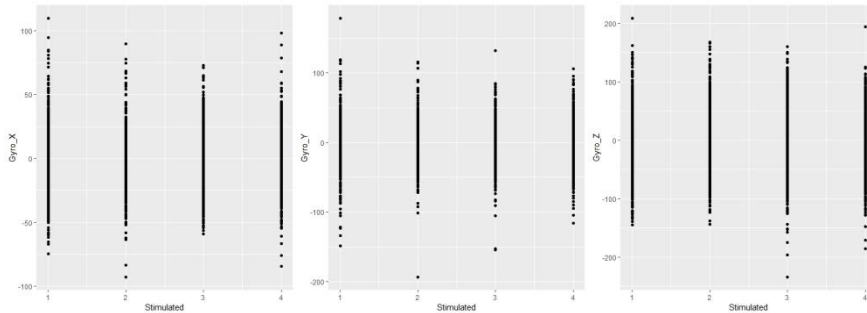
Short-Time Fourier Transform (STFT) was used to decompose the raw signal from each electrode into 5 bands (delta, theta, alpha, beta, gamma) giving a total of 20 features obtained from EEG data. Inertial sensing data was obtained in three axes (x,y,z) from both the accelerometer and gyroscope of the Muse headset giving a total of 6 features obtained from inertial sensors. The raw EEG signal from each of the four electrodes was also made available to the classifier as an additional input features. Therefore, each deep neural network was able to receive up to a grand total of 30 input features from both the EEG and inertial sensing data depending on the which experimental setup was being tested shown in table 1.

Table 1. Ten-fold cross-validation classification.

Setup No.	4-Quadrant Subject-Independent Classification	
	Setup Description	Accuracy (%)
1	5-band STFT EEG only	26.07
2	Raw EEG only	26.46
3	Inertial Sensing only	40.21
4	5-band STFT EEG + Raw EEG	40.40
5	Raw EEG + Inertial Sensing	40.92
6	5-band STFT EEG + Inertial Sensing	41.16
7	5-band STFT EEG + Raw EEG + Inertial Sensing	47.44

Table 1 shows the classification accuracies obtained from seven different combinations of features used as input to the deep neural network. The classifier produced the highest classification accuracy of 47.44% when all 30 features were used comprising the inertial sensing, raw and STFT EEG data.

The second and third best performing systems were also setups that utilized inertial sensing data to augment either the STFT or raw EEG data producing accuracies of 41.16% and 40.92% respectively. The two worst systems with accuracies of only slightly better than random for a four-class classification problem at 26.07% and 26.46% were the setups that exclusively used either STFT EEG or raw EEG data only and which excluded the use of inertial sensing data.



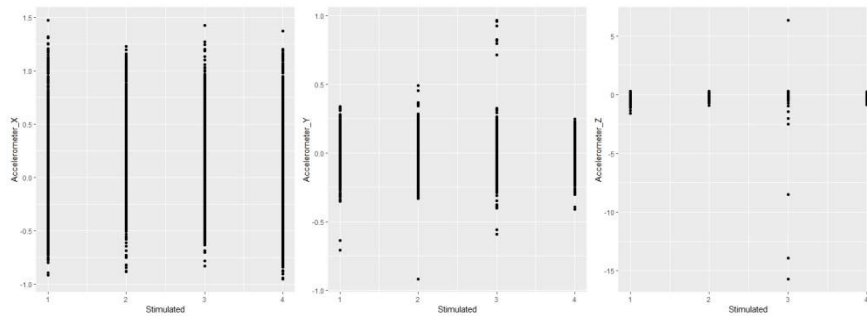


Fig. 6. Distribution of inertial sensor data against stimuli quadrant

A plot of the distribution of acquired signals obtained from each of the 6 inertial sensors is presented below in Fig. 6. As can be seen from the plots, there are observable although fairly small differences in the distribution of the values when comparing against the different axes and different sensors. It is perhaps these small but observable differences that are able to be used by the deep neural networks to make an improved distinction between the different quadrants that represent the 4 distinct emotional responses to the VR stimuli.

5 Conclusion and Future Work

This paper investigated the first use of inertial sensing for detecting emotions experienced through immersive virtual reality stimuli in a four-class, subject-independent approach. Augmenting raw and power spectral densities of EEG data obtained from using a wearable BCI device with accelerometer and gyroscopic data enabled a deep learning neural network to produce classification rates of 47% compared to only 26% obtained from non-augmented setups that used either only EEG raw or power spectral densities. As such, this study has demonstrated promising results from the use of inertial sensing as a method to augment existing EEG-based emotion classification systems, particularly when such inertial sensing data can be deployed for immersive VR environments.

Useful avenues of further investigation leading from this current explorative study include the use of additional modalities such as galvanic skin response as well as ECG data to further augment inertial sensing and EEG data. Other deep learning approaches such as Long-Short Term Memory architectures that are able to capture the dynamics of the emotional response may also be beneficial in further improving the classification accuracies.

Acknowledgements

This work was supported in part by the research grant ref: IF0318M1003 received from the Ministry of Energy, Science, Technology, Environment & Climate Change (MESTECC), Malaysia.

References

- [1] G. Devanur, : "What If Wearable Technologies Can Track Our Emotions?", Forbes, 2 Jan. (2019).
- [2] S. M. Alarcao, and M. J. Fonseca.: "Emotions recognition using EEG signals: A survey.", IEEE Transactions on Affective Computing, , DOI: 10.1109/TAFFC.2017.2714671. (2017)
- [3] M. Hamada, B. B. Zaidan, and A. A. Zaidan.: "A systematic review for human EEG brain signals based emotion classification, feature extraction, brain condition, group comparison.", Journal of medical systems, 42(9):162, (2018).
- [4] M. L. R. Menezes, A. Samara, L. Galway, A. Sant'Anna, A. Verikas, F. Alonso-Fernandez, H. Wang, and R. Bond, : "Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset.", Personal and Ubiquitous Computing 21(6): 1003-1013, (2017).
- [5] S. C. Mukhopadhyay, K. P. Jayasundera, and O. A. Postolache,: Modern Sensing Technologies, Springer, (2019).
- [6] Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil: "Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review." Applied Sciences 7(12):1239, (2017).
- [7] T. Christy and L. I. Kuncheva.: "Technological advancements in affective gaming: A historical survey.", GSTF Journal on Computing (JoC) 3(4), (2018).
- [8] G. Bernal, and P. Maes.: "Emotional Beasts: Visually Expressing Emotions through Avatars in VR."

Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, ACM,(2017).

- [9] I. Mavridou, E. Seiss, T. Kostoulas, C. Nduka, and E. Balaguer-Ballester.: "Towards an effective arousal detection system for virtual reality.", Proceedings of the Workshop on Human-Habitat for Health (H3): Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era, ACM, (2018).
- [10] D. Quesnel, S. DiPaola, and B. E. Riecke.: "Deep Learning for Classification of Peak Emotions within Virtual Reality Systems." International SERIES on Information Systems and Management in Creative eMedia (CreMedia) 2017(2):6-11, (2018).
- [11] A. Jain, and V. Kanhangad.: "Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors." IEEE Sensors Journal 18(3):1169-1177, (2018).
- [12] A. Jordao, L. A. B. Torres, and W. R. Schwartz.: "Novel approaches to human activity recognition based on accelerometer data.", Signal, Image and Video Processing, 1-8,(2018).
- [13] M.W. McCarthy, D.A. James, J. B. Lee, and D. D. Rowlands.: "Decision-tree-based human activity classification algorithm using single-channel foot-mounted gyroscope.", Electronics Letters, 51(9):675-676, (2015).
- [14] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu.: "Deep learning for sensor-based activity recognition: A survey.", Pattern Recognition Letters (2018).
- [15] T. Zebin, M. Sperrin, N. Peek, and A. J. Casson.: "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks.", 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018.
- [16] M. M. Hassan, M. A. Uddin, A. Mohamed, and A. Almogren.: "A robust human activity recognition system using smartphone sensors and deep learning.", Future Generation Computer Systems 81:307-313, (2018).