Computational Analysis of Valence and Arousal in Virtual Reality Gaming using Lower Arm Electromyograms

Ilia Shumailov
Department of Computer Science & Technology
University of Cambridge
is410@cam.ac.uk

Hatice Gunes
Department of Computer Science & Technology
University of Cambridge
Hatice.Gunes@cl.cam.ac.uk

Abstract—Progress in the affective computing field has led to the creation of affect-aware games that aim to adapt to the emotions experienced by the players. In this paper we focus on affect recognition in virtual reality (VR) gaming, a problem that to the best of our knowledge has not yet been sufficiently explored. We aim to answer two research questions: (i) Is it possible to reliably capture and recognize the affective state of a person based on EMG sensors placed on their lower arms, while they interact with the virtual environment? and (ii) Is EMG signal from one arm sufficient for detecting affect? We conducted a study in which 8 people were playing a set of VR games with two EMG sensors placed on their arms. We analysed the EMG signals and extracted a number of features to infer the affective states of the players. Our experimental results show that the EMG measures from left and right arms provide sufficient information to detect emotions experienced by a player of a VR game. Our results also show that classifying a DWT-*db1* signal with Support Vector Machine (SVM) yields F1=0.91 for predicting low/high arousal and F1=0.85 for predicting positive/negative valence when using just the left-arm EMG signal. To the best of our knowledge, this is the first work that uses EMG data from arm movements as a single source of affective information and addresses affect recognition in VR gaming.

1. Introduction

Electromyogram (EMG) measures the electrical activity that takes place in the muscles. Over the past hundred years, EMG signal has been researched and applied to many different problems. EMG-based control relies on the existence of patterns in the way muscles activate for a particular action and the detection of those patterns to infer what action was performed.

Recently, EMG-based device control has been gaining popularity. For example, Hernandez et al. showed that it is possible to control a prosthetic system using an EMG signal that is connected directly to the skin [6]. Another popular application of EMG is gesture recognition. Zhang et al. showed that it is possible to make use of EMG and accelerometer data to classify which gesture a person makes and use that gesture to control a video game [21]. Using

EMG data, it is also possible to detect a wider set of actions and distinguish between different types of grips [9].

In all of the above-mentioned research, the overall approach is similar - i.e., an action is performed, a sensor collects the signals representing this action, the signals get filtered discarding some information, and the patterns get detected to be attributed to the actions performed. However, we argue that the discarded data can be effectively used for analysing the affect expressed and creating adaptation in gaming (see [8] for a review on affective gaming).

Recent advancements in technology led to the development of fully functional VR games. Their potential is enormous - two of the Virtual Reality headsets, Oculus Rift and HTC Vive, have already found their way to homes of thousands of people and VR gaming is already being explored for education, training and therapy. In this paper, we hypothesise that EMG data obtained from the lower arms controlling the VR environment can be effectively used to analyse and recognise the player's affect in terms of valence and arousal. We aim to answer two research questions: (i) Is it possible to reliably capture and recognise the affective state of a person based on EMG sensors placed on their lower arms, while they interact with the virtual environment? and (ii) Is EMG signal from one arm sufficient for detecting affect, or do we need to use multiple sensors?

We conducted a study in which 8 people were playing a set of VR games with two EMG sensors placed on their arms. The participants were then asked to provide self-reported arousal and valence annotations using the scheme originally developed by Thyer [18], and further adapted by Grekow and Ras [5]. We analysed the recorded EMG signals and extracted a number of features to infer the affective states of the players. Our results show that the EMG measures from left and right arms provide sufficient information to detect emotions experienced by a player of a Virtual Reality game. Our results also show that classifying a DWT-*db1* signal with SVM-RBF yields F1=0.91 for predicting low/high arousal and F1=0.85 for predicting positive/negative valence when using just the left-arm EMG signal.

The rest of the paper is structured as follows. Section 2 discusses the related work in the field. It is followed by Section 3, which outlines the motivation and describes the

methodology adopted. Section 4 describes the feature extraction process while Section 5 presents the classification experiments conducted and the results obtained. Section 6 analyses and discusses the findings and Section 7 concludes the paper.

2. Related Work

One of the early attempts of using EMG to classify emotions was performed by Fridlund et al. [4]. They used EMG to capture the emotional response to images shown to a person. Benedek and Hazlett used facial EMG to perform a real-time evaluation of software design [1]. In their study, the users attempted to use the system, while their reactions were collected and classified as either positive or negative. Wagner et al. compared different methods for emotion classification using a combination of signals [20]. The work was further extended by Cheng and Liu to include EMG [2]. It was shown that back-propagation based neural network improved by Levenberg-Marquardt performs much better than Support Vector Machine (SVM) as was presented originally by Kim et al. [10]. Overall, EMG is most commonly processed using Discrete Wavelet Transforms (DWT) as, was done in the works of Cheng and Liu [2], and Kakoty and Hazarika [9]. Further analysis of different wavelets for DWT and their performance for EMG analysis was presented by Phinyomark et al. [13]. Similarly to the choice of wavelets, different features were also analysed in order to understand how much information these features can capture about the signal and when one should use one feature over another [15].

The rise in popularity of VR is associated with a very engaging and realistic experience provided by the virtual reality equipment. The realism was captured in the work of Slater et al. [23] who observed the physiological response to the presence of virtual reality characters. The power of that realism was also recognised by the medical research community. For example, Opdyke et al. showed that one can actually treat the fear of heights using VR [24]. Similarly, Welch et al. have shown that it is possible to build models of social interaction to study autism [22]. They presented a system where a number of sensors were used to capture the affective state of a child during the interaction with a virtual character. Similarly to the research by Slater et al., their system utilised multiple physiological sensors [23].

However, all of the above-mentioned systems use a number of complex and expensive sensors. In this work, we hypothesise that affective response in virtual reality gaming can be captured using a single sensor which is relatively cheap and is available in the consumer market. The creation of affect recognition systems built upon such cheap and widely available sensors could positively impact various areas such as social welfare, and improve the effect of platforms such as 'VR for Good'[1].

1. Accessible at: https://www.oculus.com/vr-for-good/

Figure 1. One of the participants interacting with the VR environment during the pilot study.

3. The Study

3.1. Motivation and Pilot Study

Our motivation for focusing on the EMG data was two-fold. Firstly, in a VR setting the face of the player is mostly occluded due to the headset they wear. Therefore, although the face is considered to be the most obvious and reliable source of affective information, it is not yet possible to utilise it for behaviour analysis in VR settings. Secondly, in current VR settings, virtually all of the interaction with games happens through the use of controllers and hand movements. As the amount of possible interactions is limited to just a number of buttons and hand movements, we believe that it is possible to reliably capture the patterns in muscle activations corresponding to states of change.

With these motivations, we designed a study to answer the two research questions we set out to explore (as described in Section) 1. Before we proceeded with the actual study, we ran a pilot study (see Figure 1) to explore a number of issues inherent to what we wanted to investigate, e.g., exploring how the EMG sensors could best be used, choosing a number of games that would capture a wide range of emotions in the time that the game was played, and finding the most convenient way for the participants to report on their emotions. Two people participated in our pilot study.

The data collected during the pilot study illustrated the need for using two Myo armbands per player to collect data from both arms. There were two reasons for this. First, one of the pilot study participants was left-handed, whereas the other one was right-handed, resulting in potential discrepancies in collected signals. Second, some games mostly made use of just one main hand (e.g. Google Earth), whereas others required the use of both hands simultaneously (e.g. Fruit Ninja VR). Using only one armband would not capture the richness of such information. However, it should be noted that we could not find a way of correctly using two Myo armbands connected to the same computer with the 0.9

Thalmic Myo Software Development Kit (SDK)[2]. Instead, two computers were used for data collection, with software set-up in such a way to start the data collection at the same time.

We tried multiple configurations for obtaining data annotation. Initially, we asked one participant to play a game for a duration of 30, 60 and 90 seconds and annotate the affect they experienced during the game. However, the participant reported that he/she could not recall the emotions they experienced during the game, and when they could recall the emotion they would struggle to tell when exactly they felt a particular emotion. Due to these reasons, we decided to record the gameplay and the speech of the player to present it to them during the post-gaming annotation phase.

3.2. The Chosen VR Games

The game choice was driven by the desire to find games that could induce a wide variety of emotions in a short time period. For this reason we chose to use the following games: 1) *Eggs time*[3] - The game is very engaging, but at the same time has multiple graphics bugs, which cause the users to experience a range of negative and positive emotions; 2) *Google Earth*[4] - The game has quite confusing controls that can cause a range of negative emotions; 3) *Spell Fighter*[5] - The game was specifically chosen, as it is practically impossible to play it without speech input and in the game it is possible to get attacked by a mannequin, without any ability to defend. The game was therefore chosen to induce a variety of negative emotions; 4) *Museum of Fine Arts*[6] - The game presents participants with a range of artwork and was chosen to induce a range of calm-positive and calm-negative emotions; 5) *Fruit Ninja*[7] - The game is very easy to play in terms of controls and is very engaging.

In order to find the appropriate time for a game round, we investigated how long it takes for a player to get into the game and start generating emotionally rich EMG data. We ended up with a time period of 60 seconds, as this was the duration that participants believed they had experienced a rich range of emotions.

3.3. Sensors and Measures

For the VR gaming, as can be seen in the Figure 1, HTC Vive headset[8] was used. The configuration was default and games were accessed through the Steam VR marketplace[9].

In the updated Myo SDK a new interface was introduced that gives developers an ability to stream raw EMG data from the Myo armband [11]. Each armband provides the following sensor data: 8 EMG channels, gyroscope and orientation. Data is streamed at a frequency of 200Hz for EMG channels and of 50Hz for gyroscope and orientation. Unfortunately, those sampling rates are default and can not be changed. We have decided to only use EMG for the emotion classification task. As we set the time for a game round to be 60 seconds, this resulted in $60 * 200 = 12000$ samples per player, each with 8 EMG channels per arm.

3.4. Data Collection and Annotation

8 participants (4 female and 4 male) with different cultural and educational backgrounds volunteered to participate. The participants also came from different educational backgrounds and different English proficiency levels (from intermediate to native speakers). The participants were informed about the experiment, its goals, safety procedures and their rights with a verbal introduction and through a signed consent form. Participants were given time to get themselves familiar with both the Myo armbands and the VR headset. Participants were also introduced to the meaning of *arousal* and *valence*. Prior to the collection phase the participants were given time to freely play the games to get used to the controls. Once the experiment was over, the logic behind the game choice was explained to reassure that there is nothing wrong with them feeling confused/annoyed at particular points in time.

In order to remove bias between game rounds, a break of at least one minute was taken. Running a single game round took around one minute with about 20 seconds preparation, followed by about 10 minutes discussion about the emotions felt, averaging out to about 1 hour 10 minutes time overall per participant. In total, we obtained $8 * 5 * 60 = 2400$ EMG-seconds from 8 participants. As we were collecting data from both arms, this provided us with a total of 4800 EMG-seconds with around 200 EMG samples per second. The participants were asked to provide one label for arousal and one label for valance, for every second of the gameplay.

For the purpose of emotion reporting it was necessary to pick an emotion model that would be simple enough for non-professional labelers to describe their feelings but at the same time flexible enough to allow participants to express a whole range of emotions. When designing the experiment, we expected that there is going to be a strong positive bias, as most people had no interaction with Virtual Reality-based gaming before the study. For these reasons, an Arousal-Valence scheme over the continuous space in the range of $[-1, 1]$ was chosen. This scheme with continuous values allowed participants to express the extent to which a particular emotion was experienced. Discrete schemes, similar to the one used by Cheng and Liu were considered but were dismissed due to the limitations they impose [2].

In order to help participants describe their affective state, a printed out copy of the scheme was provided to them. This scheme is based on the original one developed by Thyer [18], but further adapted by Grekow and Ras [5] for emotion recognition in music. We have further adapted it by adding more labels around the Arousal-Valence space to make it

2. Accessible at https://developer.thalmic.com/
3. Accessible at http://store.steampowered.com/app/531990/
4. Accessible at http://store.steampowered.com/app/348250/
5. Accessible at http://store.steampowered.com/app/455440/
6. Accessible at http://store.steampowered.com/app/515020/
7. Accessible at http://store.steampowered.com/app/486780/
8. Accessible at https://www.vive.com/uk/
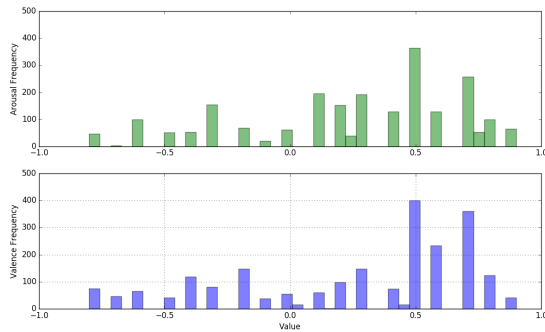9. Accessible at http://store.steampowered.com/steamvr

Figure 2. Histogram of collected labels for Arousal and Valence. Blue-valence, Green-arousal. X-axes present the value, Y-axes - the frequency of appearance.
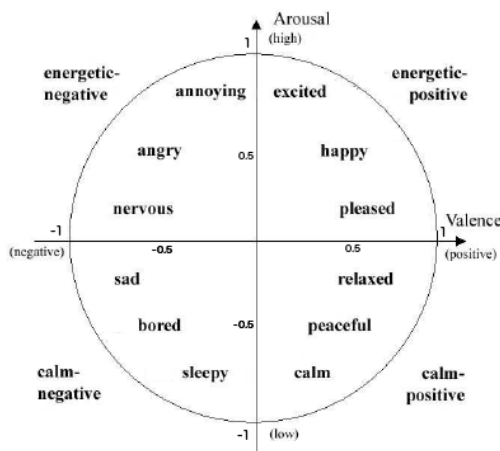

Figure 3. The annotation scheme used.

easier for participants to describe their feelings. The adapted scheme is illustrated in Figure 3.

The collected label distribution is presented in Figure 2. As can be seen from the figure, the data collected is unbalanced and there is a strong bias towards positive emotions. Classification of such data would result in low accuracy and high misrepresentation. In order to mitigate this issue, we will analyse the classification results after the commonly used under- and over- *sampling techniques* are applied to the data. These techniques are discussed in Section 5.

4. Feature extraction

While exploring the use of EMG in biomechanics De Luca [3] outlined that the relationship between actual and reported muscle contraction force is nonrigid due to multiple factors[10]. Such claims were further explored in the review by Sousa and Tavares [17]. Not only the signal is reported to be different for each individual, but it changes within the same

---

10. For example, electrode configuration, fibre type or signal crosstalk. For the full list see Figure 1 in De Luca [3].

individual over time, and therefore absolute values of the signals are not representative in group comparisons. In order to allow for comparison between different individuals and different muscles, there is a need to standardise the signal. To do this, isometric and dynamic normalisation methods are reported to be the most appropriate ones [17]. Kakoty and Hazarika used mean normalisation to process EMG to classify the gestures [9]. Shrirao et al. used the Root Mean Square normalisation for classification of the EMG data to predict the joint angle [16]. For the purpose of this work, we decided to make use of Root Mean Square (RMS) normalisation.

Due to the historic popularity of EMG, many different features were considered for classification, both individually and in groups. These different features were considered in time, frequency and time-frequency domains. Time domain features are based on EMG signal amplitude. Tkach et al. presented a review of different features that are extracted from EMG signal and are used for pattern recognition [19]. They used features utilised by Hudgins et al. [7]: Mean Absolute Value, Mean Absolute Value Slope, Zero Crossings, Slope sign changes, Waveform Length, combined with autoregressive, cepstral coefficients, Willison amplitude and sample entropy. In frequency domain, power spectrum, mean and median of signal frequencies and frequency ratio are used. Lastly, in time-frequency domain, only short-time Fourier transform, wavelet transform and wavelet packet transforms are used.

Kakoty and Hazarika claimed that out of all domains, the time-frequency domain is the most suited for EMG signal analysis [9]. The reason is that the wavelet transforms capture frequency resolution really well for low frequencies and time resolution for high frequencies. In EMG signals high-frequency components are rare and last for a very limited period of time, whereas low frequency persists for much longer. Additionally, they argued that approximations calculated by the discrete wavelet transform, although reduce the size of the signal, still contain sufficient information for analysis and feature extraction.

For the reasons outlined above, we decided to make use of the following features on the DWT approximation of the original signal: 1) Mean absolute value; 2) Mean absolute value slope; 3) Original signal zero crossing; 4) Original signal slope sign changes; 5) Waveform length; 6) Variance; 7) Root Mean Square; 8) Waveform Length; 9) Standard deviation of the amplitude of the signal. For mathematical definitions of the features outlined above please refer to the review published by Phinyomark et al. [12].

The choice of mother wavelets for initial signal approximation was based on the follow-up research by Phinyomark et al. [13]. According to their research, for decomposition and de-noising (Type I and Type III in classification) *coif5*, *db1*, *bior1.1* and *rbio1.1* should be used. Originally we wanted to use all of these wavelets, however, we discovered that *db1*, *bior1.1* and *rbio1.1* resulted in very similar classification accuracy whereas *coif5* showed relatively smaller accuracy. Therefore, in this paper, we present the results of DWT based on *db1*.

TABLE 1. SVM CLASSIFICATION RESULTS FOR LOW/HIGH AROUSAL AND NEGATIVE/POSITIVE VALENCE. THE RESULTS PRESENTED ARE AVERAGE F1 VALUES WITH SUBJECT-INDEPENDENT CROSS-VALIDATION OVER ALL 8 SUBJECTS (ALL) AND OVER 7 OUT OF 8 SUBJECTS (7/8).

Sampling	Left				Right			
	Arousal		Valence		Arousal		Valence	
	All	7/8	All	7/8	All	7/8	All	7/8
None	0.86 ± 0.16	0.91 ± 0.06	0.83 ± 0.07	0.84 ± 0.07	0.86 ± 0.16	0.91 ± 0.06	0.83 ± 0.08	0.84 ± 0.08
CNN	0.86 ± 0.16	0.91 ± 0.06	0.82 ± 0.06	0.83 ± 0.06	0.85 ± 0.16	0.90 ± 0.06	0.81 ± 0.07	0.82 ± 0.07
SMOTE	0.84 ± 0.15	0.89 ± 0.06	0.80 ± 0.06	0.81 ± 0.06	0.83 ± 0.14	0.88 ± 0.06	0.80 ± 0.06	0.81 ± 0.06
SMOTE+CNN	0.86 ± 0.16	$\mathbf{0.91 \pm 0.06}$	0.83 ± 0.08	$\mathbf{0.85 \pm 0.07}$	0.86 ± 0.16	$\mathbf{0.91 \pm 0.06}$	0.83 ± 0.08	$\mathbf{0.84 \pm 0.08}$

5. Emotion Classification

As the majority of the affective computing research using EMG focused on the prediction/classification of four classes, for comparison purposes we also decided to approach the problem as a four-class classification problem using RBF-based Support Vector Machine with $C = 1$ implemented in *sklearn*[11] python framework. Alongside, we used one under/oversampling technique to mitigate the bias presented in the data implemented in the *imbalanced-learn* python framework. The techniques are used to mitigate the biased nature of data discussed previously.

Table 1 presents the classification results for the collected data in terms of average F1 values calculated with leave-one-subject-out (LOSO) cross-validation. The table presents the classification results for original data along with the data that is over-sampled with SMOTE [25] and under-sampled with Condensed Nearest Neighbor [26]. SMOTE generates synthetic examples while operating in the feature space. In particular, it takes each minority class sample and introduces a synthetic sample in the region between a given number of nearest samples from the same minority class. Those synthetic neighbours are generated by computing the difference between the features of the samples and randomly picking a number and adding it to the feature vector. It adds a random point in the region in-between already existing feature vectors.

Condensed Nearest Neighbours (CNN) is an under-sampling technique presented by Hart to mitigate the storage constraints of the nearest neighbours classifiers [26]. The technique focuses on picking out a subset of samples from the original set of samples in such a way that the subset provides sufficient information to classify the points in the original set that are not present in the subset.

6. Analysis and Discussion

Table 1 shows the SVM classification results for low/high arousal and negative/positive valence. for the whole dataset, along with the results obtained for a reduced subset. The latter is presented to explain the high standard deviation that can be seen in the classification results of both left and right arms. The subset excludes one of the participants who during the annotation phase stated that he/she was too tired to reflect on the experienced affect during the game and provided the same annotations for

11. See http://scikit-learn.org/stable/

both arousal and valence. By excluding the data of this participant, one could see a much smaller standard deviation for the F1 measure.

As it can be seen from the table, when using the original dataset, both arousal and valence can be classified with high accuracy (F1=0.85), which further increases when the subset is used (F1=0.91). This suggests that even with unbalanced data we were able to fairly reliably recover the affective state of a person playing VR games. To explore how the lack of balance in the data has further affected the classification results, we calculated F1 values from over- and under-*sampled data* by reporting on the following: 1) CCN of the data; 2) SMOTE of the data; and 3) first SMOTE of the data followed by the CNN. As it can be seen from Table 1, for CNN the F1 values are very similar to those of the original dataset but decrease slightly for both left and right arms. This is due to the fact that CNN omits some of the data, that could later be of use for the SVM classifier. A more interesting trend has been observed for the data after SMOTE. There is a slight drop in the F1 value, suggesting that the problem is not the undersampled class, but rather the oversampled class that suffers from misclassification due to an increased confidence in prediction of the samples on the boundary between the classes. Finally, similarly to Chawla et al. we oversampled the original data with SMOTE and then undersampled it [25]. The authors found that such a procedure results in a more generalised classifier. In our SMOTE+CNN case, we can see that the obtained F1 value is the same, except for the valence reported for the left arm. There we can see a slightly higher standard deviation in the case of the whole dataset, and a slightly better F1 value for the subset.

As we have hypothesised, if the misclassification is due to the high variation within the majority class, there is indeed a need for more data. Finally, we could see very similar accuracies for both left and right arms, suggesting that one source of EMG could be sufficient for effective emotion recognition in VR gaming. To make sure that we were not classifying game-specific emotional responses, we ran an additional experiment with leave-one-game-out evaluation, which yielded very similar results to the ones presented in Table 1.

7. Conclusions and Future Work

Our results showed that the EMG measures from both (lower) arms provide sufficient information to detect emotions experienced by a player of a Virtual Reality game.

Although not in a VR setting, Cheng and Liu reported on emotion recognition from surface EMG in [2] using Back-propagation Neural Network. The recognition accuracy they reported was as follows: 67% for joy, 83.33% for anger, 83.33% for sadness and 67% for pleasure, with an overall recognition accuracy of 75%. Our results showed that it is possible to obtain comparable and even higher classification accuracy when using lower arm EMG. We obtained F1=0.91 for the classification of arousal and F1=0.85 for the classification of valence, in both left and right arms. Furthermore, we found that data from both left and right arms result in comparable classification accuracy, suggesting that a single sensor is sufficient to detect affect experienced during VR gaming. Although this result is surprising it could be explained by the fact that all of the games required the participants to keep the hands active during the gameplay. To the best of our knowledge, this is the first work that uses EMG data from arm movement as a single source of information and addresses affect recognition in Virtual Reality gaming. However, to make firm conclusions an additional study should explore the generalizability and effectiveness of the technique described here, with a higher number of participants.

The development of a virtual reality affective game, which would change its difficulty in response to the affect experienced by the players, is one of the avenues we would like to explore as an extension of the work presented in this paper.

References

[1] J. Benedek and R. Hazlett. Incorporating facial emg emotion measures as feedback in the software design process. *Proc. of Human Computer Interaction Consortium*, 2005.

[2] B. Cheng and G.-Y. Liu. Emotion recognition from surface emg signal using wavelet transform and neural network. *Proc. of International Conference on Bioinformatics and Biomedical Engineering*, pp. 1363–1366, 2008.

[3] C. J. De Luca. The use of surface electromyography in biomechanics. In *Journal of Applied Biomechanics*, 13, pp. 135–163, 1997.

[4] A. J. Fridlund, G. E. Schwartz, and S. C. Fowler. Pattern recognition of self-reported emotional state from multiple-site facial emg activity during affective imagery. In *Psychophysiology*, 21(6), pp. 622–637, 1984.

[5] J. Grekow and Z. W. Raś. Detecting emotions in classical music from midi files. *Proc of the International Symposium on Foundations of Intelligent Systems*, pp. 261–270, 2009.

[6] A. Hernandez Arieta, R. Katoh, H. Yokoi, and Y. Wenwei. Development of a multi-dof electromyography prosthetic system using the adaptive joint mechanism. In *Applied Bionics and Biomechanics*, 3(2), pp. 101–111, 2006.

[7] B. Hudgins, P. Parker, and R. N. Scott. A new strategy for multifunction myoelectric control. *Proc. of IEEE Transactions on Biomedical Engineering*, 40(1), pp. 82–94, 1993.

[8] E. Hudlicka. Affective gaming in education, training and therapy. In *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches*, pp. 482–511.

[9] N. M. Kakoty and S. M. Hazarika. Recognition of grasp types through principal components of dwt based emg features. *Proc. of IEEE International Conference on Rehabilitation Robotics*, pp. 1–6, 2011.

[10] K. H. Kim, S. Bang, and S. Kim. Emotion recognition system using short-term monitoring of physiological signals. In *Medical and biological engineering and computing*, 42(3), pp. 419–427, 2004.

[11] S. Lake, M. Bailey, and A. Grant. Announcing raw emg data for developers from the myo armband, 2014. Accessible at: developerblog.myo.com/big-data/

[12] A. Phinyomark, C. Limsakul, and P. Phukpattaranont. A novel feature extraction for robust emg pattern recognition. In *Journal of Computing*, 1(1), pp. 71–80, December 2009

[13] A. Phinyomark, C. Limsakul, and P. Phukpattaranont. Optimal wavelet functions in wavelet denoising for multifunction myoelectric control. In *ECTI Transactions on Electrical Eng., Electronics, and Communications.*, 8(1), pp. 43–52, 2010.

[14] A. Phinyomark, C. Limsakul, and P. Phukpattaranont. Application of wavelet analysis in emg feature extraction for pattern classification. In *Measurement Science Review*, 11(2), pp. 45–52, 2011.

[15] E. Scheme and K. Englehart. On the robustness of emg features for pattern recognition based myoelectric control; a multi-dataset comparison. *Proc. of Engineering in Medicine and Biology Society*, pp. 650–653, 2014.

[16] N. A. Shrirao, N. P. Reddy, and D. R. Kosuri. Neural network committees for finger joint angle estimation from surface emg signals. In *Biomedical Engineering Online*, 8(2), pp. 2–13, 2009.

[17] A. S. Sousa and J. M. R. Tavares. Surface electromyographic amplitude normalization methods: a review. In *Electromyography: New Developments, Procedures and Applications*, 2012.

[18] R. E. Thayer. *The biopsychology of mood and arousal*. Oxford University Press, 1990.

[19] D. Tkach, H. Huang, and T. A. Kuiken. Study of stability of time-domain features for electromyographic pattern recognition. In *Journal of Neuroengineering and Rehabilitation*, 7(1), pp. 1–13, 2010.

[20] J. Wagner, J. Kim, and E. André. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 940–943, 2005.

[21] X. Zhang, X. Chen, W.-h. Wang, J.-h. Yang, V. Lantz, and K.-q. Wang. Hand gesture recognition and virtual game control based on 3d accelerometer and emg sensors. *Proc. of the International Conference on Intelligent User Interfaces*, pp. 401–406, 2009.

[22] K. Welch, U. Lahiri, C. Liu, R. Weller, N. Sarkar, and Z. Warren. An affect-sensitive social interaction paradigm utilizing virtual reality environments for autism intervention. In *Ambient, Ubiquitous and Intelligent Interaction*, pp. 703–712, 2009.

[23] M. Slater, C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, and D. Friedman. Analysis of physiological responses to a social situation in an immersive virtual environment. In *Presence: Teleoperators and Virtual Environments* , 15(5), pp. 553–569, 2006.

[24] Opdyke, Dan, James S. Williford, and Max North. Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. In *Am J Psychiatry*, 152(4), pp. 626-62-628, 1995.

[25] Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. In *Journal of Artificial Intelligence Research*, 16, pp. 321–357, 2002.

[26] P. Hart. The condensed nearest neighbor rule, In *Information Theory, IEEE Transactions on Information Theory*, 14(3), pp. 515–516, 1968.