NCSP'22
RISP International Workshop on Nonlinear circuits, Communications and Signal Processing 2022
February 28 - March 1

Sensor Fusion-based Emotion Classification In Virtual Reality Using Machine Learning

Yi Liu[1] Eri Sato-shimokawara[1] and Toru Yamaguchi[1]

[1]Tokyo Metropolitan University
6-6, Asahigaoka, Hino, Tokyo, 191-0065, Japan
Phone/FAX:+81-42-585-8649
E-mail: liu-yi@ed.tmu.ac.jp, eri@tmu.ac.jp, yamachan@tmu.ac.jp

Abstract

In recent years, the development of virtual reality technology has made virtual reality content available to the average household. This paper primarily focused on finding the relationship between individual emotion and biological information while immersed under virtual reality toward estimating group-level dialogue atmosphere. Our findings validate the use of biometric information to automatically recognize different emotional states and dialogue atmospheres.

1. Introduction

Affective computing has emerged as an important interdisciplinary field of study based on psychology, computer science, and biomedical engineering, which Rosalind Picard proposed in 1997. In recent years, virtual reality(VR) technology development has made VR content available to the average household. VRChat, Neos VR, and Horizon Workrooms are popular applications in VR, which offer an endless collection of social VR experiences. To help conversation with other users or intelligent virtual humans on those online virtual world platforms, we need to estimate users' inner moods and evaluate dialogue atmosphere based on affective computing methodology. However, it is not easy to estimate affective state while wearing a VR headset, not only for computers but even for humans themselves. Psychology researchers consider that emotions are sometimes not expressed, and information expressed, such as facial expressions and voices, can differ from the actual emotion [1]. For example, people can fake a smile when feeling sad or angry inside. Based on this understanding, our research aims to understand true feelings toward estimating group-level atmosphere in multi-person Dialogues while immersed in the virtual reality world.

In this paper, to classify the emotion, we use biological features extracted from EEG and ECG data gathered from wearable sensors and combined through Support Vector Machine(SVM), Naive Bayes, and K-Nearest Neighbor(KNN) algorithms. The experiment was conducted in VRChat with a single collaborator. To evaluate the dialogue atmosphere, we used biological features and audio features. The experiment was conducted in Horizon Workrooms with two collaborators talking on a specific topic in the same virtual meeting room.

The structure of this paper is as follows: In section 2, the background and related work for this research paper will be briefed; In section 3, we propose the first experiment and evaluation; In section 4, we propose the second experiment and evaluation; In section 5, we will present the conclusion and future work.

2. Background and Related Work

As described in the introduction, we embody the technique that estimates emotion and dialogue atmosphere using biometric information. In the following, we will brief through the related research of this paper.

2.1 Automatic emotion recognition

Automatic emotion recognition (AEE) is an essential issue in various fields of activities, typically performed by measuring various human body parameters or electric impulses in the nervous system and analyzing their changes [2].For emotion recognition, the most popular biological information used are electroencephalography (EEG), electrocardiography (ECG), galvanic skin response (GSR), respiration rate analysis (RR), skin temperature measurements (SKT), Electromyography (EMG), Electrooculography (EOG). In addition, blood pressure, heart rate variability (HRV) based on ECG or PPG is also common [2]. External features, such as eye gazes, facial expressions, body motions, and voices, are also usually used for multimodal emotion recognition.

The most common application scenario for AEE is human-computer interaction(HCI). In Yuhei Ikeda's research, they used brain waves and heart rate to Estimate Emotion for Robot Interaction [1]. AEE research in VR is also popular since 2016. In Marín-Morales's research, they researched emotion recognition from brain and heartbeat dynamics using wearable sensors in virtual reality [3]. In NS Suhaimi's research, they tried emotion state classification of EEG brain signal collected in Virtual Reality using SVM and KNN [4].

2.2 Mood Engineering

When people communicate with each other, they sense the "atmosphere" by the words, facial expressions, gazes, nods, etc., of multiple people and consciously or unconsciously anticipate the next possible atmosphere. In 2013, the Japanese Society for Artificial Intelligence started activities in mood engineering, to discuss the atmosphere created by communication from an engineering point of view. Mood engineering is a theory that studies the atmosphere generated by multi-person conversations. The specific research targets of mood engineering include conversations among multi-people, conversations with artificial verbal and non-verbal information, and the atmosphere on the Internet inferred from the verbal activities of multiple people in web media [5].

In 2019, DNP developed a "next-generation stealth space". It's a system that can change the color of light and background music in a room by using sensors to detect the status of people's presence and conversations and the indoor environment. In Shohei Kato's research, they researched Dialogue Mood Estimation Focusing on Intervals of Utterance State [6]. They also used Speech Sounds considering the personality trait of the speaker for dialogue mood estimation [7].

3. Preliminary Experiment

In the first stage, single-person emotional arousal experiments were conducted to determine whether biological information can be used for affective computing toward estimating group-level dialogue atmosphere in the VR environment.

3.1 Methodology of the Preliminary Experiment

We conducted emotional arousal experiments in VRChat, a famous online VR platform. Ten participants aged between 20 through 28 were asked to wear polar H10, muse2 headband, and quest2 VR headset. Polar H10 is a highly reliable workhorse in the chest strap category. Muse2 headband is a multi-sensor meditation device that provides real-time feedback on brain activity, heartbeat, breathing, and movement. Quest2, designed by Facebook, does not need a PC or console, is lightweight and powerful enough to run impressively detailed virtual reality experiences.

Table 1: Six worlds in VRChat

Relax	Sky	Light rain room	Sad	Heavy rain room	/
Fear	SCP/WDK	/	Disgust	Dead Body room	/
Joy	Water park	/			

We selected six popular and quest2 available worlds in VRChat(Table 1), which correspond to five of the six basic emotions(relax, sad, fear, disgust, and joy).

The collaborators were asked to experience the six VR worlds listed in Table 1 and intentionally arouse their own emotions by immersing them in VR worlds and music. Meanwhile, ECG and EEG will be recorded. Every section will last about 3 minutes.

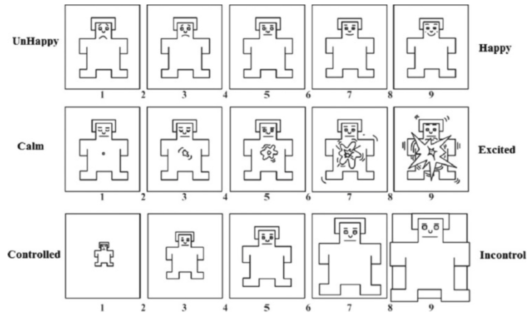

Figure 1: Self-Assessment Manikin questionnaire [8]

	Weak 1～7 Strong
Relax	
Sad	
Excited	
Fear	
Disgust	
Joy	

Figure 2: Six emotions questionnaire

After each section, collaborators needed to answer the Self-Assessment Manikin questionnaire(Fig.1) and six emotions(Relax, Sad, Fear, Excited, Disgust, Joy) questionnaire(Fig.2) and rest for a few minutes.

3.2 Result of the Preliminary Experiment

3.2.1 Data Analysis and Processing

The raw data collected by polar H10 and muse2 headband were recorded to CSV format files for use in Python programming. Every section lasted about 3 minutes, and we selected last-minute data. We converted raw ECG data to HRV data and calculated Attention, Meditation values using raw EEG data. When converting ECG to HRV, ECG data up to approximately 120-300 seconds earlier is necessary. HRV for the last minute can be calculated by sliding through the time window. Then we reduce the data to approximately 1 data point per second, including the features of HRV, Brainwave spectrum, and head Accelerator. We analyzed the quality and distribution of the data and processed the data as appropriate. According to the results of data analysis, some collaborators' biological information shows individual differences. Moreover, due to the poor contact problem with Muse2, some channels of brainwaves were not collected successfully. So we discarded the data that was not typical enough or had a large number of missing values.

According to the self-report results in Table 2, we believe that the emotional arousal experiment in VRChat can

Table 2: Self-report information

	Valence	Arousal	Dominance	Emotion Value
Relax1	5.25	1.75	6.125	5.56
Relax2	4	1.75	5.5	6
Sad	2	3.75	2.875	5.25
Fear	1.875	6.25	1.5	6
Disgust	2.5	5.375	3.25	6
Joy	7.14	3.71	5.86	5.86

effectively arouse collaborators' emotions. We also found there are situations where collaborators are happy&relaxed (Fear&Disgust, Excited&Joy) simultaneously. It suggests that those emotions are not discrete in some cases.

3.2.2 Classification Result

After normalization, PCA, one-hot encoding, we trained SVM, Naive Bayes and KNN for emotion classification. The Cross-Validation results in Table 3 suggest that the features of ECG and EEG are valid for emotion classification in VR. If we want to improve accuracy, we need to increase the amount of typical data and find ways to reduce the impact of individual differences. However, because of the poor contact problem of Muse2, it is not easy to perform experiments wearing Muse2 brainwave headband, not to mention conducting multi-person experiments. So we decided not to use Muse2 in the second experiment.

Table 3: Cross-Validation Classification results using SVM, Naive Bayes and KNN

	SVM	Naive Baiyes	KNN
Relax	0.6875±0.15	0.625±0.15	0.5625±0.15
Sad	0.6875±0.15	0.625±0.15	0.625±0.15
Fear	0.6875±0.07	0.625±0.07	0.6875±0.15
Disgust	0.5625±0.15	0.625±0.15	0.5625±0.07
Joy	0.625±0.15	0.5625±0.07	0.625±0.07

4. Second Experiment

In the second stage, we want to use biological information and audio to evaluate the dialogue atmosphere in VR. Since we abandoned the use of Muse2 because of the contact problem, we are considering using HRV and conversational features to assess the atmosphere of two-person conversations.

4.1 Methodology of the Second Experiment

The two collaborators were sitting in separate quiet classrooms. If they did not wear VR headsets, they could not hear each other. After wearing a Quest2 headset, they were asked to access the same virtual meeting room in Horizon Workrooms customizing as a virtual avatar and have conversations on a specific topic.

Six pairs of collaborators were recruited(4 females and 8 males, aged from 20 to 28). Each pair was asked to talk on 4 or 5 different topics. We chose the topics on the principle that they are likely to cause emotional fluctuations or are likely to produce opposing positions. The topic of the conversation was chosen randomly, and each conversation lasted approximately three minutes. Between each different topic, they have a few minutes to answer the self-report questionnaire and take a break.

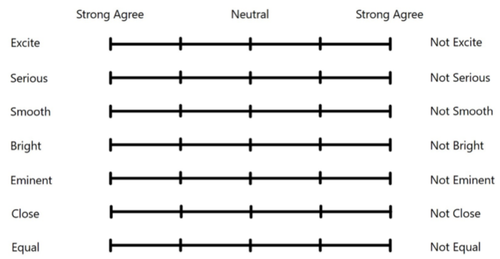

Figure 3: Dialogue Moods Questionnaire [7]

For the self-report section, collaborators answered the six emotions self-report questionnaire(Fig.2) and dialogue mood evaluation questionnaire(Fig.3). We mainly focus on the questionnaire about dialogue mood, while the six emotion questionnaires are used as references to better understand the collaborators' status during the experiment.

If there is an apparent inconsistency between their self-report and their actual performance, we will confirm with the collaborators to ensure the correctness of the questionnaire answer. We also found that despite two people being in the same conversational setting, there may be opposite results in evaluating the dialogue atmosphere by both parties.

4.2 Result of the Second Experiment

4.2.1 Data Analysis and Processing

In the data pre-processing stage, we converted raw ECG data collected by polar H10 to HRV features. And the audio data collected by the recorder were noise-reduced by frequency domain filtering and then calculated to mid-term features using pyAudioAnalysis [9].

We analyzed the quality and distribution of the data and listened to the audio to judge whether the dialogue accorded with the collaborators' self-report from the perspective of the third party. According to the result of data analysis, we manually select the period that best fits the target atmosphere and slice it by the length of 1 minute. Then we reduced the data of each section to one row of data per second. Sections that are obviously inconsistent with self-report will not be used.

For labels, we mapped the value of the five-point scale self-report of dialogue mood into two classes, namely True and False. Based on the ratio of labels, we used oversampling and undersampling to balance uneven datasets.

4.2.2 Classification Result

After normalization, PCA, one-hot encoding, we trained SVM, Naive Bayes and KNN for dialogue atmosphere and emotion classification. The Cross-Validation results in Table 4 suggest that the features of ECG and Voice can be used for dialogue atmosphere classification in VR. Moreover, We find that Smooth has better results, while Eminent, Close and Equal have relatively poor classification results. The Cross-Validation results in Table 5 suggest that the features of ECG and audio can be used for emotion classification in VR. Due to insufficient samples, the classification results of Fear and Disgust have large fluctuations. So we need more topics that cause fear or disgust, albeit with some difficulty.

Table 4: Cross-Validation results of dialogue atmosphere classification using SVM, Naive Bayes and KNN

	SVM	Naive Baiyes	KNN
Excite	0.730±0.12	0.672±0.12	0.656±0.12
Serious	0.708±0.12	0.762±0.09	0.729±0.09
Smooth	0.796±0.15	0.775±0.18	0.789±0.18
Bright	0.712±0.15	0.726±0.18	0.683±0.18
Eminent	0.694±0.18	0.677±0.18	0.622±0.16
Close	0.688±0.12	0.689±0.16	0.638±0.22
Equal	0.667±0.12	0.687±0.18	0.621±0.22

Table 5: Cross-Validation results of emotion classification using SVM, Naive Bayes and KNN

	SVM	Naive Baiyes	KNN
Relax	0.684±0.12	0.648±0.22	0.614±0.16
Sadness	0.694±0.12	0.675±0.18	0.576±0.26
Excited	0.667±0.09	0.632±0.18	0.532±0.09
Fear	0.714±0.22	0.739±0.25	0.634±0.18
Disgust	0.578±0.25	0.643±0.25	0.611±0.22
Joy	0.624±0.14	0.591±0.18	0.639±0.14

5. Conclusions

This study uses machine learning algorithms like SVM, KNN, and Naive Bayes to classify emotion and dialogue atmosphere in VR. A limitation of this study is that individual differences appear in biological information, which may relate to the surrounding environment and individual physique.

For future recommendations, we expect to increase the amount of experimental data by conducting experiments with more than three people. Furthermore, applying deep learning algorithms such as CNN and LSTM and combined networks such as CNN-SVM and CNN-LSTM to take advantage of feature extraction and time series. It is also interesting to combine the status of individuals and groups to apply to various VR application scenarios, such as ubiquitous VR innovative spaces and VR virtual humans.

Since there may be some degree of individual differences in biological information, a well-generalizing model that can continuously optimize by extracting personal features can be considered. In the next stage, we tend to recognize the degree of confidence, interest, and psychological fluctuations in multi-person conversations or collaborations in VR.

References

[1] Yuhei Ikeda, Ryota Horie, and Midori Sugaya. Estimating emotion with biological information for robot interaction. *Procedia computer science*, 112:1589–1600, 2017.

[2] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. Human emotion recognition: Review of sensors and methods. *Sensors*, 20(3):592, 2020.

[3] Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific reports*, 8(1):1–15, 2018.

[4] NS Suhaimi, J Teo, and J Mountstephens. Emotional state classification in virtual reality using wearable electroencephalography. In *IOP Conference Series: Materials Science and Engineering*, volume 341, page 012016. IOP Publishing, 2018.

[5] Daisuke Katagami. Could we artificially create atmosphere? In *The 27th Annual Conference of the Japanese Society for Artificial Intelligence(2013)*, page 2 pages. The Japanese Society for Artificial Intelligence, 2013.

[6] Kaoru Toyoda, Yoshihiro Miyakoshi, Ryosuke Yamanishi, and Shohei Kato. Dialogue mood estimation focusing on intervals of utterance state. *Transactions of the Japanese Society for Artificial Intelligence*, 27(2), 2012.

[7] Tatsuya Hasegawa and Shohei Kato. Dialogue mood estimation from speech sounds clusterized with speakers' personality traits. In *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pages 399–401. IEEE, 2019.

[8] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[9] Theodoros Giannakopoulos. pyaudioanalysis: An open-source python library for audio signal analysis. *PloS one*, 10(12):e0144610, 2015.