POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

A New Virtual Reality Experimental Framework for Targeted Emotion Elicitation

Laurea Magistrale in Biomedical Engineering - Ingegneria Biomedica

Authors: Beatrice Prato, Andrea Signorelli

Advisor: Prof. Riccardo Barbieri

Co-advisors: Edoardo Maria Polo, Maximiliano Mollura

Academic year: 2021-2022

1. Introduction

Emotions serve a pivotal role in numerous cognitive functions, including rational decision-making, perception, and learning, among others. The physiological and psychological statuses of humans are profoundly influenced by emotions, thus underscoring the critical importance of recognizing and comprehending them in the context of human behavior research. To facilitate the study of human emotions, researchers must evoke affective states in controlled laboratory environments through various elicitation methods such as images, audio, videos, and more recently, virtual reality (VR) [1].

1.1. Affective Computing

Affective computing, discipline introduced in 1997 by Rosalind Picard, has emerged in the last few decades as a significant area of research that aims at facilitating automatic quantification and recognition of human emotions. The interdisciplinary nature of this field encompasses psychophysiology, computer science, biomedical engineering, and artificial intelligence.

Emotional characterization can be performed using discrete and continuous models. Discrete models rely on the universality of emotional facial expressions to define a set of basic emotions, such as the Pick-A-Mood Model. Continuous models use instead a multidimensional space to represent fundamental emotional characteristics. The Circumplex Model of Affect (CMA) is one of the most used continuous model that represents emotions in a three-dimensional space defined by valence, arousal, and dominance.

Affective computing classifies emotions using biometric signals and machine-learning algorithms. Various signals have been employed, including voice, face, neuroimaging, and physiological measures. The Central Nervous System (CNS) and Autonomic Nervous System (ANS) dynamics are commonly used to classify emotions. EEG measures CNS activity, while signals like HRV, GSR, EMG and respiration have been used to analyse ANS changes.

Identifying characteristic physiological behaviors associated with specific emotions has been the focus of multiple studies [2]. Highly arousing emotional states, such as happiness and fear, are typically associated with sympathetic activation and consequently to an increase in heart rate (HR), a decrease in pressure waves, and galvanic skin response (GSR) activation. Meanwhile low arousal affective states such as sadness and relaxation are linked to correlated to parasympathetic activation, resulting in GSR inhibition, lower HR, and higher pressure waves due to vasodilation.

1.2. Affective Elicitation Methods

Emotion elicitation is crucial for developing systems that detect, interpret and adapt to human affect. Two methods are available: active and passive [3]. Active methods involve directly influencing subjects, while passive methods present external stimuli such as images, sound, or video. Some of the most widely used examples are the IAPS (pictures), IADS (audio) and OASIS (which is used in this study) databases. However, these methods are limited in studying emotions in simulated real-world situations. Immersive VR is a new method that could overcome these limitations.

2. Methods

2.1. Protocol Design

In order to develop a system that can detect, interpret and adapt to human emotions, the elicitation of emotions is crucial. This thesis proposes an emotion elicitation protocol consisting of two techniques: a traditional picture-based elicitation method that shows content through a Flat Screen (FS), and a fully immersive VR elicitation method, the latter accessed through a Meta Quest 2 head-mounted display. Both techniques were developed using the Unity graphic engine to elicit four emotions in an increasing arousal strategy: sadness, relaxation, happiness and fear.

During the experiment, physiological signals were measured from the subjects, including electrocardiogram (ECG), peripheral pulse pressure (ABP), galvanic skin response (GSR), and respiration, and their collection was managed by a ProComp Infiniti device. After each section of the protocol, the subjects also compiled a subjective assessment survey. The FS technique employed images taken from the OASIS database, selected by their pre-validated valence and arousal values, that allowed their positioning on the CMA and determined the emotion that they are expected to elicit in the subjects. Using this criteria, 5 images for each target emotion were selected, and they were identically used for all subjects (to give an example, the five images used for the relax elicitation part are shown in Fig.1).

These same selected images were used as an inspiration to create the immersive VR scenes deputed to eliciting the same emotion, generating continuity between the two methods and allowing a comparison between them (the example continues in fig. 2). In both techniques, to prevent biased reactions from subjects to the elicited stimuli, the emotional sections were not only alternated with neutral sections, but also ordered according to an increasing arousal strategy.

Figure 1: The OASIS images of the FS relax stimulation

Figure 2: The VR relax scene

2.2. Subjective Emotional Assessment

Following each portion of affective elicitation, an emotional assessment survey was conducted through the Microsoft Forms platform. These surveys aimed to collect additional information about the test subjects and to qualitatively evaluate the coherence of the resulting subjective emotional state with the administered elicitation. The surveys were structured into two sections: the first section gathered general information about the subjects, such as age and any previous VR experience, while the second section asked subjects to evaluate emotions elicited by individual sections using SAM and PAM models. To facilitate this, all OASIS images were presented for FS method, and a picture of the VR scene from the subject's point of view during the experiment was shown. Finally, an estimation of the perceived duration of the whole experimental part was enquired.

2.3. Feature Extraction from the Physiological Signals

From the four acquired physiological signals, a total of 30 features were extracted. Starting from the GSR signal, 17 features were extracted both from the Tonic and Phasic component of the signal, and using a Band-pass filtering. As for the cardiac features a Point Process Model of Heart Rate Variability was created [4] and 9 features were computed, both regarding the time component and spectral component of the signal; other cardiac features computed were the Pulse Arrival Time (PAT) and Pulse Pressure (PP). As for the recorded respiration signal, two features were calculated: respiratory frequency and amplitude.

2.4. Statistical Analysis

Statistical analysis was conducted separately for the questionnaire results and extracted physiological features, with both intra- and inter-elicitation methods being evaluated using appropriate statistical tests.
Regarding subjective evaluations, for the flat screen method, SAM-derived arousal and valence coordinates were plotted (considering their mean and confidence interval) against the expected mean points from the OASIS pictures. For the VR method, these coordinates and their confidence intervals were plotted against those of the FS method. This allowed for an assessment of any differences in positioning and accuracy of the emotional elicitation. PAM results were used to qualitatively assess mood changes induced by the experiment.
Regarding the extracted features, their trends throughout the experiment were first observed to confirm the de-biasing effect of baselines. These trends could then allow for depolarization of the features, making it possible to assess their actual response to the single target stimulation in comparison with their preceding baseline. Median trends were plotted and compared not only between the four elicitating parts, but also between the two stimulation methods. Various statistical tests were then performed on the depolarized features, including a Friedman's test followed by a post-hoc with Bonferroni's correction to determine which features were significantly influenced by individual emotion elicitation, and a Wilcoxon signed rank's test to assess method-related differences among the features. In particular, the idea was to get to a first identification of the features that separate low from high levels of arousal or valence, a research topic that was carried out in the next sections of the thesis.

2.5. Feature Selection

Additional analysis was conducted to identify the best features capable of distinguishing between the four emotional states for each elicitation method. A Square Method feature selection approach was applied for this purpose [5]. The best three features for FS and VR methods were selected and their values plotted in a 3D graph that allowed for an easy intuitive analysis of their separation capacity among emotional states.
The same features were also used to analyse their separation ability among low and high arousal and valence states. Eventually, a cross-analysis was carried out by using the best FS features on the VR database and viceversa.

2.6. Machine Learning Classification

The two datasets, each one referring to an elicitation technique, were divided into a train (80%) and a test (20%) sets. Before experimenting with the classification model, additional tests were conducted for feature selection using different techniques besides the square method. These methods included variance evaluation, correlation level between features, and K-Best selection based on an ANOVA F-value.
The sets that resulted from the aforementioned methods were used to train a K-Neighbors classification model, and finally the sets that lead to the highest accuracy were selected as the best ones.
Then, various machine learning models were trained to maximize their classification ability for three different tasks:

- **Multiclass separation**: aims at correctly classifying all 4 emotional states (sadness vs relax vs happiness vs fear);
- **Level of arousal**: binary classification between low and high arousal states (sadness and relax vs happiness and fear);
- **Level of valence**: binary classification between low and high valence states (sadness and fear vs relax and happiness).

3. Results

3.1. Subjective Assessment Results

49 subjects filled out the qualitative assessment questionnaires, their medium age resulted of 23±2.56 years. Starting from the SAM results of the post-FS stimulation, reported in Table 1, for each elicitation part the delta between the mean arousal/valence values and the respective OASIS mean values for each emotion can be seen.

Scene	ΔArousal	ΔValence
Sad	0.48(↑)	0.29(↑)
Rlx	0.28(↑)	0.55(↑)
Hap	-0.65(↓)	-0.26(↓)
Fear	0.04(≃)	0.19(↑)

Table 1: Variations of arousal-valence scores between FS stimulation and used OASIS.

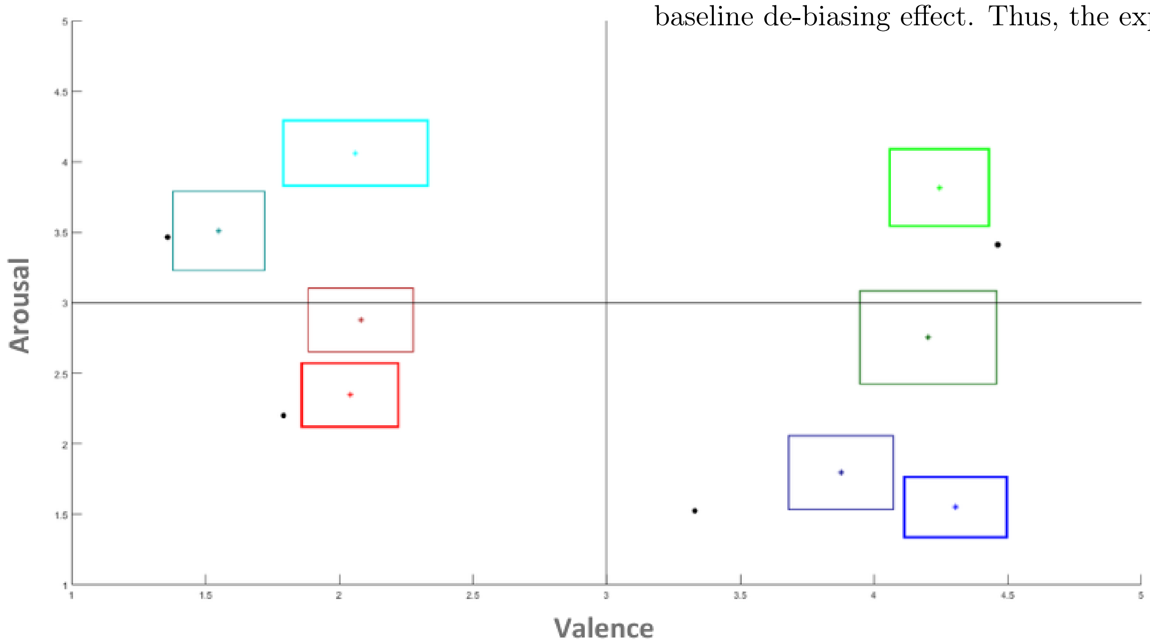

Figure 3: OASIS-FS-VR placement on the CMA

The results of the SAM assessment post-VR are reported in Table 2, following the same organization as the previous comparison, but the deltas are computed between the VR and the respective FS scores.

Scene	ΔArousal	ΔValence
Sad	-0.53 (↓)	-0.04(≃)
Rlx	0.25(↑)	0.43(↑)
Hap	1.05(↑↑)	0.05(≃)
Fear	0.55(↑)	0.51(↑)

Table 2: Variations of arousal-valence scores between VR and FS elicitation.

In figure 3 a graphical representation between the OASIS mean values (black dots), the FS confidence intervals (dark colors) and VR confidence intervals (bright colors).
Talking about the PAM results, they were highly on-point for the post-FS stimulation, while they appear to be more distributed in the post-VR survey. At the same time, they remain coherent with the expected moods and highlight the elicitation of a higher level of arousal, compared with the flat screen stimulation.

3.2. Objective Feature Analysis

44 complete signals were used for this analysis. All 30 features were extracted, and then their median trends were observed, confirming the baseline de-biasing effect. Thus, the expected

depolarization of all elicitating scenes compared to their previous baseline was possible. Median trends of the depolarized experiments showed that VR-induced deltas in the features were generally more pronounced than the FS-induced ones, and that the VR stimulation generated a higher level of arousal in the subjects.
The results of the Friedman's test showed that none of the extracted features for the Flat Screen section were effective in distinguishing the four different elicitated emotions. However, for the VR section, the test was significant for 20 out of 30 extracted features. Further analysis using a Multcompare test with Bonferroni correction identified 17 VR features as capable of distinguishing different emotional states. Specifically, all significant features were able to distinguish Fear from the lower arousal state Relax, while six features were significant for distinguishing Fear from Sadness.
To assess whether there were significant differences between the two elicitation techniques, a Wilcoxon Signed-rank test was conducted between corresponding stimulation parts of the Flat Screen and VR sections of the protocol. The only feature that resulted significantly different for all parts of the two elicitation methods is GSR slope, while others showed a significant difference between two or more scenes. Finally, some features showed highly pronounced differences for only a single emotion.

3.3. Feature Selection and Classification Models

Using the Square Method, the three features that were best in separating the 4 emotional states were identified for each half of the protocol. The values of all subjects of those features, divided for the emotional part in which they were acquired, were plotted in a 3D graph in order to visually assess their separation capacity. The result for FS can be seen in figure 4, while the ones of VR are reported in figure 5.

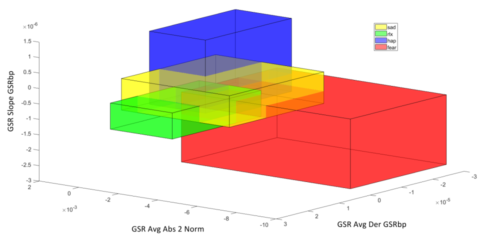

Figure 4: 3D emotion separation based on FS best features

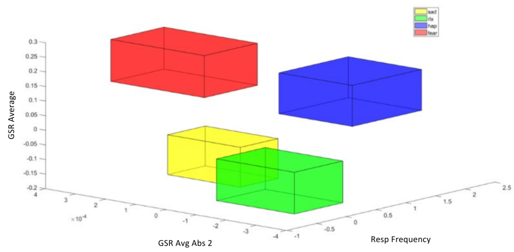

Figure 5: 3D emotion separation based on VR best features

The K-Neighbors classification model achieved the highest accuracy for all tasks in the machine learning classification model, after selecting a different set of best features for each dataset. Arousal separation achieved the highest accuracy, followed by valence separation and multiclass separation for both elicitation methods. The model trained on VR data consistently outperformed the one trained on FS data in all cases.
The accuracy scores for all the performed classifications are reported in Table 3.

Classification	Acc. FS	Acc. VR
Multiclass	25%	48.5%
Binary Valence	47.2%	65.7%
Binary Arousal	61.1%	80%

Table 3: Accuracy scores of classification models

4. Discussion

This work explores the integration of Virtual Reality and Affective Computing using advanced signal processing. A novel experimental protocol was developed to test the validity of VR technology as a medium for targeted emotional triggers in a fully immersive virtual environment, and it was compared against an already validated picture-based method for emotion elicitation.
The results of the study showed that none of the four different emotional parts tested had mean expected values of valence-arousal that fell within the 95% confidence interval. Specifically, the evoked response for happiness fell into the quadrant associated with relax, indicating diffi-

culty in inducing high arousal states. VR didn't show this limitation, probably also because of the induced feeling of presence in the subjects.
Physiological signals and their features were consistent with the literature for both elicitation methods, but those acquired during the VR part showed higher variability. This variability may be helpful in subsequent classifications of emotional states based on their values.
The Friedman's test results, along with subsequent feature selection, provided a deeper understanding of which VR features are more useful for recognizing and separating specific emotional states. This insight was applied in the machine learning classification section, which offered good results despite smaller than optimal training and test datasets.

5. Innovations

This thesis not only proposes a novel Virtual Reality experimental framework for targeted emotion elicitation, but also makes a direct comparison of the new method and a traditional emotion elicitation method. The result is a validation of the use of VR technology in the field of Affective Computing. This work was also one of the first to apply Point Process algorithms to compute cardiac features in the Affective Computing field. A novel definition of physiological feature sets that can differentiate between emotional states was identified, providing insights for emotional classifiers. A last important innovation that was introduced is that the environments used in the VR stimulation were designed to create lifelike scenarios by implementing coherent and realistic audio and visual cues, both static and dynamic, rather than relying on single isolated stimuli as vastly done in literature.

6. Conclusions

In conclusion, this work defines a new VR-based protocol for targeted emotion elicitation that induces more specific responses than traditional Flat Screen picture-based methods. Although further studies are needed to confirm and expand on these insights, this thesis demonstrates the efficacy of Virtual Reality as an emotional elicitation tool and its potential in enhancing the field of Affective Computing. This claim is justified by the more pronounced amplitudes and higher variability of the VR-extracted features, that better explain the changes in the arousal of the subjects and, on a smaller scale, also in in their valence. The 3D emotion separation graphs of FS (fig. 4) and VR (fig. 5) show graphically and intuitively this enhanced ability of the VR-derived features, also confirmed by the machine learning models' performances. This innovative approach proved its ability to better elicit specific emotional responses in the subjects, compared with a traditional Flat Screen picture-based method. VR provides a more immersive and realistic experience for the subjects that is sure to enhance the field of Affective Computing, opening up exciting new opportunities.

References

[1] Javier Marín-Morales, Carmen Llinares, Jaime Guixeres, and Mariano Alcañiz. Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors (Switzerland)*, 20(18):1–26, 2020. ISSN 14248220. doi: 10.3390/s20185163.

[2] Sylvia D. Kreibig. Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3):394–421, 2010. ISSN 03010511. doi: 10.1016/j.biopsycho.2010.03.010.

[3] Rafael Calvo, Jacqueline Kory, and Sidney D Mello. Affect Elicitation for Affective Computing. 2014.

[4] Riccardo Barbieri, Eric C. Matten, Abdul Rasheed A. Alabi, and Emery N. Brown. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. *American journal of physiology. Heart and circulatory physiology*, 288, 1 2005. ISSN 0363-6135. doi: 10.1152/AJPHEART.00482.2003.

[5] Edoardo Maria Polo, Maximiliano Mollura, Marco Zanet, Marta Lenatti, Alessia Paglialonga, and Riccardo Barbieri. Analysis of the Effect of Emotion Elicitation on the Cardiovascular System. *Computing in Cardiology*, 2021-September:2021–2024, 2021. ISSN 2325887X. doi: 10.23919/CinC53138.2021.9662859.